Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1079–1097.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1251–1271.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Particle Spectrum in the Modified Nonminimal Supersymmetric Standard Model in the Strong Yukawa Coupling Regime

R. B. Nevzorov* and M. A. Trusov**

Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia
**e-mail: nevzorov@heron.itep.ru; vxitep.itep.ru*
***e-mail: trusov@heron.itep.ru*

Received April 12, 2000

Abstract—A theoretical analysis of solutions of renormalization group equations in the minimal supersymmetric standard model, which lead to a quasi-fixed point has shown that the mass of the lightest Higgs boson in these models does not exceed 94 ± 5 GeV. This implies that a considerable part of the parameter space in the minimal supersymmetric model is in fact eliminated by existing LEPII experimental data. In the nonminimal supersymmetric standard model the upper bound on the mass of the lightest Higgs boson reaches its maximum in the strong Yukawa coupling regime when the Yukawa constants are substantially greater than the gauge constants on the grand unification scale. In the present paper the particle spectrum is studied using the simplest modification of the nonminimal supersymmetric standard model which gives a self-consistent solution in this region of parameter space. This model can give $m_h \sim 125$ GeV even for comparatively low values of $\tan\beta \geq 1.9$. The spectrum of Higgs bosons and neutralinos is analyzed using the method of diagonalizing mass matrices proposed earlier. In this model the mass of the lightest Higgs boson does not exceed 130.5 ± 3.5 GeV. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The search for the Higgs boson remains one of the top priorities for existing accelerators as well as for those still at the design stage. This is because this boson plays a key role in the standard model which describes all currently available experimental data with a high degree of accuracy. As a result of the spontaneous symmetry breaking $SU(2) \times U(1)$ the Higgs scalar acquires a nonzero vacuum expectation value without destroying the Lorentz invariance, and generates the masses of all fermions and vector bosons. An analysis of the experimental data using the standard model has shown that there is a 95% probability that its mass will not exceed 210 GeV [1]. At the same time, assuming that there are no new fields and interactions and also no Landau pole in the solution of the renormalization-group equations for the self-action constant of Higgs fields up to the scale $M_{\rm Pl} \approx 2.4 \times 10^{18}$ GeV, we can show that $m_h < 180$ GeV [2, 3]. In this case, physical vacuum is only stable provided that the mass of the Higgs boson is greater than 135 GeV [2–6]. However, it should be noted that this simplified model does not lead to unification of the gauge constants [7] and a solution of the hierarchy problem [8]. As a result, the construction of a realistic theory which combines all the fields and interactions is extremely difficult in this case.

Unification of the gauge constants occurs naturally on the scale $M_X \approx 3 \times 10^{16}$ GeV within the supersymmetric generalization of the standard model, i.e., the minimal supersymmetric standard model (MSSM) [7]. In order that all the fundamental fermions acquire mass in the MSSM, not one but two Higgs doublets H_1 and H_2 must be introduced in the theory, each acquiring the nonzero vacuum expectation value v_1 and v_2 where $v^2 = v_1^2 + v_2^2 = (246\ \text{GeV})^2$. The spectrum of the Higgs sector of the MSSM contains four massy states: two *CP*-even, one *CP*-odd, and one charged. An important distinguishing feature of the supersymmetric model is the existence of a light Higgs boson in the *CP*-odd sector. The upper bound on its mass is determined to a considerable extent by the value of $\tan\beta = v_2/v_1$. In the tree approximation the mass of the lightest Higgs boson in the MSSM does not exceed the mass of a Z boson ($M_Z \approx 91.2$ GeV): $m_h \leq M_Z \cos 2\beta$ [9]. Allowance for the contribution of loop corrections to the effective interaction potential of the Higgs fields from a t quark and its superpartners significantly raises the upper bound on its mass:

$$m_h \leq \sqrt{M_Z^2 \cos^2 2\beta + \Delta_{11}^{(1)} + \Delta_{11}^{(2)}}. \qquad (1)$$

Here $\Delta_{11}^{(1)}$ and $\Delta_{11}^{(2)}$ are the single-loop [10] and two-loop [11] corrections, respectively. The values of these corrections are proportional to m_t^4, where m_t is the running mass of a t quark which depends logarithmically on the supersymmetry breaking scale M_S and is almost independent of the choice of $\tan\beta$. In [3, 5, 6] bounds on the mass of the Higgs boson were compared in the

minimal standard and supersymmetric models. The upper bound on the mass of a light *CP*-even Higgs boson in the MSSM increases with increasing $\tan\beta$ and for $\tan\beta \gg 1$ in realistic supersymmetric models with $M_S \leq 1000$ GeV reaches 125–128 GeV.

However, a considerable fraction of the solutions of the system of MSSM renormalization-group equations is focused near the infrared quasi-fixed point at $\tan\beta \sim 1$. In the region of parameter space of interest to us ($\tan\beta \ll 50$) the Yukawa constants of a *b* quark (h_b) and a τ lepton (h_τ) are negligible so that an exact analytic solution can be obtained for the single-loop renormalization-group equations [12]. For the Yukawa constant $h_t(t)$ of a *t* quark and the gauge constants $g_i(t)$ its solution has the following form:

$$Y_t(t) = \frac{\dfrac{E(t)}{6F(t)}}{1 + \dfrac{1}{6Y_t(0)F(t)}}, \quad \tilde{\alpha}_i(t) = \frac{\tilde{\alpha}_i(0)}{1 + b_i\tilde{\alpha}_i(0)t},$$

$$E(t) = \left[\frac{\tilde{\alpha}_3(t)}{\tilde{\alpha}_3(0)}\right]^{16/9}\left[\frac{\tilde{\alpha}_2(t)}{\tilde{\alpha}_2(0)}\right]^{-3}\left[\frac{\tilde{\alpha}_1(t)}{\tilde{\alpha}_1(0)}\right]^{-13/89}, \qquad (2)$$

$$F(t) = \int_0^t E(t')dt',$$

where the index *i* has values between 1 and 3,

$$b_i = 33/5, \quad b_2 = 1, \quad b_3 = -3,$$

$$\tilde{\alpha}_i(t) = \left(\frac{g_i(t)}{4\pi}\right)^2, \quad Y_i(t) = \left(\frac{h_t(t)}{4\pi}\right)^2.$$

The variable *t i*s determined by a standard method: $t = \ln(M_X^2/q^2)$. The boundary conditions for the renormalization-group equations are usually set at the grand unification scale $M_X(t = 0)$ where the values of all three Yukawa constants are the same:

$$\tilde{\alpha}_1(0) = \tilde{\alpha}_2(0) = \tilde{\alpha}_3(0) = \tilde{\alpha}(0).$$

On the electroweak scale where $h_t^2(0) \gg 1$ the second term in the denominator of the expression describing the evolution of $Y_t(t)$ is much smaller than unity and all the solutions are concentrated in a narrow interval near the quasi-fixed point $Y_{\mathrm{QFP}}(t) = E(t)/6F(t)$ [13]. In other words in the low-energy range the dependence of $Y_t(t)$ on the initial conditions on the scale M_X disappears. In addition to the Yukawa constant of the *t*-quark, the corresponding trilinear interaction constant of the scalar fields A_t and the combination of the scalar particle masses $\mathfrak{M}_t^2 = m_Q^2 + m_U^2 + m_2^2$ also cease to depend on $A_t(0)$ and $\mathfrak{M}_t^2(0)$ as $Y_t(0)$ increases. Then on the electroweak scale near the infrared quasi-fixed point $A_t(t)$ and $\mathfrak{M}_t^2(t)$ are only expressed in terms of the gaugino mass on the grand unification scale. Formally this type of solution can be obtained if $Y_t(0)$ is made to go to infinity. Deviations from this solution are determined by the ratio $1/6F(t)Y_t(0)$ which is of the order of $1/10h_t^2(0)$ on the electroweak scale.

The properties of the solutions of the system of MSSM renormalization-group equations and also the particle spectrum near the infrared quasi-fixed point for $\tan\beta \sim 1$ have been studied by many authors [14, 15]. Recent investigations [15–17] have shown that for solutions $Y_t(t)$ corresponding to the quasi-fixed point regime the value of $\tan\beta$ is between 1.3 and 1.8. These comparatively low values of $\tan\beta$ yield significantly more stringent bounds on the mass of the lightest Higgs boson. The weak dependence of the soft supersymmetry breaking parameters $A_t(t)$ and $\mathfrak{M}_t^2(t)$ on the boundary conditions near the quasi-fixed point means that the upper bound on its mass can be calculated fairly accurately. A theoretical analysis made in [15, 16] showed that m_h does not exceed 94 ± 5 GeV. This bound is 25–30 GeV below the absolute upper bound in the minimal supersymmetric model. Since the lower bound on the mass of the lightest Higgs boson from LEPII data is 113 GeV [1], which for the spectrum of heavy supersymmetric particles is the same as the corresponding bound on the mass of the Higgs boson in the standard model, a considerable fraction of the solutions which come out to a quasi-fixed point in the MSSM, are almost eliminated by existing experimental data. This provides the stimulus for theoretical analyses of the Higgs sector in more complex supersymmetric models.

The simplest expansion of the MSSM which can conserve the unification of the gauge constants and raise the upper bound on the mass of the lightest Higgs boson is the nonminimal supersymmetric standard model (NMSSM) [18–20]. In addition to the doublets H_1 and H_2, the Higgs sector of this model contains the additional singlet superfield *Y* relative to the gauge $\mathrm{SU}(2) \times \mathrm{U}(1)$ interactions. The most attractive region of the NMSSM parameter space from the point of view of theoretical analysis is that corresponding to the limit of strong Yukawa coupling when the Yukawa constants on the grand unification scale M_X are substantially larger than the gauge constant g_{GUT}. This is the region where the upper bound on the mass of the lightest Higgs boson reaches its maximum, which is several gigaelectronvolts larger than the corresponding absolute bound in the MSSM. In addition, in this particular case it is possible to select the interaction constants so as to achieve the unification of the Yukawa constants of a *b* quark and a τ lepton on the scale M_X [21, 22] which usually occurs in grand unification theories [23].

However, the mass of the lightest Higgs boson in the NMSSM differs substantially from its upper bound [24]. In this connection, the present paper examines a very sim-

ple model in which m_h reaches its upper theoretical bound for a specific choice of fundamental parameters. This model is obtained by modifying the nonminimal supersymmetric model and yields a self-consistent solution in the strong Yukawa coupling regime where, even for comparatively low values of $\tan\beta \geq 1/9$, the mass of the lightest Higgs boson in the modified NMSSM may reach 125–127 GeV. Although the parameter space of this model is enlarged considerably, the theory does not lose its predictive capacity. The proposed model is used to study characteristics of the spectrum of superpartners of observable particles and Higgs bosons. The mass of the lightest Higgs boson in this model does not exceed 130.5 ± 3.5 GeV.

This bound on the mass of the lightest Higgs boson is not the absolute upper bound in supersymmetric models. For instance, it was shown in [25] that by introducing four or five additional $5 + \bar{5}$ multiplets of matter, the upper bound on m_h in the NMSSM is increased to 155 GeV. Recently the upper bound on the mass of the lightest Higgs boson has been actively discussed using more complex expansions of MSSM theory [26–28]. In particular, in addition to the singlet it is also possible to introduce several SU(2) triplets into the Higgs sector of supersymmetric models. Their appearance destroys the unification of the gauge constants at high energies. In order to reconstruct this, in addition to triplets we also need to add several multiplets of matter which carry color charge in the SU(3) group but do not participate in SU(2) × U(1) interactions, for example $4(3 + \bar{3})$. A numerical analysis made in [27] shows that unification of the gauge constants then occurs on the scale $\tilde{M}_X \sim 10^{17}$ GeV and the mass of the lightest Higgs boson does not exceed 190 GeV. The existence of a fourth generation of particles in the MSSM [28], which is extremely problematical from the point of view of the known experimental data, also leads to an appreciable increase in the upper bound on m_h. Consequently, an increase in the upper bound on the mass of the lightest Higgs boson in the supersymmetric models is usually accompanied by a substantial increase in the number of particles in the models which may be counted as a serious disadvantage of this type of model. In the present study, unlike those noted above [25–28], we examine the dependence of m_h and the particle spectrum on the fundamental parameters of the modified nonminimal supersymmetric model in the strong Yukawa coupling regime.

2. PARAMETERS OF THE NONMINIMAL SUPERSYMMETRIC STANDARD MODEL IN THE STRONG YUKAWA COUPLING REGIME

By definition the superpotential of the nonminimal supersymmetric model is invariant with respect to the discrete transformations $y'_\alpha = \exp(2\pi i/3)y_\alpha$ of the Z_3 group [19] which means that we can avoid the problem of the μ-term in supergravity models. Z_3-symmetry usually occurs in string models in which all the fields of the observable sector remain massless in the exact supersymmetry limit. In addition to observable superfields y_α, supergravity theories also contain a hidden sector in which local supersymmetry is broken. In modern supergravity theories this sector includes singlet dilaton S and moduli T_m fields with respect to gauge interactions. These fields always appear in four-dimensional theory and they occur as a result of the compacting of additional dimensions. The vacuum-averaged dilaton and moduli fields determine the values of the gauge constants on the grand unification scale and also the dimensions and shape of compacted space. The superpotential in supergravity models is usually represented as an expansion in terms of superfields of the observable sector [29]:

$$W = \hat{W}_0(S, T_m) + \frac{1}{2}\mu_{\alpha\beta}(S, T_m)y_\alpha y_\beta + \frac{1}{6}h_{\alpha\beta\gamma}(S, T_m)y_\alpha y_\beta y_\gamma, \quad (3)$$

where $\hat{W}_0(S, T_m)$ is the superpotential of the hidden sector. In Eq. (3) summation is performed over the recurrent Greek subscripts. The requirements for conservation of R-parity [30] and gauge invariance have the result that the single parameter μ is retained in the MSSM which corresponds to the term $\mu(H_1H_2)$ in the superpotential (3). However, the expansion (3) assumes that this fundamental parameter should be of the order of $M_{\rm Pl}$ since this scale is the only dimensional parameter characterizing the hidden (gravity) sector of the theory. In this case, however, the Higgs bosons H_1 and H_2 acquire an enormous mass $m^2_{H_1, H_2} \simeq \mu^2 \simeq M^2_{\rm Pl}$ and no breaking of SU(2) ⊗ U(1) symmetry occurs. In the NMSSM the term $\mu(H_1H_2)$ in the superpotential (3) is not invariant with respect to discrete transformations of the Z_3 group and for this reason should be eliminated from the analysis ($\mu = 0$). As a result of the multiplicative nature of the renormalization of this parameter, the term $\mu(q)$ remains zero on any scale $q \leq M_X$–$M_{\rm Pl}$. However, the absence of mixing of the Higgs doublets on the electroweak scale has the result that H_1 acquires no vacuum expectation value as a result of the spontaneous symmetry breaking and d-type quarks and charged leptons remain massless. In order to ensure that all quarks and charged leptons acquire nonzero masses, an additional singlet superfield Y with respect to gauge SU(2) ⊗ U(1) transformations is introduced in the NMSSM. The superpotential of the Higgs sector of the nonminimal supersymmetric model [18–20] has the form

$$W_h = \lambda Y(H_1H_2) + \frac{\kappa}{3}Y^3. \quad (4)$$

As a result of the spontaneous breaking of SU(2) ⊗ U(1) symmetry, the field Y acquires a vacuum expectation value ($\langle Y\rangle = y/\sqrt{2}$) and the effective μ term ($\mu = \lambda y/\sqrt{2}$) is generated.

In addition to the Yukawa constants λ and κ, and also the standard-model constants, the nonminimal supersymmetric model contains a large number of unknown parameters. These are the so-called soft supersymmetry breaking parameters which are required to obtain an acceptable spectrum of superpartners of observable particles from the phenomenological point of view. The hypothesis on the universal nature of these constants on the grand unification scale allows us to reduce their number in the NMSSM to three: the mass of all the scalar particles m_0, the gaugino mass $M_{1/2}$, and the trilinear interaction constant of the scalar fields A. In order to avoid strong violation of *CP* parity and also spontaneous breaking of gauge symmetry at high energies ($M_{\rm Pl} \gg E \gg m_t$) as a result of which the scalar superpartners of leptons and quarks would acquire nonzero vacuum expectation values, the complex phases of the soft supersymmetry breaking parameters are assumed to be zero and only positive values of m_0^2 are considered. Naturally universal supersymmetry breaking parameters appear in the minimal supergravity model [31] and also in various string models [29, 32]. In the low-energy region the hypothesis of universal fundamental parameters can avoid the appearance of neutral currents with changes in flavor and can simplify the analysis of the particle spectrum as far as possible. The fundamental parameters thus determined on the grand unification scale should be considered as boundary conditions for the system of renormalization-group equations which describes the evolution of these constants as far as the electroweak scale or the supersymmetry breaking scale. The complete system of renormalization-group equations of the nonminimal supersymmetric model can be found in [33, 34]. These experimental data impose various constraints on the NMSSM parameter space which were analyzed in [35, 36].

The introduction of the neutral field Y in the NMSSM potential leads to the appearance of a corresponding F term in the interaction potential of the Higgs fields. As a consequence, the upper bound on the mass of the lightest Higgs boson is increased:

$$m_h \le \sqrt{\frac{\lambda^2}{2} v^2 \sin^2 2\beta + M_Z^2 \cos^2 2\beta + \Delta_{11}^{(1)} + \Delta_{11}^{(2)}}. \quad (5)$$

The relationship (5) was obtained in the tree approximation ($\Delta_{11} = 0$) in [20]. However, loop corrections to the effective interaction potential of the Higgs fields from the *t*-quark and its superpartners play a very significant role. In terms of absolute value their contribution to the upper bound on the mass of a Higgs boson remains approximately the same as in the minimal supersymmetric model. When calculating the corrections $\Delta_{11}^{(1)}$ and $\Delta_{11}^{(2)}$ we need to replace the parameter μ by $\lambda y/\sqrt{2}$. Studies of the Higgs sector in the nonminimal supersymmetric model and the single-loop corrections to it were reported in [24, 33, 36–39]. In [6] the upper bound on the mass of the lightest Higgs boson in the NMSSM was compared with the corresponding bounds on m_h in the minimal standard and supersymmetric models. The possibility of a spontaneous loss of *CP* parity in the Higgs sector of the NMSSM was studied in [39, 40].

It follows from condition (5) that the upper bound on m_h increases as λ increases. Moreover, it only differs substantially from the corresponding bound in the MSSM in the range of small $\tan\beta$. For high values ($\tan\beta \gg 1$) the value of $\sin 2\beta$ tends to zero and the upper bounds on the mass of the lightest Higgs boson in the MSSM and NMSSM are almost the same. The case of small $\tan\beta$ is only achieved for fairly high values of the Yukawa constant of a *t*-quark h_t on the electroweak scale [$h_t(t_0) \ge 1$ where $t_0 = \ln(M_X^2/m_t^2)$ and $\tan\beta$ decreases with increasing $h_t(t_0)$. However, an analysis of the renormalization-group equations in the NMSSM shows that an increase in the Yukawa constants on the electroweak scale is accompanied by an increase in $h_t(0)$ and $\lambda(0)$ on the grand unification scale. It thus becomes obvious that the upper bound on the mass of the lightest Higgs boson in the nonminimal supersymmetric model reaches its maximum in the strong Yukawa coupling limit, i.e., when $h_t(0)$ and $\lambda(0) \gg q_i(0)$.

In our previous two studies [21, 41] we analyzed the renormalization of the NMSSM parameters in the strong Yukawa coupling regime. We showed [21] that as the values of the Yukawa constants on the scale M_X increase, the solutions of the renormalization-group equations on the electroweak scale are pulled toward a quasi-fixed (Hill) line ($\kappa = 0$) or surface ($\kappa \ne 0$) in Yukawa constant space, which limit the range of permissible values of h_t, λ, and κ. Outside this range in the solutions of the renormalization-group equations for $Y_i(t)$ where $Y_t(t) = h_t^2(t)/(4\pi)^2$, $Y_\lambda(t) = \lambda^2(t)/(4\pi)^2$, and $Y_\kappa(t) = \kappa^2(t)/(4\pi)^2$, a Landau pole appears below the grand unification scale and perturbation theory cannot be applied when $q^2 \sim M_X^2$. Along the Hill line or surface the values of $Y_i(t)$ are distributed nonuniformly. As $Y_i(0)$ increases, the region in which the solutions of the renormalization-group equations are concentrated on the electroweak scale in the strong Yukawa coupling regime becomes narrower and in the limit $Y_i(0) \longrightarrow \infty$ all the solutions are focused near the quasi-fixed points. These points are formed as a result of intersection of the Hill line or surface with the infrared fixed (invariant) line. This line connects the stable fixed point in the strong Yukawa coupling regime [42] with the infrared stable fixed point of the system of NMSSM renormal-

ization-group equations [43]. The invariant lines and their properties in the minimal standard and supersymmetric models were studied in detail in [44].

As, with increasing $Y_i(0)$, the Yukawa constants approach the quasi-fixed points, corresponding solutions for the trilinear interaction constants $A_i(t)$ and the following combinations of scalar particle masses $\mathfrak{M}_i^2(t)$:

$$\mathfrak{M}_t^2 = m_Q^2 + m_U^2 + m_2^2,$$
$$\mathfrak{M}_\lambda^2 = m_2^2 + m_1^2 + m_y^2,$$
$$\mathfrak{M}_\kappa^2 = 3m_y^2,$$

cease to depend on their initial values on the scale M_X. If the evolution of the gauge and Yukawa constants is known, the rest of the renormalization-group equations of the nonminimal supersymmetry model can be considered as a system of linear equations for the soft symmetry breaking parameters. In order to solve this system of equations we first need to integrate the equations for the gaugino masses and for the trilinear interaction constants of the scalar fields $A_i(t)$ and then use the results to calculate $\mathfrak{M}_i^2(t)$. Since this system of differential equations for $A_i(t)$ and $\mathfrak{M}_i^2(t)$ is linear, under universal boundary conditions we can obtain the dependence of the soft supersymmetry breaking parameters on the electroweak scale on A_i, $M_{1/2}$, and m_0^2 [45, 46]:

$$A_i(t) = e_i(t)A + f_i(t)M_{1/2},$$
$$\mathfrak{M}_i^2(t) = a_i(t)m_0^2 + b_i(t)M_{1/2}^2 + c_i(t)AM_{1/2} + d_i(t)A^2. \quad (6)$$

The functions $e_i(t)$, $f_i(t)$, $a_i(t)$, $b_i(t)$, $c_i(t)$, and $d_i(t)$ remain unknown since no analytic solution of the complete system of NMSSM renormalization-group equations exists. It was shown in [41] that as the quasi-fixed points are approached, the values of the functions $e_i(t_0)$, $a_i(t_0)$, $c_i(t_0)$, and $d_i(t_0)$ tend to zero whereas for $Y_i(0) \longrightarrow \infty$ all $A_i(t)$ are proportional to $M_{1/2}$ and all $\mathfrak{M}_i^2(t) \propto M_{1/2}^2$. The weak dependence of $A_i(t)$ and $\mathfrak{M}_i^2(t)$ in the strong Yukawa coupling regime on the initial conditions has the result that the solutions of the renormalization-group equations for trilinear interaction constants and combinations of scalar particle masses and also the solutions for $Y_i(t)$ are focused on the electroweak scale near the quasi-fixed points. In general under nonuniversal boundary conditions the solutions for $A_i(t)$ and $\mathfrak{M}_i^2(t)$ are grouped near certain lines ($\kappa = 0$) or planes ($\kappa \neq 0$) in the soft supersymmetry breaking parameter space. These lines and planes are almost perpendicular to the axes A_t and $\mathfrak{M}_t^2$ whereas the planes in the spaces $(A_t, A_\lambda, A_\kappa)$ and $(\mathfrak{M}_t^2, \mathfrak{M}_\lambda^2, \mathfrak{M}_\kappa^2)$ are also almost parallel to the axes A_κ and $\mathfrak{M}_\kappa^2$. Along these lines and planes as $Y_i(0)$ increases, the trilinear interaction constants and combinations of scalar particle masses go to quasi-fixed points.

3. CHOICE OF MODEL

The soft supersymmetry breaking parameters play a key role in an analysis of the particle spectrum in modern supersymmetric models. They destroy the Bose–Fermi degeneracy of the spectrum in supersymmetric theories so that the superpartners of observable particles are substantially heavier than quarks and leptons. However, it should be noted that a study of the particle spectrum in the NMSSM is considerably more complex than a study of this spectrum in the MSSM for $\tan\beta \sim 1$ since two new Yukawa constants λ and κ appear in the nonminimal supersymmetric model for which the boundary conditions are unknown. In turn, the renormalization of the trilinear interaction constants and the scalar particle masses, i.e., the values of the functions $e_i(t_0)$, $f_i(t_0)$, $a_i(t_0)$, $b_i(t_0)$, $c_i(t_0)$, and $d_i(t_0)$, where $t_0 = 2\ln(M_X/M_t^{\text{pole}})$ depends on the choice of $h_t(0)$, $\lambda(0)$, and $\kappa(0)$ on the grand unification scale.

The most interesting from the point of view of a theoretical analysis is a study of the spectrum of heavy supersymmetric particles when the scale of the supersymmetry breaking is $M_S^2 \gg M_Z^2$. This is primarily because in this limit the contribution of new particles to the electroweak observable ones is negligible (see, for example [47]). As has been noted, the standard model highly accurately describes all the existing experimental data. Additional Higgs fields and superpartners of observable particles interacting with vector $W^\pm$ and Z bosons make a nonzero contribution to the electroweak observables. However, for $M_S^2 \gg M_Z^2$ their contribution is suppressed in a power fashion as $(M_Z/M_S)^2$, where any increase in the scale of the supersymmetry breaking leads to convergence of the theoretical predictions for the strong interaction constant $\alpha_s(M_Z)$ which may be obtained assuming unification of the gauge constants [48], with the results of an analysis of the experimental data [49]. In addition, it should be noted that the mass of the lightest Higgs boson which is one of the central objects of investigation in any supersymmetric model reaches its highest value for $M_S \sim 1$–3 TeV.

Unfortunately, in the strong Yukawa coupling regime in the NMSSM with a minimal set of fundamental parameters it is impossible to obtain a self-consistent solution which on the one hand would lead to a spectrum including heavy superparticles and on the other could give a mass of the lightest Higgs boson greater than that in the MSSM. When calculating the particle spectrum, the

fundamental parameters A, m_0^2, and $M_{1/2}$ on the scale M_X should be selected so that the derivatives of the interaction potential of the scalar fields $V(H_1, H_2, Y)$ with respect to the vacuum expectation values v_1, v_2, and y would be zero at the minimum:

$$\frac{\partial V(v_1, v_2, y)}{\partial v_1} = 0, \quad \frac{\partial V(v_1, v_2, y)}{\partial v_2} = 0, \quad \frac{\partial V(v_1, v_2, y)}{\partial y} = 0. \tag{7}$$

Since the trilinear interaction constants and the scalar particle masses in the strong Yukawa coupling regime are almost independent of A, Eqs. (7) link the vacuum expectation value of the neutral scalar field $\langle Y\rangle$, and the parameters m_0^2 and $M_{1/2}$. The value of $\tan\beta$ is determined using the Yukawa constant of a t quark on the electroweak scale (see below). Then a spectrum of heavy supersymmetric particles is only achieved when $\lambda/\kappa \gg 1$. However, in this region of parameter space the value of m_h^2 becomes negative and the physical vacuum is unstable which can be attributed to the strong mixing of the *CP*-even components of the neutral field Y and the superposition of Higgs doublets $h = H_1\cos\beta + H_2\sin\beta$.

Studies of the particle spectrum in the nonminimal supersymmetric model [33, 45, 46, 50] have shown that a self-consistent nontrivial solution of the system of nonlinear algebraic Eqs. (7) for $|Y| \le 10$ TeV which determines the position of the minimum of the interaction potential of the scalar fields, only exists for $\lambda^2(t_0)$, $\kappa^2(t_0) \lesssim 0.1$. In this case a strict correlation exists between the fundamental parameters of the NMSSM. In particular, in order to ensure that spontaneous breaking of SU(2) $\otimes$ U(1) symmetry occurs and the field Y has a nonzero vacuum expectation value of the field, the condition $|A_\kappa/m_y| \ge 3$ must be satisfied. However, the following inequalities must also be satisfied:

$$A_l^2 \le 3(m_1^2 + m_{E_L}^2 + m_{E_R}^2),$$
$$A_d^2 \le 3(m_1^2 + m_{D_L}^2 + m_{D_R}^2),$$
$$A_u^2 \le 3(m_2^2 + m_{U_L}^2 + m_{U_R}^2).$$

Otherwise, the superpartners of leptons and quarks acquire vacuum expectation values [51]. All these constraints have the result that the ratio $|A/m_0|$ varies between 3 and 4. In [52, 53] the particle spectrum in the NMSSM is analyzed separately for $\tan\beta = m_t(m_t)/m_b(m_t)$ and under nonuniversal boundary conditions.

The limit $h_t^2 \gg \lambda^2, \kappa^2$ in the nonminimal supersymmetric model corresponds to the MSSM [33]. For $\kappa = 0$ the Lagrangian of the Higgs sector of the NMSSM is invariant with respect to the global SU(2) $\otimes$ U(1) $\otimes$ U(1) transformations. As a result of the spontaneous symmetry breaking, only the $U(1)$ symmetry corresponding to electromagnetic interaction remains unbroken, which leads to four massless degrees of freedom. Two of these are eaten by a charged $W^\pm$ boson and one by a Z boson. Ultimately, the spectrum of the nonminimal supersymmetric model for $\kappa = 0$ contains one physical massless state which corresponds to the *CP*-odd component of the field Y. For low values of $\kappa^2 \ll \lambda^2 \ll h_t^2$ the mass of the lightest *CP*-odd boson is nonzero and is proportional to the self-action constant of the neutral superfield Y. If the Yukawa constants are λ and $\kappa \sim 10^{-3}$–10^{-4}, for a certain choice of fundamental parameters the mass of the lightest *CP*-even Higgs boson may be only a few gigaelectronvolts [33, 46, 54]. The main contribution to its wave function is made by the neutral scalar field Y which makes it very difficult to search for this on existing accelerators and those at the design stage since the interaction constants of this type of Higgs boson with gauge bosons and fermions are small. In this limiting case, the lightest stable supersymmetric particle having R parity of -1 is usually the superpartner of the neutral scalar field Y [33, 46].

However, unlike the minimal supersymmetric model, the discrete Z_3 symmetry which can avoid problems of the μ term in the NMSSM has the result that three degenerate vacuums appear in the theory because of the breaking of gauge symmetry. Immediately after a phase transition on the electroweak scale the Universe is filled equally with three degenerate phases. The entire space is then divided into separate regions in each of which a particular phase is achieved. The regions are separated by domain walls with the surface energy density $\sigma \sim v^3$. Data from cosmological observations eliminate the existence of domain walls. The domain structure of vacuum in the NMSSM is destroyed if the vacuum degeneracy [55] caused by the Z_3 symmetry disappears. It was shown in [56] that breaking of Z_3 symmetry by introducing into the NMSSM Lagrangian nonrenormalizable operators of dimension $d = 5$ which do not break the SU(2) $\otimes$ U(1) symmetry can be used to obtain splitting of initially degenerate vacuums such that the domain walls disappear before the beginning of the nucleosynthesis era ($T \sim 1$ MeV). Although operators of dimension $d = 5$ are suppressed with respect to M_{Pl} in supergravity models, their introduction leads to quadratic divergences in the two-loop approximation, i.e., to the problem of hierarchies. Consequently, linear and quadratic terms with respect to superfields are generated in the superpotential of this theory and the vacuum expectation value of the neutral scalar field Y is of the order of 10^{11} GeV.

In order to avoid a vacuum domain structure and obtain a self-consistent solution in the strong Yukawa coupling regime, we need to modify the nonminimal supersymmetric model. The NMSSM can be modified most simply by introducing additional terms in the super-

potential of the Higgs sector: $\mu(H_1H_2)$ and $\mu' Y^2$ which are not forbidden by gauge SU(2) ⊗ U(1) and *R* symmetries. The additional bilinear terms in the NMSSM superpotential destroy the Z_3 symmetry and no domain structures appear in this theory since no system of degenerate vacuums exists. The introduction of the parameter μ ensures that it is possible to obtain a spectrum of heavy supersymmetric particles in the strong Yukawa coupling regime in the modified model and for a certain choice of μ' the mass of the lightest Higgs boson reaches its upper bound. In this case the mass of the lightest Higgs boson has its highest value for $\kappa = 0$ since as $\kappa(t_0)$ increases, the upper bound on its mass is reduced as a result of a decrease in $\lambda(t_0)$. In the limit $\kappa = 0$ the *CP*-odd Higgs sector of the modified NMSSM contains no physical massless states since in this case, the global symmetry of the Lagrangian is the same as the local symmetry which eliminates the Yukawa self-action constant of the neutral field *Y* from the analysis. Assuming that this is zero and neglecting all the Yukawa constants except for λ and h_t, the complete superpotential of the modified NMSSM can be expressed in the following form:

$$W_{\text{NMSSM}} = \mu(H_1H_2) + \mu' Y^2 + \lambda Y(H_1H_2) + h_t(H_2Q)U_R^C, \quad (8)$$

where U_R^C is the charge-coupled right superfield of a *t* quark and *Q* is a doublet of left superfields of *b* and *t* quarks.

In supergravity models, bilinear terms with respect to the superfields may be generated in the superpotential (8) as a result of the additional term $Z(H_1H_2)$ + h.c. in the Keller potential [57, 58] or nonrenormalizable interaction of the fields of the observable and hidden sectors. The appearance of nonrenormalizable operators of this type in the superpotential of supergravity models may be attributed to nonperturbative effects (for instance, gaugino condensation) [58, 59]. In addition to the parameters μ and μ', this model also sees the appearance of the corresponding bilinear interaction constants of the scalar fields *B* and *B*' which for a minimal choice of fundamental parameters should be assumed to the equal on the grand unification scale. Thus, the nonminimal supersymmetric model may include seven fundamental parameters in addition to the constants of the standard model:

$$\lambda,\ \mu,\ \mu',\ A,\ B,\ m_0,\ M_{1/2}.$$

4. CONSTRAINTS ON THE PARAMETER SPACE OF THE MODIFIED NONMINIMAL SUPERSYMMETRIC STANDARD MODEL

Despite a substantial expansion of the parameter space, the theory does not lose its predictive capacity. An analysis of the behavior of the solutions of the NMSSM renormalization-group equations in the strong Yukawa coupling limit for $\kappa = 0$ showed that for $Y_i(0) \longrightarrow \infty$ all the solutions are concentrated near the quasi-fixed point:

$$\rho_t^{QFP}(t_0) = 0.803, \quad \rho_{A_t}^{QFP}(t_0) = 1.77,$$
$$\rho_{\mathfrak{M}_t^2}^{QFP}(t_0) = 6.09, \quad \rho_\lambda^{QFP}(t_0) = 0.224, \quad (9)$$
$$\rho_{A_\lambda}^{QFP}(t_0) = -0.42, \quad \rho_{\mathfrak{M}_\lambda^2}^{QFP}(t_0) = -2.28,$$

where

$$\rho_t(t) = \frac{Y_t(t)}{\tilde{\alpha}_3(t)}, \quad \rho_\lambda(t) = \frac{Y_\lambda(t)}{\tilde{\alpha}_3(t)},$$
$$\rho_{A_i}(t) = \frac{A_i(t)}{M_{1/2}}, \quad \rho_{\mathfrak{M}_t^2}(t) = \frac{\mathfrak{M}_\lambda^2(t)}{M_{1/2}^2}.$$

Thus, at the first stage of the analysis we fixed the initial values of the Yukawa constants $\lambda^2(0) = h_t^2(0) = 10$ corresponding to the quasi-fixed point regime (9) of the renormalization-group equations and also the supersymmetry breaking scale $M_3(1000\ \text{GeV}) = 1000$ GeV which determines the mass scale of all the supersymmetric particles.

Existing FNAL experimental data from measurements of the mass of a *t* quark can uniquely relate $\tan\beta$ to the Yukawa constant h_t of a *t* quark. The running mass of a *t* quark generated when the SU(2) × U(1) symmetry is broken is directly proportional to $h_t(t_0)$:

$$m_t(M_t^{\text{pole}}) = \frac{h_t(M_t^{\text{pole}})}{\sqrt{2}} v \sin\beta. \quad (10)$$

However, the value of $m_t(M_t^{\text{pole}})$ calculated in the $\overline{MS}$ scheme [60] is equal to $m_t(M_t^{\text{pole}}) = 165 \pm 5$ GeV. The inaccuracy in determining the running mass of a *t*-quark is primarily attributable to the experimental error with which its pole mass is measured ($M_t^{\text{pole}} = 174.3 \pm 5.1$ GeV [61]). For each fixed set of boundary conditions $h_t(0)$ and $\lambda(0)$, using renormalization-group equations we can calculate the Yukawa constant of a *t*-quark on the electroweak scale and then, substituting the value obtained $h_t(t_0)$ into formula (10), we can determine the value of $\tan\beta$. In the infrared quasi-fixed point regime we obtain $\tan\beta \approx 1.88$ for $m_t(M_t^{\text{pole}}) = 165$ GeV (Tables 1 and 2).

An additional constraint which fixes M_3 can be used to determine one of the supersymmetry breaking parameters $M_{1/2}$. The values of all the other dimensional parameters μ, μ', *A*, *B*, and m_0 should be selected so that spontaneous breaking of gauge SU(2) ⊗ U(1) symmetry occurs on the electroweak scale. The complete inter-

Table 1. Mass spectrum of superpartners of observable particles and Higgs bosons for $\lambda^2(0) = h_t^2(0) = 10$ and $\mu_{\text{eff}} > 0$ as a function of the choice of fundamental parameters A, m_0, and $M_{1/2}$ (all the parameters and masses are in gigaelectronvolts)

m_0^2	0	$M_{1/2}^2$	0	0	0	0
A	0	0	$-M_{1/2}$	$0.5M_{1/2}$	0	0
$M_{1/2}$	–392.8	–392.8	–392.8	–392.8	–785.5	–196.4
$m_t(t_0)$	165	165	165	165	165	165
$\tan\beta$	1.883	1.883	1.883	1.883	1.883	1.883
μ_{eff}	728.6	841.7	726.8	730.1	1361.2	380.4
B_0	–1629.1	–1935.4	–1260.0	–1813.2	–3064.4	–861.8
y	–0.00037	–0.00021	–0.00043	–0.00035	–0.00006	–0.00233
$\mu'(t_0)$	–1899.8	–2176.7	–1905.9	–1898.3	–3544.6	–993.1
$m_h(t_0)$*	**125.0**	**125.1**	**125.0**	**125.0**	**134.9**	**114.8**
$m_h(t_0)$**	**118.4**	**118.5**	**118.4**	**118.4**	**123.2**	**111.9**
M_3 (1 TeV)	1000	1000	1000	1000	2000	500
$m_{\tilde{t}_1}$ (1 TeV)	840.6	889.7	841.1	840.3	1652.0	447.4
$m_{\tilde{t}_2}$ (1 TeV)	695.1	713.6	696.6	694.3	1366.2	371.6
m_H (1 TeV)	898.5	1080.5	895.4	900.3	1691.0	468.8
m_S (1 TeV)	2623.4	3034.3	2452.2	2706.0	4901.7	1378.0
m_{A_1} (1 TeV)	953.9	1113.8	1245.7	925.2	1722.6	538.2
m_{A_2} (1 TeV)	704.3	762.7	872.0	318.2	1366.2	302.2
$m_{\tilde{\chi}_1}(t_0)$	164.6	164.4	164.6	164.6	326.9	84.3
$m_{\tilde{\chi}_2}(t_0)$	327.8	327.6	327.8	327.8	649.4	170.1
$m_{\tilde{\chi}_3}$ (1 TeV)	755.1	870.8	753.3	756.7	1404.2	400.9
$\lvert m_{\tilde{\chi}_4}$ (1 TeV)$\rvert$	755.9	872.6	755.1	758.4	1405.0	404.3
$\lvert m_{\tilde{\chi}_5}$ (1 TeV)$\rvert$	1931.8	2212.3	1938	1930.3	3599.0	1015.4
$m_{\tilde{\chi}_1^\pm}(t_0)$	327.8	327.6	327.8	327.8	649.4	169.9
$m_{\tilde{\chi}_2^\pm}$ (1 TeV)	757.0	872.6	755.2	758.5	1405.2	404.5

*Single-loop approximation.
**Two-loop approximation.

action potential of the Higgs fields in the modified NMSSM can be expressed as the sum:

$$\begin{aligned} V(H_1, H_2, Y) &= \mu_1^2|H_1|^2 + \mu_2^2|H_2|^2 + \mu_y|Y|^2 \\ &+ [\mu_3^2(H_1H_2) + \mu_4^2Y^2 + \lambda A_\lambda Y(H_1H_2) \\ &+ \lambda\mu' Y^*(H_1H_2) + \lambda\mu Y(|H_1|^2 + |H_2|^2) + \text{h.c.}] \\ &+ \lambda^2|(H_1H_2)|^2 + \lambda^2Y^2(|H_1|^2 + |H_2|^2) \\ &+ \frac{q'^2}{8}(|H_2|^2 - |H_1|^2)^2 \\ &+ \frac{q^2}{8}(H_1^+\sigma_a H_1 + H_2^+\sigma_a H_2)^2 + \Delta V(H_1, H_2, Y), \end{aligned} \tag{11}$$

where $\Delta V(H_1, H_2, Y)$ are the single-loop corrections to the effective interaction potential; g and g' are the constants of the gauge SU(2) and U(1) interactions ($g_1 = \sqrt{5/3}\,q'$). The constants μ_i^2 in the interaction potential (11) are related to the soft supersymmetry breaking parameters as follows:

$$\mu_1^2 = m_1^2 + \mu^2, \quad \mu_2^2 = m_2^2 + \mu^2, \quad \mu_y^2 = m_y^2 + \mu'^2,$$

$$\mu_3^2 = B\mu, \quad \mu_4^2 = \frac{1}{2}B'\mu',$$

where

$$m_1^2(M_X) = m_2^2(M_X) = m_y^2(M_X) = m_0^2,$$
$$B(M_X) = B'(M_X) = B_0.$$

Table 2. As Table 1 but for $\mu_{\text{eff}} < 0$

m_0^2	0	$M_{1/2}^2$	0	0	0	0
A	0	0	$-M_{1/2}$	$M_{1/2}$	0	0
$M_{1/2}$	–392.8	–392.8	–392.8	–392.8	–785.5	–196.4
$m_t(t_0)$	165	165	165	165	165	165
$\tan\beta$	1.883	1.883	1.883	1.883	1.883	1.883
μ_{eff}	–727.8	–840.9	–726.0	–731.2	–1360.7	–378.9
B_0	1008	1320.3	1366.7	647.9	2050.4	495.8
y	–0.00149	–0.001	–0.00128	–0.00177	–0.00020	–0.0112
$\mu'(t_0)$	1671.5	1950.6	1656.8	1690.3	3172.7	857.8
$m_h(t_0)$*	**134.1**	**134.9**	**134.0**	**134.2**	**143.1**	**124.1**
$m_h(t_0)$**	**124.4**	**124.8**	**124.3**	**124.5**	**127.2**	**119.6**
M_3 (1 TeV)	1000	1000	1000	1000	2000	500
$m_{\tilde{t}_1}$ (1 TeV)	890.2	935.6	890.5	889.8	1682.8	507.9
$m_{\tilde{t}_2}$ (1 TeV)	630.3	652.2	632.2	628.0	1328.1	283.5
m_H (1 TeV)	896.2	1078.5	893.5	899.3	1689.9	464.4
m_S (1 TeV)	2147.4	2565.9	2309.2	1972.3	4126.5	1097.7
m_{A_1} (1 TeV)	1123.2	1219.3	931.0	1437.9	1984.8	623.1
m_{A_2} (1 TeV)	857.6	1017.8	545.0	886.9	1657.5	412.8
$m_{\tilde{\chi}_1}(t_0)$	160.0	160.5	160.0	160.0	324.4	74.9
$m_{\tilde{\chi}_2}(t_0)$	311.1	313.7	311.0	311.2	639.9	141.4
$\lvert m_{\tilde{\chi}_3}$ (1 TeV)$\rvert$	753.7	896.6	751.9	757.2	1403.4	398.5
$m_{\tilde{\chi}_4}$ (1 TeV)	764.7	878.1	763.0	768.1	1410.0	416.7
$m_{\tilde{\chi}_5}$ (1 TeV)	1700.7	1983.2	1685.8	1719.6	3221.8	879.1
$m_{\tilde{\chi}_1^\pm}(t_0)$	310.7	313.4	310.7	310.8	639.8	139.4
$m_{\tilde{\chi}_2^\pm}$ (1 TeV)	763.3	877.0	761.6	766.7	1409.1	414.5

*Single-loop approximation.
**Two-loop approximation.

The position of the physical minimum of the interaction potential of the Higgs fields (11) is determined by Eqs. (7). Since the vacuum expectation value v and $\tan\beta$ are known, the system of Eqs. (7) can be used to find μ and B_0. Then, it is convenient to introduce $\mu_{\text{eff}} = \mu + \lambda y/\sqrt{2}$ instead of μ. After various transformations we obtain

$$\mu_{\text{eff}}^2 = \frac{m_1^2 - m_2^2\tan^2\beta + \Delta_Z(\mu_{\text{eff}})}{\tan^2\beta - 1} - \frac{1}{2}M_Z^2,$$

$$\left[m_1^2 + m_2^2 + 2\mu_{\text{eff}}^2 + \frac{\lambda^2}{2}v^2 + \Delta_\beta(\mu_{\text{eff}})\right]\sin 2\beta = -2\left(B\mu + \frac{\lambda y X_2}{\cos 2\beta}\right), \quad (12)$$

$$y(m_y^2 + \mu'^2 + B'\mu') = \frac{\lambda}{2}v^2 X_1 - \Delta_y(\mu_{\text{eff}}),$$

where Δ_i corresponds to the contribution of the single-loop corrections:

$$\Delta_\beta = \frac{2}{v^2\tan^2 2\beta}\frac{\partial \Delta V}{\partial \beta} + 4\frac{\partial \Delta V}{\partial v^2}, \quad \Delta_y = \frac{\partial \Delta V}{\partial y},$$

$$\Delta_Z = \frac{1}{\cos^2\beta}\left(2\frac{\partial V}{\partial v^2}\cos 2\beta - \frac{1}{v^2}\frac{\partial \Delta V}{\partial \beta}\sin 2\beta\right),$$

and

$$X_1 = \frac{1}{\sqrt{2}}[2\mu_{\text{eff}} + (\mu' + A_\lambda)\sin 2\beta],$$

$$X_2 = \frac{1}{\sqrt{2}}(\mu' + A_\lambda)\cos 2\beta.$$

When calculating the single-loop corrections we shall only take into account the contribution from loops containing a t quark and its superpartners since their

contribution is dominant. In supersymmetric theories each fermion state with a specific chirality has a scalar superpartner. Thus, a t-quark incorporating left and right chiral components has two scalar superpartners, right $\tilde{t}_R$ and $\tilde{t}_L$, which become mixed as a result of the spontaneous breaking of SU(2) ⊗ U(1) symmetry, and this results in the formation of two charged scalar particles having masses $m_{\tilde{t}_1}$ and $m_{\tilde{t}_2}$:

$$m^2_{\tilde{t}_1,\tilde{t}_2} = \frac{1}{2}[m_Q^2 + m_U^2 + 2m_t^2 \pm \sqrt{(m_Q^2 - m_U^2)^2 + 4m_t^2 X_t^2}], \quad (13)$$

where $X_t = A_t + \mu_{\text{eff}}/\tan\beta$. Since $m^2_{\tilde{t}_1}$ and $m^2_{\tilde{t}_2}$ should be positive we have

$$X_t^2 < \frac{1}{m_t^2}(m_Q^2 + m_t^2)(m_U^2 + m_t^2).$$

Otherwise, the quark fields acquire nonzero vacuum expectation values and the gauge SU(2) ⊗ U(1) symmetry of the initial Lagrangian is completely broken, which leads to the appearance of nonzero masses for gluons and photons. The contribution of the single-loop corrections from the t quark and its superpartners to the effective interaction potential of the Higgs fields is expressed only in terms of their masses:

$$\Delta V(H_1, H_2, Y) = \frac{3}{32\pi^2}\left[m^4_{\tilde{t}_1}\left(\ln\frac{m^2_{\tilde{t}_1}}{q^2} - \frac{3}{2}\right) + m^4_{\tilde{t}_1}\left(\ln\frac{m^2_{\tilde{t}_2}}{q^2} - \frac{3}{2}\right) - 2m_t^4\left(\ln\frac{m_t^2}{q^2} - \frac{3}{2}\right)\right]. \quad (14)$$

For this reason all Δ_i are merely functions of μ_{eff} and do not depend on B_0 and y.

Using the first equation of the system (12) we can find μ_{eff}. In this case, the sign of μ_{eff} is not fixed and must be considered as a free parameter in the theory. Substituting this value of μ_{eff} into the two remaining equations of the system (12), we can eliminate B_0 from the number of independent fundamental parameters and calculate the vacuum expectation value of the field Y: $\langle Y\rangle = y/\sqrt{2}$ and in order to find B_0 we need to bear in mind the relationships linking the bilinear interaction constants B and B' in the electroweak scale with B_0 obtained by solving the renormalization-group equations in the modified NMSSM (see Appendix):

$$B(t) = \zeta(t)B_0 + \sigma(t)A + \omega(t)M_{1/2},$$
$$B'(t) = B_0 + \sigma_1(t)A\frac{\mu_0}{\mu_0'} + \omega_1(t)M_{1/2}\frac{\mu_0}{\mu_0'}, \quad (15)$$

where $\zeta(t)$, $\sigma(t)$, $\sigma_1(t)$, $\omega(t)$, and $\omega_1(t)$ are various functions of t, $h_t(0)$, and $\lambda(0)$ which do not depend on the choice of fundamental soft supersymmetry breaking parameters on the grand unification scale, and also on μ_0 and μ_0'. For a fixed sign of μ_{eff} three different solutions of the system of Eqs. (12) exist. However, only one of these is of interest from the physical point of view. It follows from the last equation in (12) that the vacuum expectation value of the field Y is of the order of $y \sim \lambda v^2/M_S$ and for the case of heavy supersymmetric particles $y \ll v$. The other two solutions give an excessively light CP-even Higgs boson which corresponds to fine tuning between the fundamental parameters B_0 and μ'.

The parameters μ_{eff} and B_0 thus determined and also the vacuum expectation value y depend on the choice of A, m_0, and μ'. Thus, at the next stage of the analysis of the modified NMSSM we studied the dependence of the particle spectrum on these fundamental parameters using Eqs. (6) linking $A_i(t_0)$ and $m_i^2(t_0)$ to A and m_0^2. Similarly we investigated the spectrum of superpartners of observable particles and Higgs bosons for other values of the Yukawa constants from the vicinity of the infrared quasi-fixed point. Although for $\tan\beta \le 2$ in the strong Yukawa coupling regime the parameters h_t and λ can be selected so that the Yukawa constants of a b quark and a τ lepton would be the same on the grand unification scale [21, 22], when studying the particle spectrum in the modified NMSSM we do not confine ourselves to the case $R_{b\tau}(0) = 1$, where $R_{b\tau} = h_b(t)/h_\tau(t)$. This condition arises in minimal schemes of gauge interaction unification [23] and imposes very stringent constraints on the parameter space of the model being studied. However, since h_b and h_τ for $\tan\beta \sim 1$ have small absolute values, they can be generated by means of nonrenormalizable operators as a result of the spontaneous symmetry breaking on the scale M_X and in this case, the Yukawa constants of a b quark and a τ lepton may differ.

5. CALCULATIONS OF MASSES OF HIGGS BOSONS AND NEUTRALINOS

We shall first consider the Higgs sector in the modified NMSSM which includes three CP-even states, two CP-odd, and one charged Higgs boson. The determinants of the mass matrices of the CP-odd and charged Higgs bosons go to zero which corresponds to the appearance of two Goldstone bosons which are eaten by massive vector $W^\pm$ and Z bosons during spontaneous breaking of SU(2) × U(1) symmetry. The (3×3) mass matrix of the CP-odd sector is formed by mixing the imaginary parts of the neutral components of the Higgs doublets with the imaginary part of the field Y. However, since the determinant of this matrix is zero, the problem of finding the eigenvalues reduces to solving an

ordinary quadratic equation. The calculated masses of the *CP*-odd states in the modified NMSSM are

$$m^2_{A_1,A_2} = \frac{1}{2}\Big[m_A^2 + m_B^2 \pm \sqrt{(m_A^2 - m_B^2)^2 + 4\Big(\frac{\lambda v}{\sqrt{2}}(\mu' + A_\lambda) + \Delta_0\Big)^2}\Big],$$
$$m_A^2 = m_1^2 + m_2^2 + 2\mu_{\text{eff}}^2 + \frac{\lambda^2}{2}v^2 + \Delta_A, \quad (16)$$
$$m_B^2 = m_y^2 + \mu'^2 - B'\mu' + \frac{\lambda^2}{2}v^2 + \Delta_3,$$

where Δ_i are the single-loop corrections [33, 36, 38].

A more complex situation is encountered in the sector of the *CP*-even Higgs fields which appears as a result of mixing of the real parts of the neutral components of the Higgs doublets and the field *Y*. The determinant of the mass matrix of the *CP*-even sector is nonzero and thus in order to calculate its eigenvalues we need to solve a cubic equation. However, for the case of heavy supersymmetric particles ($M_S \gg M_Z$) in the Higgs field basis

$$\chi_1 = \frac{1}{\sqrt{2}}\cos\beta\,\mathrm{Re}H_1^0 + \frac{1}{\sqrt{2}}\sin\beta\,\mathrm{Re}H_2^0,$$
$$\chi_2 = -\frac{1}{\sqrt{2}}\sin\beta\,\mathrm{Re}H_1^0 + \frac{1}{\sqrt{2}}\cos\beta\,\mathrm{Re}H_2^0, \quad (17)$$
$$\chi_3 = \frac{1}{\sqrt{2}}\mathrm{Re}Y$$

this matrix has a hierarchical structure:

$$M_{ij}^2 = \begin{pmatrix} E_1^2 & 0 & 0 \\ 0 & E_2^2 & 0 \\ 0 & 0 & E_3^2 \end{pmatrix} + \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix}, \quad (18)$$

where

$$E_1^2 = 0, \quad E_2^2 = m_1^2 + m_2^2 + 2\mu_{\text{eff}}^2,$$
$$E_3^2 = m_y^2 + \mu'^2 + B'\mu',$$
$$V_{11} = M_Z^2\cos^2 2\beta + \frac{1}{2}\lambda^2 v^2 \sin^2 2\beta + \Delta_{11},$$
$$V_{12} = V_{21} = \Big(\frac{1}{4}\lambda^2 v^2 - \frac{1}{2}M_Z^2\Big)\sin 4\beta + \Delta_{12},$$
$$V_{13} = V_{31} = \lambda v X_1 + \Delta_{13},$$
$$V_{22} = M_Z^2\sin^2 2\beta + \frac{1}{2}\lambda^2 v^2\cos^2 2\beta + \Delta_A + \Delta_{22},$$
$$V_{23} = V_{32} = \lambda v X_2 + \Delta_{23},$$
$$V_{33} = \frac{1}{2}\lambda^2 v^2 + \Delta_{33}.$$

Here Δ_A and Δ_{ij} are the corrections from loops containing a *t*-quark and its superpartners. The hierarchical structure of the mass matrix means that perturbation theory from quantum mechanics can be used to diagonalize it. The role of the smallness parameters in the perturbation theory are played by the ratios M_Z^2/E_2^2 and M_Z^2/E_3^2. This method of calculating the masses of Higgs bosons in supersymmetric theories was developed in [24]. Also discussed there is the simplest method of obtaining a hierarchical mass matrix in the Higgs field basis (17). A numerical analysis made in this study showed that perturbation theory can be used to calculate the masses of Higgs bosons in the modified NMSSM to within 1 GeV (~1%).

This method can be used to diagonalize the mass matrix of a neutralino which occurs as a result of the mixing of superpartners of gauge bosons W_3 and *B* (or *Z* and γ) with superpartners of neutral Higgs fields. In the basis ($\tilde{B}^0$, $\tilde{W}^3$, $\tilde{H}_1^0$, $\tilde{H}_2^0$, $\tilde{Y}$), this matrix has the following form:

$$\tilde{M}_{ij} = \begin{pmatrix} M_1 & 0 & -A' & B' & 0 \\ 0 & M_2 & C' & -D' & 0 \\ -A' & C' & 0 & \mu_{\text{eff}} & \frac{\lambda}{\sqrt{2}}v\sin\beta \\ B' & -D' & \mu_{\text{eff}} & 0 & \frac{\lambda}{\sqrt{2}}v\cos\beta \\ 0 & 0 & \frac{\lambda}{\sqrt{2}}v\sin\beta & \frac{\lambda}{\sqrt{2}}v\cos\beta & \mu' \end{pmatrix}. \quad (19)$$

The fragment of the 4×4 matrix which includes the first four columns and four rows is the same as the mass matrix of a neutralino in the MSSM with *A*', *B*', *C*', and *D*' given by

$$A' = M_Z\cos\beta\sin\theta_W, \quad B' = M_Z\sin\beta\sin\theta_W,$$
$$C' = M_Z\cos\beta\cos\theta_W, \quad D' = M_Z\sin\beta\cos\theta_W.$$

Using the unitary transformation *U*:

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

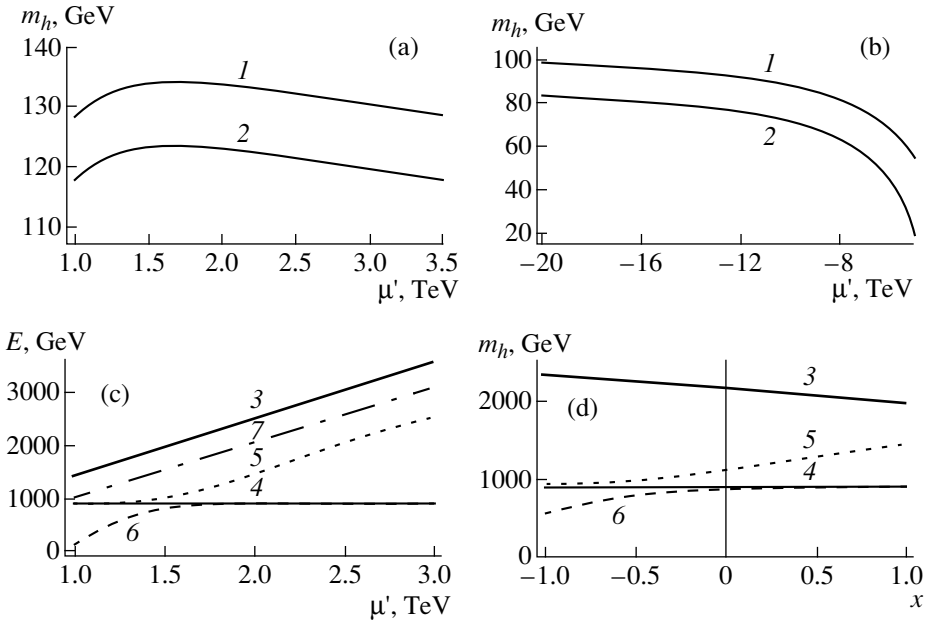


Fig. 1. Mass spectrum in modified NMSSM as a function of μ' and $x = A/M_{1/2}$ for $h_t^2(0) = \lambda^2(0) = 10$, $m_0^2 = 0$, $M_3 = 1$ TeV, and $\mu_{\text{eff}} \le 0$. Curves *1* and *2* give the mass of the lightest Higgs boson calculated in the single-loop and two-loop approximations, respectively. Curves *3* and *4* gives the masses of heavy *CP*-even Higgs bosons m_S and m_H, curves *5* and *6* give the masses of *CP*-odd Higgs bosons m_{A_1} and m_{A_2}, and curve *7* gives the mass of the heaviest neutralino.

the matrix (19) can be reduced to the form (18) $\tilde{M}' = U\tilde{M}U^+$ and then using the ratios $(M_Z/\mu_{\text{eff}})^2$ and $(M_Z/\mu')^2$ as the small parameters in the first order of perturbation theory for the spectrum of supersymmetric particles we obtain

$$m_{\tilde{\chi}_1} = M_1 + \frac{(A'-B')^2}{2(M_1-\mu_{\text{eff}})} + \frac{(A'+B')^2}{2(M_1+\mu_{\text{eff}})},$$

$$m_{\tilde{\chi}_2} = M_2 + \frac{(C'-D')^2}{2(M_2-\mu_{\text{eff}})} + \frac{(C'+D')^2}{2(M_2+\mu_{\text{eff}})},$$

$$m_{\tilde{\chi}_3} = \mu_{\text{eff}} + \frac{(A'-B')^2}{2(\mu_{\text{eff}}-M_1)} + \frac{(C'-D')^2}{2(\mu_{\text{eff}}-M_2)} + \frac{\lambda^2 v^2 \sin^2(\beta+\pi/4)}{2(\mu_{\text{eff}}-\mu')}, \quad (20)$$

$$m_{\tilde{\chi}_4} = -\mu_{\text{eff}} - \frac{(A'+B')^2}{2(M_1+\mu_{\text{eff}})} - \frac{(C'+D')^2}{2(M_2+\mu_{\text{eff}})} - \frac{\lambda^2 v^2 \sin^2(\beta-\pi/4)}{2(\mu_{\text{eff}}+\mu')},$$

$$m_{\tilde{\chi}_5} = \mu' + \frac{\lambda^2 v^2 \sin^2(\beta+\pi/4)}{2(\mu'-\mu_{\text{eff}})} + \frac{\lambda^2 v^2 \sin^2(\beta-\pi/4)}{2(\mu'+\mu_{\text{eff}})}.$$

The accuracy with which $m_{\tilde{\chi}_i}$ is calculated is slightly lower than that for the *CP*-even Higgs sector. This is primarily because the parameters used for the expansion when diagonalizing the neutralino mass matrix according to perturbation theory are larger. At this point we shall not discuss the spectrum of squarks, sleptons, and charginos in greater detail since the analytic expressions for the masses of these particles remain the same as in MSSM. In this case, in all the formulas we need to replace μ with μ_{eff}. We merely note that in the principal approximation the masses of two Dirac charginos and neutralinos $\tilde{\chi}_2$, $\tilde{\chi}_3$ are the same: $m_{\tilde{\chi}_1^\pm} \approx \mu_{\text{eff}}$ and $m_{\tilde{\chi}_2^\pm} \approx M_2$.

6. RESULTS OF A NUMERICAL ANALYSIS

Results of a numerical analysis of the spectrum of Higgs bosons and superpartners of observable particles in the modified NMSSM are given in Figs. 1–3 and Tables 1–3. We first need to note that for a fixed sign of μ_{eff} there are two allowed regions of parameter space. In one of these the mass of the lightest Higgs boson is greater than in the MSSM (see Figs. 1a and 2a) whereas in the other it is smaller (see Figs. 1b and 2b). The mass of the lightest Higgs boson calculated in the first order with respect to perturbation theory has the following form:

$$m_h^2 \approx V_{11} - \frac{|V_{13}|^2}{E_3^2} = M_Z^2\cos^2 2\beta + \frac{1}{2}\lambda^2 v^2 \sin^2 2\beta + \Delta_{11} - \frac{(\lambda v X_1 + \Delta_{13})^2}{m_y^2 + \mu'^2 + B'\mu'}. \quad (21)$$

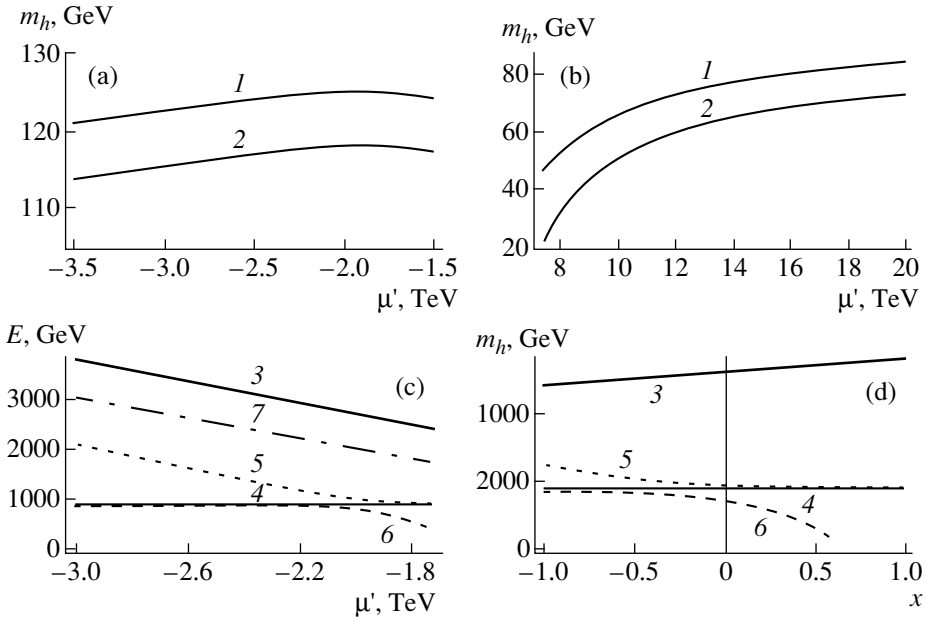


Fig. 2. Mass spectrum in modified NMSSM as a function of μ' and x for $h_t^2(0) = \lambda^2(0) = 10$, $m_0^2 = 0$, $M_3 = 1$ TeV, and $\mu_{\text{eff}} \geq 0$. The notation is the same as in Fig. 1.

Since the matrix element is $V_{12} \sim M_Z^2$, we neglected its contribution to m_h. The mass of the lightest *CP*-even Higgs boson reaches its highest value when

$$\mu' = -\frac{2\mu_{\text{eff}}}{\sin 2\beta} - A_\lambda - \frac{\sqrt{2}\Delta_{13}}{\lambda v \sin 2\beta},$$

where $V_{13} = 0$ (see Figs. 1a and 2a). Thus, m_h is larger in that region of parameter space where the signs of μ' and μ_{eff} are opposed. In the limit $\mu' \longrightarrow \pm\infty$ the masses of the *CP*-even and *CP*-odd Higgs bosons corresponding to the field Y become much larger than the scale of the supersymmetry breaking. In the low energy region their contribution to the effective interaction potential of the Higgs fields H_1 and H_2 disappears and the mass of the lightest Higgs boson is the same as in the MSSM. For this reason, as can be seen from the graphs plotted in Figs. 1a, 1b, and 2a, 2b, m_h reaches a constant value when $\mu' \longrightarrow \pm\infty$. The results of the numerical calculations plotted in these figures indicate that the two-loop corrections [11] play a significant role in the calculations of the mass of the lightest Higgs boson. In this particular case, they reduce its mass approximately by 10 GeV. Although the single-loop corrections increase logarithmically as the scale of the supersymmetry breaking increases, their increase for $M_S \gg M_Z$ is completely compensated by the log–log asymptotic form of the two-loop corrections and m_h remains almost constant. Allowance for the loop corrections has the result that for $\mu_{\text{eff}} < 0$ the mass of the lightest Higgs boson is greater than in that in the case $\mu_{\text{eff}} > 0$. This can be attributed to the fact that m_h increases with increasing mixing in the superpartner sector of the t quark ($\tilde{t}_R$ and $\tilde{t}_L$) which is determined by the value of $X_t = A_t + \mu_{\text{eff}}/\tan\beta$. Since $A_t < 0$, the absolute value of the mixing between $\tilde{t}_R$ and $\tilde{t}_L$ is greater than that for $\mu_{\text{eff}} < 0$. It should be noted that the mass of the lightest Higgs boson is almost independent of A and m_0 because of the weak dependence of the squark mass on the corresponding fundamental parameters (see Tables 1 and 2).

Since that part of the parameter space in which μ_{eff} and μ' have the same sign is almost eliminated by the existing experimental data, it is most interesting to study the spectrum of Higgs bosons in the region where the mass of the lightest Higgs boson is greater than that in the minimal supersymmetric model. In this particular region of parameter space the bilinear soft supersymmetry breaking constants and μ_{eff} have opposite signs and near the maximum of m_h the parameter is $\mu' \sim -2\mu_{\text{eff}}/\sin 2\beta$ ($|\mu'| > |\mu_{\text{eff}}| \sim M_S$). For this reason the heaviest particle in the modified NMSSM spectrum is the *CP*-even Higgs boson which corresponds to the neutral field Y since its mass in the principal approximation with respect to perturbation theory is $m_S^2 \approx E_3^2 > \mu'^2$ and is substantially larger than the scale of the supersymmetry breaking. It can be seen from Figs. 1c and 2c that the mass of the other heavy *CP*-even Higgs boson (m_H) is almost independent of μ' since $m_S^2 \gg m_H^2$.

However, the spectrum of the *CP*-odd Higgs sector is determined to a considerable extent by the choice of fundamental parameters. As μ' increases, the mass of the *CP*-odd *Y* increases and the latter becomes one of the heaviest particles. For values of $\mu' \sim B'$ the mass of the lightest *CP*-odd Higgs boson $m_{A_2}^2 \approx m_B^2$ is very low (see Figs. 1c, 2c), which leads to the appearance of a constraint on μ'. Nevertheless, with this choice of fundamental parameters this Higgs boson is negligibly involved in electromagnetic and weak interactions since the main contribution to its wave function is made by the *CP*-odd component of the field *Y*. Thus, even when its mass is relatively low, it is extremely difficult to detect this particle experimentally. The heaviest fermion in this model is the superpartner of the field *Y*. Its mass $m_{\tilde{\chi}_5}$ is proportional to μ' [see (20)]. The spectrum of remaining neutralinos, charginos, squarks, and sleptons does not depend on the choice of μ'.

Since the dependence of the soft supersymmetry breaking parameters on *A* disappears in the strong Yukawa coupling regime on the electroweak scale, the spectrum of superpartners of the observable particles and also μ_{eff} and *B* whose numerical values are determined by solving the system of Eqs. (12), vary weakly when the trilinear interaction constant of the scalar fields varies between $-M_{1/2}$ and $M_{1/2}$. Despite this, the dependence of the Higgs boson spectrum on *A* is conserved. This is mainly because the bilinear interaction constant of the neutral scalar fields B' is proportional to *A*. Using the relations (15), we obtain

$$B'(t_0) = \frac{1}{\xi(t_0)}B(t_0) + \left[\left(\sigma_1(t_0)\frac{\mu_0}{\mu_0'} - \frac{\sigma(t_0)}{\zeta(t_0)}\right)x + \left(\omega_1(t_0)\frac{\mu_0}{\mu_0'} - \frac{\omega(t_0)}{\zeta(t_0)}\right)\right]M_{1/2},$$

where $x = A/M_{1/2}$. As *x* increases for $\mu_{\text{eff}} < 0$ ($\mu_{\text{eff}} > 0$) the bilinear interaction constant B' increases (decreases) in absolute value and conversely. As $|B'|$ decreases, the masses of the *CP*-even and *CP*-odd states corresponding to the neutral field *Y* converge. At the same time, an increase in the absolute value of B' leads to a decrease in $m_{A_2}^2$ which disappears when $B' \sim \mu'$. The dependence of the Higgs boson spectrum on the parameter *A* for $m_0 = 0$ is studied in Figs. 1d and 2d. The parameter μ' in this particular case is selected so that the mass of the lightest Higgs boson coincides with the upper bound on m_h for $A = 0$.

Although in some cases we assumed $m_0 = 0$ when analyzing the modified NMSSM, this limit is unacceptable from the physical point of view since in this case the lightest (and consequently stable) supersymmetric particle is the superpartner of the right τ lepton which contradicts existing astrophysical observations. However, as m_0 increases, the mass of the superpartner of the right τ lepton increases and even for comparatively low values of $m_0/M_{1/2}$ the lightest particle in the spectrum of superpartners of observable particles becomes the neutralino. The results of the numerical calculations presented in Tables 1 and 2 can be used to assess the influence of the fundamental constants *A*, m_0, and $M_{1/2}$ on the superpartner spectrum of the *t* quark, gluinos, neutralinos, charginos, and Higgs bosons. For each set of parameters listed above we give the values of the upper bound on the mass of the lightest Higgs boson calculated in the single-loop and two-loop approximations and also the corresponding μ_{eff}, B_0, *y*, and μ' for which $V_{13} = 0$. It can be seen from the data presented in Table 1 that the qualitative pattern of the spectrum remains unchanged if the parameters *A* and m_0 vary within reasonable limits. It should also be noted that as m_0^2 increases, the masses of squarks, sleptons, Higgs bosons, and also heavy charginos and neutralinos increases whereas the spectrum of the lightest particles remains unchanged. The mass of a charged Higgs boson which has not been mentioned before is almost independent of *A* and μ' and numerically similar results are obtained for the mass of the *CP*-even state m_H and for the mass of a charged Higgs boson $m_{H^\pm}$.

In the present paper we have made a detailed study of the superpartner and Higgs boson spectrum for initial values of the Yukawa constants $h_t^2(0) = \lambda^2(0) = 10$ corresponding to the scenario of an infrared quasi-fixed point in the NMSSM. The results of the numerical calculations presented in Tables 1 and 2 indicate that for $m_t(M_t^{\text{pole}}) = 165$ GeV and $M_3 \leq 2$ TeV the mass of the lightest Higgs boson does not exceed 127 GeV. Other data presented in Table 3 indicate that the distinguishing features of the supersymmetric particle spectrum are conserved for $h_t^2(0) \gg \lambda^2(0)$ and $h_t^2(0) \ll \lambda^2(0)$ as long as the Yukawa constants on the grand unification scale are substantially larger than the gauge constants. Nevertheless, the upper bound on the mass of the lightest Higgs boson, the value of $\tan\beta$, and the particle masses calculated for

$$\mu' = -\frac{2\mu_{\text{eff}}}{\sin 2\beta} - A_\lambda - \frac{\sqrt{2}\Delta_{13}}{\lambda v \sin 2\beta},$$

when $V_{13} = 0$ vary as a function of the choice of $h_t^2(0)$ and $\lambda^2(0)$. Nevertheless, as $\lambda^2(0)$ decreases from 10 to 2, the upper bound on m_h for $M_3 = 1$ TeV drops from 128 to 113 GeV (see Table 3). Thus, at the concluding stage of the analysis of the modified NMSSM for each fixed $\tan\beta$ we selected the Yukawa constant $\lambda(t_0)$ so that m_h reached its highest value on condition that perturbation theory can be applied as far as the grand uni-

Table 3. As Tables 1 and 2 but for various initial values $h_t^2(0)$ and $\lambda^2(0)$ for $A = m_0 = 0$ and m_t (174 GeV) = 165 GeV

	$\mu_{\text{eff}} < 0$				$\mu_{\text{eff}} > 0$			
$\lambda^2(0)$	0	2	10	10	0	2	10	10
$h_t^2(0)$	10	10	10	2	10	10	10	2
$M_{1/2}$	–392.8	–392.8	–392.8	–392.8	–392.8	–392.8	–392.8	–392.8
$\tan\beta$	1.614	1.736	1.883	2.439	1.614	1.736	1.883	2.439
μ_{eff}	–821.5	–771.4	–727.8	–641.8	822.7	772.4	728.6	642.3
B_0	471.7	622.5	1008.0	886.2	–743.1	–988.1	–1629.1	–1583.3
y	–	–0.0014	–0.0015	–0.0012	–	–0.0003	–0.0004	–0.0005
$\mu'(t_0)$	–	1693.9	1671.5	1749.8	–	–1941.4	–1899.8	–1943.1
$m_h(t_0)$*	**103.5**	**123.6**	**134.1**	**137.6**	**88.1**	**112.4**	**125.0**	**131.2**
$m_h(t_0)$**	**90.3**	**113.0**	**124.4**	**127.8**	**79.7**	**105.5**	**118.4**	**123.6**
M_3 (1 TeV)	1000	1000	1000	1000	1000	1000	1000	1000
$m_{\tilde{t}_1}$ (1 TeV)	894.0	891.6	890.2	890.5	834.6	837.0	840.6	853.5
$m_{\tilde{t}_2}$ (1 TeV)	613.5	622.2	630.3	648.5	692.2	693.8	695.1	696.4
m_H (1 TeV)	1033.4	961.0	896.2	758.5	1035.7	963.3	898.5	761.1
m_S (1 TeV)	–	1999.8	2147.4	2187.2	–	2405.3	2623.4	2663.8
m_{A_1} (1 TeV)	1029.7	1374.8	1123.2	1294.0	1031.3	1390.6	953.9	965.1
m_{A_2} (1 TeV)	–	949.8	857.6	735.6	–	951.6	704.3	674.3
$m_{\tilde{\chi}_1}(t_0)$	160.3	160.1	160.0	159.9	164.6	164.6	164.6	164.4
$m_{\tilde{\chi}_2}(t_0)$	312.7	311.9	311.1	309.4	328.2	328.1	327.8	326.4
$\lvert m_{\tilde{\chi}_3}$ (1 TeV)$\rvert$	842.8	795.8	753.7	665.8	844.4	797.2	755.1	668.1
$\lvert m_{\tilde{\chi}_4}$ (1 TeV)$\rvert$	856.4	807.8	764.7	677.1	850.6	800.9	755.9	666.7
$\lvert m_{\tilde{\chi}_5}$ (1 TeV)$\rvert$	–	1711.2	1700.7	1790.0	–	1960.7	1931.8	1986.5
$m_{\tilde{\chi}_1^\pm}(t_0)$	312.4	311.6	310.7	309.0	328.2	328.1	327.8	326.4
$m_{\tilde{\chi}_2^\pm}$ (1 TeV)	854.2	806.0	763.3	676.7	849.5	800.4	757.0	669.0

*Single-loop approximation.
**Two-loop approximation.

fication scale. The dependence $m_h(\tan\beta)$ thus obtained is plotted in Fig. 3 where we also plotted the upper bound on m_h in the MSSM as a function of $\tan\beta$. As was to be expected, the two bounds on the mass of the lightest Higgs boson are almost the same for large $\tan\beta$ when the term $(1/2)\lambda^2 v^2 \sin^2 2\beta$ in Eq. (5) tends to zero. The curve $m_h(\tan\beta)$ in the NMSSM reaches is maximum when $\tan\beta \sim 2.5$ which corresponds to the strong Yukawa coupling regime. Both bounds on the mass of the lightest Higgs boson were obtained for $M_3 \leq 2$ TeV. By varying the scale of supersymmetry breaking we can show that the mass of the lightest Higgs boson in the NMSSM does not exceed 130.5 ± 3.5 GeV. The indeterminacy observed in calculations of the upper bound on m_h is mainly attributable to the experimental error with which the mass of a *t* quark is measured.

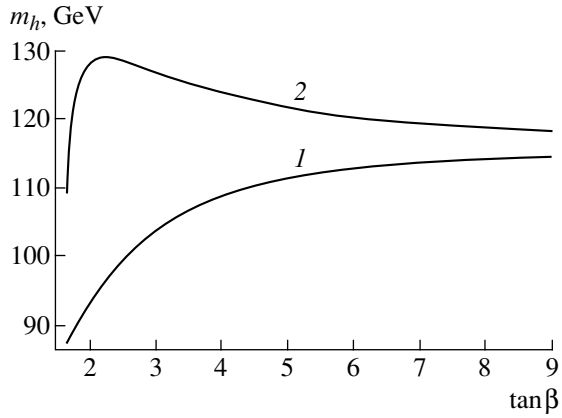


Fig. 3. Upper bound on the mass of the lightest Higgs boson in the MSSM (curve *1*) and in the modified NMSSM (curve *2*) as a function of $\tan\beta$ for M_3 = 2 Tev.

7. CONCLUSIONS

In the nonminimal supersymmetric model the mass of the lightest *CP*-even Higgs boson reaches its highest value in the strong Yukawa coupling regime when all the solutions of the renormalization-group equations are grouped near the infrared quasi-fixed point. However, in this region of the parameter space using the NMSSM with a minimal set of fundamental parameters it is not possible to obtain a self-consistent solution which on the one hand would give a spectrum of heavy supersymmetric particles and on the other could give a mass of the lightest Higgs boson greater than that in the MSSM. In order to find such a solution, we need to modify the nonminimal supersymmetric model. In the present paper we studied the spectrum of superpartners and Higgs bosons using a very simple expansion of the NMSSM which can give a self-consistent solution in the strong Yukawa coupling regime. Although the parameter space of this model is expanded substantially, the theory does not lose its predictive capacity.

The mass matrix of the *CP*-even Higgs sector in the modified NMSSM has a hierarchical structure which means that it can be diagonalized using a method of calculating the spectrum of Higgs bosons proposed earlier, which is based on the ordinary perturbation theory of quantum mechanics. This method can be used to calculate the mass of Higgs bosons to within 1 GeV (~1%). In this case the mass of the lightest Higgs boson near the infrared quasi-fixed point for $m_t(M_t^{\text{pole}}) = 165$ GeV and $M_3 \le 2$ TeV does not exceed 127 GeV. By varying the ratio of the Yukawa constants on the grand unification scale, we can show that $m_h \le 130.5 \pm 3.5$ GeV where the indeterminacy observed when calculating the upper bound on m_h is mainly attributable to the experimental error with which the mass of the *t*-quark is measured. The heaviest particle in the region of parameter space of interest is the *CP*-even Higgs boson corresponding to the neutral field *Y*.

In the present study we used the same method of diagonalizing the mass matrices to calculate and study neutralino masses. As a result we showed that the heaviest fermion in the dominant region of parameter space is $\tilde{Y}$, the superpartner of the neutral scalar field *Y*. For values of $m_0^2 \le M_{1/2}^2$ gluinos, squarks, heavy *CP*-even and *CP*-odd Higgs bosons are substantially heavier than sleptons, lightest charginos, and neutralinos. The only exception is one of the *CP*-odd Higgs bosons whose mass varies substantially depending on the choice of parameters of the model. However, even if it is relatively low, for example, of the order of M_Z, there are certain problems involved in recording it experimentally since the main contribution to its wave function is made by the *CP*-odd component of the field *Y*.

The upper bound on the mass of the lightest Higgs boson in the nonminimal supersymmetric model was also studied in recent publications [25] and [62]. The predictions obtained in these studies are 5–6 GeV higher than the bound given above. The difference in the predictions can be attributed to the fact that the authors of [25] and [62] used the value $|X_t/M_S| = \sqrt{6}$ where $M_S = \sqrt{m_{t_1} m_{t_2}}$ to calculate the upper bound on m_h since the mass of the lightest Higgs boson reaches its highest value for this value of X_t. However, in the strong Yukawa coupling regime in the modified NMSSM the ratio $|X_t/M_S|$ is 1.4–1.5. Since the mass of the lightest Higgs boson increases with increasing mixing between the *t*-quark superpartners for $0 \le |X_t/M_S| \le \sqrt{6}$, and the ratio $|X_t/M_S|$ is considerably less than $\sqrt{6}$, the upper bound on m_h in the realistic expansion of the NMSSM is more stringent that the absolute bound in the nonminimal supersymmetric model.

ACKNOWLEDGMENTS

We are grateful to M.I. Vysotskiĭ, D.I. Kazakov, and K.A. Ter-Martirosyan for stimulating questions, useful discussions, and comments. One of us (R.B.N.) thanks the Italian National Institute of Nuclear Physics (Ferrara Division) for their hospitality. This work was supported by the Russian Foundation for Basic Research (project nos. 98-02-17372, 98-02-17453, 00-15-96786, and 00-15-96562).

APPENDIX

Renormalization-Group Equations for the Parameters μ, μ', B, and B' in the Modified Nonminimal Supersymmetric Standard Model and Their Solution

In addition to the trilinear interaction constants and the scalar particle masses, the modified NMSSM used to study the particle spectrum includes μ, μ', *B*, and *B*'.

The evolution of these constants is described by four renormalization-group equations:

$$\frac{d\mu}{dt} = -\frac{\mu}{2}\left(2Y_\lambda + 3Y_t - 3\tilde{\alpha}_2 - \frac{3}{5}\tilde{\alpha}_1\right),$$

$$\frac{d\mu'}{dt} = -2\mu'(Y_\lambda + Y_\kappa),$$

$$\frac{dB}{dt} = -\left(2Y_\lambda B + \sqrt{Y_\lambda Y_\kappa}B'\frac{\mu'}{\mu} + 3Y_t A_t + 2Y_\lambda A_\lambda - 3\tilde{\alpha}_2 M_2 - \frac{3}{5}\tilde{\alpha}_1 M_1\right), \quad \text{(A.1)}$$

$$\frac{dB'}{dt} = -\left(2Y_\kappa B' + 4\sqrt{Y_\lambda Y_\kappa}B\frac{\mu}{\mu'} + 4Y_\lambda A_\lambda\frac{\mu}{\mu'} + 4Y_\kappa A_\kappa\right).$$

For $\kappa = 0$ and the minimal set of fundamental parameters $B(0) = B'(0) = B_0$, $A_i(0) = A$, $M_i(0) = M_{1/2}$, $\mu'(0) = \mu_0'$, and $\mu(0) = \mu_0$ we can show using the general solution of the system of linear differential equations that $B(t)$, $B'(t)$, $\mu(t)$, and $\mu'(t)$ are given by

$$\mu(t) = \xi(t)\mu_0,$$

$$\mu'(t) = \xi_1(t)\mu_0',$$

$$B(t) = \zeta(t)B_0 + \sigma(t)A + \omega(t)M_{1/2}, \quad \text{(A.2)}$$

$$B'(t) = B_0 + \sigma_1(t)A\frac{\mu_0}{\mu'} + \omega_1(t)M_{1/2}\frac{\mu_0}{\mu_0'},$$

where the functions $\xi(t)$, $\xi_1(t)$, $\zeta(t)$, $\sigma(t)$, $\sigma_1(t)$, $\omega(t)$, and $\omega_1(t)$ determining the evolution of the fundamental parameters depend mainly on the choice of Yukawa constants on the grand unification scale and do not depend on the initial values of the soft supersymmetry breaking parameters μ_0 and μ_0'.

REFERENCES

1. P. Igo-Kemenes, in *Proceedings of XXX International Conference on High-Energy Physics, Osaka, 2000* (in press).
2. N. Cabibbo, L. Maiani, G. Parisi, and R. Petronzio, Nucl. Phys. B **158**, 295 (1979); M. A. Beg, C. Panagiotakopolus, and A. Sirlin, Phys. Rev. Lett. **52**, 883 (1984); M. Lindner, Z. Phys. C **31**, 295 (1986); B. Schrempp and F. Schrempp, Phys. Lett. B **299**, 321 (1993); B. Schrempp and M. Wimmer, Prog. Part. Nucl. Phys. **37**, 1 (1996); T. Hambye and K. Reisselmann, Phys. Rev. D **55**, 7255 (1997).
3. P. Q. Hung and G. Isidori, Phys. Lett. B **402**, 122 (1997); D. I. Kazakov, Phys. Rep. **320**, 187 (1999).
4. M. Sher, Phys. Rep. **179**, 273 (1989); M. Lindner, M. Sher, and H. W. Zaglauer, Phys. Lett. B **228**, 139 (1989); M. Sher, Phys. Lett. B **317**, 159 (1993); **331**, 448 (1994); C. Ford, D. R. T. Jones, P. W. Stephenson, and M. B. Einhorn, Nucl. Phys. B **395**, 17 (1993); G. Altarelli and G. Isidori, Phys. Lett. B **337**, 141 (1994).
5. N. V. Krasnikov and S. Pokorski, Phys. Lett. B **288**, 184 (1992); J. A. Casas, J. R. Espinosa, and M. Quiros, Phys. Lett. B **342**, 171 (1995).
6. M. A. Díaz, T. A. ter Veldhius, and T. J. Weiler, Phys. Rev. D **54**, 5855 (1996).
7. U. Amaldi, W. de Boer, and H. Fürstenau, Phys. Lett. B **260**, 447 (1991); P. Langaker and M. Luo, Phys. Rev. D **44**, 817 (1991); J. Ellis, S. Kelley, and D. V. Nanopoulos, Nucl. Phys. B **373**, 55 (1992).
8. E. Gildener and S. Weinberg, Phys. Rev. D **13**, 3333 (1976).
9. K. Inoue, A. Kakuto, H. Komatsu, and S. Takeshita, Prog. Theor. Phys. **67**, 1889 (1982); R. Flores and M. Sher, Ann. Phys. **148**, 95 (1983).
10. H. E. Haber and R. Hempfling, Phys. Rev. Lett. **66**, 1815 (1991); Y. Okada, M. Yamaguchi, and T. Yanagida, Prog. Theor. Phys. **85**, 1 (1991); J. Ellis, G. Ridolfi, and F. Zwirner, Phys. Lett. B **257**, 83 (1991); J. Ellis, G. Ridolfi, and F. Zwirner, Phys. Lett. B **262**, 477 (1991); R. Barbieri, M. Frigeni, and F. Caravaglios, Phys. Lett. B **258**, 167 (1991); Y. Okada, M. Yamaguchi, and T. Yanagida, Phys. Lett. B **262**, 54 (1991); M. Drees and M. Nojiri, Phys. Rev. D **45**, 2482 (1992); D. M. Pierce, A. Papadopoulos, and S. Johnson, Phys. Rev. Lett. **68**, 3678 (1992); P. H. Chankowski, S. Pokorski, and J. Rosiek, Phys. Lett. B **274**, 191 (1992); H. E. Haber and R. Hempfling, Phys. Rev. D **48**, 4280 (1993); P. H. Chankowski, S. Pokorski, and J. Rosiek, Nucl. Phys. B **423**, 437 (1994); A. Yamada, Z. Phys. C **61**, 247 (1994); A. Dabelstein, Z. Phys. C **67**, 495 (1995); D. M. Pierce, J. A. Bagger, K. Matchev, and R. Zhang, Nucl. Phys. B **491**, 3 (1997).
11. J. R. Espinosa and M. Quiros, Phys. Lett. B **266**, 389 (1991); R. Hempfling and A. H. Hoang, Phys. Lett. B **331**, 99 (1994); M. Carena, J. R. Espinosa, M. Quiros, and C. E. M. Wagner, Phys. Lett. B **355**, 209 (1995); J. A. Casas, J. R. Espinosa, M. Quiros, and A. Riotto, Nucl. Phys. B **436**, 3 (1995); M. Carena, M. Quiros, and C. E. M. Wagner, Nucl. Phys. B **461**, 407 (1996); H. E. Haber, R. Hempfling, and A. H. Hoang, Z. Phys. C **75**, 539 (1997); S. Heinemeyer, W. Hollik, and G. Weiglein, Phys. Rev. D **58**, 091701 (1998); S. Heinemeyer, W. Hollik, and G. Weiglein, Phys. Lett. B **440**, 296 (1998); **455**, 179 (1999); R. Zhang, Phys. Lett. B **447**, 89 (1999).
12. L. E. Ibáñez and C. López, Phys. Lett. B **126B**, 54 (1983); Nucl. Phys. B **233**, 511 (1984); W. de Boer, R. Ehret, and D. I. Kazakov, Z. Phys. C **67**, 647 (1995).
13. C. T. Hill, Phys. Rev. D **24**, 691 (1981); C. T. Hill, C. N. Leung, and S. Rao, Nucl. Phys. B **262**, 517 (1985).
14. V. Barger, M. S. Berger, P. Ohmann, and R. J. N. Phillips, Phys. Lett. B **314**, 351 (1993); W. A. Bardeen, M. Carena, S. Pokorski, and C. E. M. Wagner, Phys. Lett. B **320**, 110 (1994); V. Barger, M. S. Berger, and P. Ohmann, Phys.

Rev. D **49**, 4908 (1994); M. Carena, M. Olechowski, S. Pokorski, and C. E. M. Wagner, Nucl. Phys. B **419**, 213 (1994); M. Carena and C. E. M. Wagner, Nucl. Phys. B **452**, 45 (1995); S. A. Abel and B. C. Allanach, Phys. Lett. B **415**, 371 (1997); **431**, 339 (1998).

15. G. K. Yeghiyan, M. Jurcĩšin, and D. I. Kazakov, Mod. Phys. Lett. A **14**, 601 (1999); S. Codoban, M. Jurcĩšin, and D. Kazakov, hep-ph/9912504.
16. J. A. Casas, J. R. Espinosa, and H. E. Haber, Nucl. Phys. B **526**, 3 (1998).
17. B. Brahmachari, Mod. Phys. Lett. A **12**, 1969 (1997).
18. P. Fayet, Nucl. Phys. B **90**, 104 (1975); M. I. Vysotskiĭ and K. A. Ter-Martirosyan, Zh. Éksp. Teor. Fiz. **90**, 838 (1986) [Sov. Phys. JETP **63**, 489 (1986)].
19. J. Ellis, J. F. Gunion, H. E. Haber, *et al.*, Phys. Rev. D **39**, 844 (1989).
20. L. Durand and J. L. López, Phys. Lett. B **217**, 463 (1989); L. Drees, Int. J. Mod. Phys. A **4**, 3635 (1989).
21. R. B. Nevzorov and M. A. Trusov, Yad. Fiz. (2001) (in press); Preprint No. 5, ITÉF (Institute of Theoretical and Experimental Physics, Moscow, 2000).
22. B. C. Allanach and S. F. King, Phys. Lett. B **328**, 360 (1994).
23. M. S. Chanowitz, J. Ellis, and M. K. Gaillard, Nucl. Phys. B **128**, 506 (1977); A. J. Buras, J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, Nucl. Phys. B **135**, 66 (1978).
24. P. A. Kovalenko, R. B. Nevzorov, and K. A. Ter-Martirosyan, Yad. Fiz. **61**, 898 (1998) [Phys. At. Nucl. **61**, 812 (1998)].
25. M. Masip, R. Múñoz-Tapia, and A. Pomarol, Phys. Rev. D **57**, 5340 (1998).
26. J. R. Espinosa and M. Quiros, Phys. Lett. B **279**, 92 (1992); **302**, 51 (1993); T. Moroi and Y. Okada, Phys. Lett. B **295**, 73 (1992); G. L. Kane, C. Kolda, and J. D. Wells, Phys. Rev. Lett. **70**, 2686 (1993); J. R. Espinosa, Phys. Lett. B **353**, 243 (1995); D. Comelli and J. R. Espinosa, Phys. Lett. B **388**, 793 (1996); Y. Sakamura, Mod. Phys. Lett. A **14**, 721 (1999).
27. J. R. Espinosa and M. Quiros, Phys. Rev. Lett. **81**, 516 (1998); M. Masip, Phys. Lett. B **444**, 352 (1998).
28. S. K. Kang, Phys. Rev. D **54**, 7077 (1996); S. K. Kang and G. T. Park, Mod. Phys. Lett. A **12**, 553 (1997); D. Dooling, K. Kang, and S. K. Kang, Int. J. Mod. Phys. A **14**, 1605 (1999); Phys. Rev. D **60**, 017701 (1999).
29. V. S. Kaplunovsky and J. Louis, Phys. Lett. B **306**, 269 (1993); A. Brignole, L. E. Ibáñez, and C. Múñoz, Nucl. Phys. B **422**, 125 (1994); **436**, 747 (1995).
30. P. Fayet, Phys. Lett. B **69B**, 489 (1977); G. Farrar and P. Fayet, Phys. Lett. B **76B**, 575 (1978).
31. H. P. Nilles, M. Srednicki, and D. Wyler, Phys. Lett. B **120B**, 345 (1983); R. Barbieri, S. Ferrara, and C. Savoy, Phys. Lett. B **119B**, 343 (1982); A. H. Chamseddine, R. Arnowitt, and P. Nath, Phys. Rev. Lett. **49**, 970 (1982).
32. K. Choi, H. B. Kim, and C. Múñoz, Phys. Rev. D **57**, 7521 (1998); A. Lukas, B. A. Ovrut, and D. Waldram, Phys. Rev. D **57**, 7529 (1998); T. Li, Phys. Rev. D **59**, 107902 (1999).
33. S. F. King and P. L. White, Phys. Rev. D **52**, 4183 (1995).
34. J.-P. Derendinger and C. A. Savoy, Nucl. Phys. B **237**, 307 (1984).
35. F. Franke and H. Fraas, Phys. Lett. B **353**, 234 (1995); S. F. King and P. L. White, Phys. Rev. D **53**, 4049 (1996).
36. S. W. Ham, S. K. Oh, and B. R. Kim, Phys. Lett. B **414**, 305 (1997).
37. P. N. Pandita, Phys. Lett. B **318**, 338 (1993); Z. Phys. C **59**, 575 (1993); S. W. Ham, S. K. Oh, and B. R. Kim, J. Phys. G **22**, 1575 (1996).
38. T. Elliott, S. F. King, and P. L. White, Phys. Lett. B **314**, 56 (1993); Phys. Rev. D **49**, 2435 (1994); U. Ellwanger, Phys. Lett. B **303**, 271 (1993); U. Ellwanger and M. Lindner, Phys. Lett. B **301**, 365 (1993).
39. S. W. Ham, S. K. Oh, and H. S. Song, hep-ph/9910461 (1999).
40. A. Pomarol, Phys. Rev. D **47**, 273 (1993); K. S. Babu and S. M. Barr, Phys. Rev. D **49**, R2156 (1994); G. M. Asatrian and G. K. Egiian, Mod. Phys. Lett. A **10**, 2943 (1995); **11**, 2771 (1996); N. Haba, M. Matsuda, and M. Tanimoto, Phys. Rev. D **54**, 6928 (1996).
41. R. B. Nevzorov and M. A. Trusov, Yad. Fiz. (2001) (in press); Preprint No. 6, ITÉF (Institute of Theoretical and Experimental Physics, Moscow, 2000).
42. P. Binetruy and C. A. Savoy, Phys. Lett. B **277**, 453 (1992).
43. B. C. Allanach and S. F. King, Phys. Lett. B **407**, 124 (1997); I. Jack and D. R. T. Jones, Phys. Lett. B **443**, 177 (1998).
44. B. Schrempp and F. Schrempp, Phys. Lett. B **299**, 321 (1993); B. Schrempp, Phys. Lett. B **344**, 193 (1995); B. Schrempp and M. Wimmer, Prog. Part. Nucl. Phys. **37**, 1 (1996).
45. U. Ellwanger, M. Rausch de Traubenberg, and C. A. Savoy, Phys. Lett. B **315**, 331 (1993).
46. U. Ellwanger, M. Rausch de Traubenberg, and C. A. Savoy, Nucl. Phys. B **492**, 21 (1997).
47. I. V. Gaidaenko, A. V. Novikov, V. A. Novikov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 723 (1998) [JETP Lett. **67**, 761 (1998)].
48. P. Langacker and N. Polonsky, Phys. Rev. D **47**, 4028 (1993); **52**, 3081 (1995).
49. Review of Particle Properties, Eur. Phys. J. C **3**, 1 (1998).
50. T. Elliott, S. F. King, and P. L. White, Phys. Lett. B **305**, 71 (1993); **351**, 213 (1995).
51. U. Ellwanger and C. Hugonie, Phys. Lett. B **457**, 299 (1999).

52. B. Ananthanarayan and P. N. Pandita, Phys. Lett. B **371**, 245 (1996); Int. J. Mod. Phys. A **12**, 2321 (1997).

53. P. Brax, U. Ellwanger, and C. A. Savoy, Phys. Lett. B **347**, 269 (1995).

54. U. Ellwanger, M. Rausch de Traubenberg, and C. A. Savoy, Z. Phys. C **67**, 665 (1995).

55. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun', Zh. Éksp. Teor. Fiz. **67**, 3 (1974) [Sov. Phys. JETP **40**, 1 (1975)].

56. S. A. Abel, S. Sarkar, and P. L. White, Nucl. Phys. B **454**, 663 (1995).

57. G. F. Giudice and A. Masiero, Phys. Lett. B **206**, 480 (1988).

58. J. A. Casas and C. Múñoz, Phys. Lett. B **306**, 288 (1993).

59. J. E. Kim and H. P. Nilles, Phys. Lett. B **138B**, 150 (1984); **263**, 79 (1991); E. J. Chun, J. E. Kim, and H. P. Nilles, Nucl. Phys. B **370**, 105 (1992).

60. R. Tarrach, Nucl. Phys. B **183**, 384 (1981); S. Narison, Phys. Lett. B **197**, 405 (1987); N. Gray, D. J. Broadhurst, W. Grafe, and K. Schilcher, Z. Phys. C **48**, 673 (1990).

61. L. Demortier *et al.* (The Top Averaging Group for the CDF and DO Collaboration), Fermilab-TM-2084 (1999).

62. G. K. Yeghiyan, hep-ph/9904488 (1999); U. Ellwanger and C. Hugonie, hep-ph/9909260 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1098–1114.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1272–1290.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Type IIB Green–Schwarz Superstrings in $AdS_5 \times S^5$ from the Supercoset Approach¶

R. R. Metsaev[a] and A. A. Tseytlin[a, b]

[a]*Department of Theoretical Physics, Lebedev Physical Institute, Moscow, 117924 Russia*
e-mail: metsaev@lpi.ru, metsaev@pacific.mps.ohio-state.edu
[b]*Department of Physics, The Ohio State University, Columbus, OH 43210-1106, USA*
e-mail: tseytlin@mps.ohio-state.edu

Received July 21, 2000

Abstract—We consider superstrings moving in the $AdS_5 \times S^5$ space-time and find their Green–Schwarz action using the supercoset approach based on the supergroup PSU(2, 2|4). We describe several parametrizations of the relevant supercoset and present the action in different κ-symmetry gauges. In particular, we discuss a gauge where all the fermionic coordinates corresponding to the conformal (*S*) supercharges are gauged away and also a light-cone type gauge where half of the *Q* and *S* supercoordinates are gauged away. The resulting action contains terms that are quadratic and quartic in fermions. In the flat-space limit, it reduces to the standard light-cone Green–Schwarz action. We comment on the possibility of fixing the bosonic light-cone gauge and of reformulating the action in terms of two-dimensional Dirac spinors. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Further progress in clarifying the relation between strings and large-*N* (non)supersymmetric gauge theories depends on better understanding the superstrings in anti-de Sitter (AdS)-type spaces with Ramond–Ramond (RR) fluxes [1, 2]. The simplest, most symmetric example is provided by the $AdS_5 \times S^5$ background of the type IIB superstring theory. As with the flat superstring, one can hope that, although this theory is quite nontrivially dual to the large-*N* $\mathcal{N} = 4$ super Yang–Mills theory [2], it should be explicitly solvable.

To be able to treat this theory in an efficient way, one should use the Green–Schwarz (GS) approach [3], where one views the string as moving in a superspace with the coordinates (x^m, θ^I_α), where θ are space-time spinors. Then the coupling of the string to a RR background has a local form $\bar\theta\Gamma\cdots\theta\partial x\partial x F\ldots$ similar to its coupling to the curvature (it is useful to note that while the classical string "feels" the metric, it feels the RR background only through its quantum fluctuations).

There seems to be no alternative to the GS description: even the simplest one-loop computation in the GS formalism (e.g., the one described in [4]) would require a summation of an infinite number of diagrams with any number of flat-space RR vertex operator insertions if done in the Neveu–Schwarz–Ramond (NSR) formalism. In fact, the use of the GS action is very natural from the modern "solitonic" point of view. If one views fundamental strings as "electric"-type objects appearing in the $D = 10$ supergravity, the effective action for the collective coordinates of a long fundamental string naturally has a GS form as dictated by the residual space-time (super)symmetries (with the characteristic nonlinear $\partial x\bar\theta\Gamma\cdots\theta$ kinetic term originating from the standard Dirac $D = 10$ fermion action reduced to the world surface of the string, with Dirac matrices becoming the projected ones). The theory must of course contain long and short strings that must also be described by the same manifestly space-time supersymmetric action.

A simple way to construct the flat-space GS action is to view strings as propagating in the flat coset superspace (ten-dimensional super Poincaré group)/(ten-dimensional Lorentz group). The action is then a kind of Wess–Zumino–Witten action with the global supersymmetry and a local κ-symmetry ensuring the correct number of the fermionic degrees of freedom. It is defined in terms of the basic objects given by the left-invariant Cartan-1 forms on the type IIB coset superspace,

$$G^{-1}dG = L^{\hat A}P_{\hat A} + L^I Q_I, \quad L^{\hat A} = dX^M L^{\hat A}_M,$$
$$X^M = (x, \theta), \quad \hat A, \hat B = 0, \ldots, 9, \quad I = 1, 2, \tag{1.1}$$

where $P_{\hat A}$ are the translation generators and $G = G(x, \theta)$ is a coset representative. If it is chosen as

$$G(x, \theta) = \exp(x^{\hat A}P_{\hat A} + \theta^I Q_I), \quad [P_{\hat A}, P_{\hat B}] = 0,$$
$$\{Q_I, Q_J\} = -2i\delta_{IJ}(\not{C}\Gamma^{\hat A})P_{\hat A} \tag{1.2}$$

¶This article was submitted by the authors in English.

(where $\Gamma^{\hat{A}}$ are the gamma matrices and $\mathcal{C}$ is the charge conjugation matrix), the coset space vielbeins are given by

$$L^{\hat{A}} = dx^{\hat{A}} - i\bar{\theta}^I\Gamma^{\hat{A}}d\theta^I, \quad L^I = d\theta^I, \tag{1.3}$$

where $x^{\hat{A}}$ are flat bosonic coordinates and θ^I are two left Majorana–Weyl ten-dimensional spinors. The general form of the action is [5]

$$I = -\frac{1}{2}\int_{\partial M_3} d^2\sigma\sqrt{g}g^{\mu\nu}L_\mu^{\hat{A}}L_\nu^{\hat{A}} + i\int_{M_3}\mathcal{H}, \qquad \mathcal{H} = s^{IJ}L^{\hat{A}}\wedge(\bar{L}^I\Gamma^{\hat{A}}\wedge L^J), \tag{1.4}$$

where M_3 is a three-dimensional manifold with the boundary ∂M_3,

$$s^{IJ} \equiv \mathrm{diag}(1,-1), \quad d\mathcal{H} = 0,$$

and we used the convention that $2\pi\alpha' = 1$. The two-dimensional metric $g_{\mu\nu}$ ($\mu, \nu = 0, 1$) has the signature $(-, +)$ and $g \equiv -\det g_{\mu\nu}$. The first "kinetic" term in the action has a "degenerate" metric containing only the square of the translational Cartan forms, not the spinor ones, because the latter would lead to an action quadratic in fermionic derivatives and thus to a potential nonunitarity. The explicit two-dimensional component form of the GS action is [3]

$$I_0 = \int d^2\sigma\mathcal{L}_0 = \int d^2\sigma\Big[-\frac{1}{2}\sqrt{g} \times g^{\mu\nu}(\partial_\mu x^{\hat{A}} - i\bar{\theta}^I\Gamma^{\hat{A}}\partial_\mu\theta^I)(\partial_\nu x^{\hat{A}} - i\bar{\theta}^J\Gamma^{\hat{A}}\partial_\nu\theta^J) - i\epsilon^{\mu\nu}s^{IJ}\bar{\theta}^I\Gamma^{\hat{A}}\partial_\nu\theta^J\Big(\partial_\mu x^{\hat{A}} - \frac{1}{2}i\bar{\theta}^K\Gamma^{\hat{A}}\partial_\mu\theta^K\Big)\Big], \tag{1.5}$$

where $\epsilon^{\mu\nu}$ is the antisymmetric tensor. The key observation that allows us to construct the GS action for superstrings in $AdS_5 \times S^5$ is that this bosonic space is the coset

$$\frac{G}{H} = \frac{\mathrm{SO}(2,4)}{\mathrm{SO}(1,4)} \times \frac{\mathrm{SO}(6)}{\mathrm{SO}(5)}.$$

As a result, we can consider strings moving not in the flat superspace but in the supercoset space with the correct bosonic coset part and the correct number of supersymmetries: we must replace the ten-dimensional super Poincaré group by a smaller group PSU(2, 2|4) and the Lorentz group SO(1, 9) by its subgroup SO(1, 4) × SO(5); i.e., we need to consider the supercoset [6]

$$\frac{\mathrm{PSU}(2,2|4)}{\mathrm{SO}(1,4)\times\mathrm{SO}(5)}$$

that still has 10 bosonic and 32 Grassmann dimensions.

The bosonic part of the GS action is simply the standard symmetric-space sigma-model action

$$L = \mathrm{Tr}[(g^{-1}\partial g)_{G/H}]^2.$$

This is obviously not a conformal sigma model; the addition of the θ fermions converts it into a novel type of a two-dimensional conformal model. The fermions couple the AdS_5 and S^5 parts of the action together and contribute the RR 5-form F_5 to the beta-function of the metric,

$$R_{MN} - (F_5F_5)_{MN} = 0,$$

making it vanish [6].

The superalgebra psu(2, 2|4) plays the central role in the construction of the GS action in $AdS_5 \times S^5$ [6]. Its even part is the sum of the algebra so(4, 2) that is the isometry algebra of AdS_5 and the algebra so(6) that is the isometry algebra of S^5. The odd part consists of 32 supercharges corresponding to 32 Killing spinors in the $AdS_5 \times S^5$ vacuum [7] of the type IIB supergravity (see [8–10]). In the 5 + 5 splitting, its generators are given by ($\hat{P}_A$, $\hat{J}_{AB}$; $P_{A'}$, $J_{A'B'}$; $Q^{\alpha\alpha'I}$), i.e., by two sets of translations and rotations in so(2, 4) and so(6) ($A = 0, 1, \ldots, 4$; $A' = 1, \ldots, 5$) and 32 supercharges ($\alpha = 1, 2, 3, 4$; $\alpha' = 1, 2, 3, 4$).

Following the same steps as in the above construction of the flat-space GS action, one finds [6] that there exists a unique action of type (1.4) that has the required properties: its bosonic part is a sigma model action on $AdS_5 \times S^5$, it has a local κ-symmetry and the global PSU(2, 2|4) symmetry, and it reduces to the standard GS action in the flat-space limit. The part of the action in [6] that is quadratic in θ^I is a direct generalization of the quadratic term in the flat-space GS action

$$S_F^{(2)} = \frac{i}{2\pi\alpha'} \times \int d^2\sigma(\sqrt{g}g^{\mu\nu}\delta^{IJ} - \epsilon^{\mu\nu}s^{IJ})\bar{\theta}^I\rho_\mu D_\nu\theta^J, \tag{1.6}$$

where ρ_μ are projections of the ten-dimensional Dirac matrices,

$$\rho_\mu \equiv \Gamma_{\hat{m}}E_M^{\hat{m}}\partial_\mu X^M = (\Gamma_A E_M^A + \Gamma_{A'}E_M^{A'})\partial_\mu X^M$$

($E_M^{\hat{m}}$ is the vielbein of the ten-dimensional target space metric). The covariant derivative D_μ is the projection of the ten-dimensional derivative

$$D_M = \partial_M + \frac{1}{4}\omega_M^{\hat{m}\hat{n}}\Gamma_{\hat{m}\hat{n}} - \frac{1}{8\cdot 5!}\Gamma^{M_1\ldots M_5}\Gamma_M e^{\Phi}\Gamma_{M_1\ldots M_5}$$

(with the dilaton field Φ and the spinor connection $\omega_M^{\hat{m}\hat{n}}$) involved in the Killing spinor equation of the type-IIB supergravity,

$$D_\mu\theta^I \equiv \left(\delta^{IJ}D_\mu - \frac{i}{2}\epsilon^{IJ}\tilde{\rho}_\mu\right)\theta^J,$$
$$D_\mu = \partial_\mu + \frac{1}{4}\partial_\mu x^M \omega_M^{\hat{m}\hat{n}}\Gamma_{\hat{m}\hat{n}}, \quad (1.7)$$
$$\tilde{\rho}_\mu \equiv (\Gamma_A E_M^A + i\Gamma_{A'}E_M^{A'})\partial_\mu X^M,$$

where the term involving $\tilde{\rho}_\mu$ originates from the coupling to the RR 5-form field strength.

The full nonlinear form of the action containing terms of higher orders in θ is known explicitly [11, 12]. Fixing a "covariant" symmetry gauge (e.g., $\theta^1 = \Gamma_{0123}\theta^2$), one obtains a relatively simple action that contains terms of only quadratic and quartic orders in the fermions [13, 14]. A similar structure is found in the covariant *S*-gauge discussed in what follows. The resulting action has the same feature as the flat-space GS action: its fermionic kinetic term is coupled to all the bosonic string coordinates. It is nondegenerate when expanded near a "long" string configuration [15]; thus, the GS action provides a well-defined and useful tool for computing quantum corrections to long string configurations ending on Wilson loops [16] at the boundary of AdS_5 [4, 15, 17].

However, to determine the fundamental closed string spectrum in $AdS_5 \times S^5$, one must learn how to quantize the $AdS_5 \times S^5$ string action in the "short" string sector, i.e., without explicitly expanding near a nontrivial bosonic string configuration.

It is well known that for the flat-space GS action, this is achieved by choosing the light-cone gauge [3, 18]. The superstring light-cone gauge fixing consists of two steps:

(I) the fermionic light-cone gauge choice, i.e., fixing the κ-symmetry by $\Gamma^+\theta^I = 0$;

(II) the bosonic light-cone gauge choice, i.e., using the conformal gauge[1] $\sqrt{g}\,g^{\mu\nu} = \eta^{\mu\nu}$ (with the Minkowski metric $\eta^{\mu\nu}$) and fixing the residual conformal diffeomorphism symmetry by $x^+(\tau, \sigma) = p^+\tau$, where τ is the world-sheet time and p^+ is the light-cone momentum.

Fixing the fermionic light-cone gauge already produces a substantial simplification of the flat-space GS action: it becomes quadratic in θ. Choosing the bosonic light-cone gauge, i.e., using an explicit choice of x^+, may not always be necessary (cf. [19, 20]), but it makes derivation of the physical string spectrum straightforward.

In what follows, we concentrate on the first and crucial step of fixing the fermionic light-cone gauge, i.e., imposing an analog of the $\Gamma^+\theta^I = 0$ condition. The idea is to obtain a simple gauge-fixed form of the action where the nondegeneracy of the kinetic term for the fermions does not depend on the choice of a specific string background in transverse directions; i.e., similarly to the flat-space, the fermion kinetic term has the structure $\partial x^+ \bar{\theta}\,\partial\theta$.

There are other motivations for studying $AdS_5 \times S^5$ strings in the light-cone gauge. One of the primary goals is to clarify the relation between the string theory and the $\mathcal{N} = 4$ super Yang–Mills (SYM) theory at the boundary. The SYM theory does not show a $\mathcal{N} = 4$ supersymmetric Lorentz-covariant description but has a simple superspace description in the light-cone gauge $A^+ = 0$ [22]. It is based on a single chiral superfield

$$\Phi(x, \theta) = A(x) + \theta^i\psi_i(x) + \ldots,$$

where $A = A_1 + iA_2$ represents the transverse components of the gauge field and ψ_i is its fermionic partner transforming under the fundamental representation of the R-symmetry group SU(4). In addition to the standard light-cone supersymmetry (shifts of θ), the light-cone superspace SYM action $S[\Phi]$ also has a nonlinear superconformal symmetry. This suggests that it may be possible to formulate the bulk string theory in a way that is naturally related to the light-cone form of the boundary SYM theory. In particular, it may be useful to split the corresponding fermionic string coordinates into two parts possessing the manifest $SU(4) \simeq SO(6)$ transformation properties, which can be the counterparts of the linearly realized Poincaré supersymmetry supercharges, and the nonlinearly realized conformal supersymmetry supercharges of the SYM theory. Also, as was shown in [23–25], field theories in the AdS space (in particular, the IIB supergravity) have a simple light-cone description. There exists a light-cone action for a superparticle in $AdS_5 \times S^5$ that was used to formulate the AdS/CFT correspondence in the light-cone gauge. This suggests that the full superstring theory in $AdS_5 \times S^5$ also must have a natural light-cone gauge formulation, which should be useful in the context of the AdS/CFT correspondence.

The paper is organized as follows. In Section 2, we discuss the basic superalgebra psu(2, 2|4) and write its (anti)commutation relations in the light-cone basis corresponding to the light-cone decomposition of the so(4, 2) generators. We discuss the relations between the so(4, 1) $\oplus$ so(5) (or 5 + 5) basis[2] of the psu(2, 2|4) superalgebra used in [6] in the construction of the GS action in $AdS_5 \times S^5$ and the more familiar so(3, 1) $\oplus$ su(4) $\approx$ sl(2, *C*) $\oplus$ su(4) (or "4 + 6") basis (naturally appearing in the discussion of the $\mathcal{N} = 4$, $d = 4$ superconformal symmetry of the SYM theory). We use the latter basis to identify the generators of the algebra in the light-cone (or so(1, 1) $\oplus$ u(1) $\oplus$ so(2) $\oplus$ su(4)) basis. The knowledge of the explicit relations between the gener-

[1] We use Minkowski signature two-dimensional world-sheet metric $g^{\mu\nu}$ with $g \equiv -\det g_{\mu\nu}$.

[2] We label the basis by the symmetry algebras under which the supercoordinates transform linearly.

ators in the three bases is useful in finding normalizations in the forms of the string action corresponding to the so(3, 1) ⊕ su(4) and the light-cone bases.

In Section 3, we present a version of the $AdS_5 \times S^5$ superstring action using the "*S*-gauge" to fix the κ-symmetry (*S* refers to the conformal supersymmetry generator). In this gauge, all of the superconformal η-fermions are gauged away.

In Section 4, we explain the transformation of the $AdS_5 \times S^5$ string action from its original form in the so(4, 1) ⊕ so(5) basis [6] to the so(3, 1) ⊕ su(4) basis and then to the light-cone basis. We adapt the original $AdS_5 \times S^5$ GS action of [6] to the case of the light-cone basis of psu(2, 2|4). The resulting κ-symmetric action is written entirely in terms of the Cartan-1 forms corresponding to the light-cone basis and in an arbitrary (e.g., Wess–Zumino (WZ) or Killing) parametrization of the supercoset space.

In Section 5, we fix the light-cone κ-symmetry gauge and find the corresponding Cartan-1 forms. These light-cone gauge-1 forms are given in the Killing parametrization of the original superspace.

In Section 6, we find the fermionic light-cone gauge-fixed form of the action of Section 4. We present the action in the Killing parametrization and discuss some of its properties. We then explain the transformation of this action into the Wess–Zumino parametrization. Our fermionic κ-symmetry light-cone gauge (which is different from the naive $\Gamma^+\theta^I = 0$ gauge but is related to it in the flat-space limit) reduces the 32 fermionic coordinates θ^I_α (two left Majorana–Weyl ten-dimensional spinors) to 16 physical Grassmann variables: "linear" θ^i and "nonlinear" η^i and their hermitian conjugates θ_i and η_i (i = 1, 2, 3, 4), which transform according to the fundamental representations of SU(4). The superconformal algebra psu(2, 2|4) dictates that these variables must be related to the Poincaré and the conformal supersymmetry in the light-cone gauge description of the boundary theory. The action and symmetry generators have a simple (quadratic) dependence on θ^i but a complicated (quartic) dependence on η^i. We split the 10 bosonic coordinates of $AdS_5 \times S^5$ into 4 isometric coordinates along the boundary directions, the radial direction of AdS_5, and the S_5 coordinates.

In Section 7, we present the light-cone gauge-fixed action in different forms and discuss some of its properties.

2. THE psu(2, 2|4) SUPERALGEBRA

We start with the commutation relations of the psu(2, 2|4) superalgebra in the so(4, 1) ⊕ so(5) basis given in [6]:

$$[\hat{P}_A, \hat{P}_B] = \hat{J}_{AB}, \quad [P_{A'}, P_{B'}] = -J_{A'B'}, \tag{2.1}$$

$$\begin{aligned}[\hat{J}^{AB}, \hat{J}^{CE}] &= \eta^{BC}\hat{J}^{AE} + 3 \text{ terms},\\ [J^{A'B'}, J^{C'E'}] &= \eta^{B'C'}J^{A'E'} + 3 \text{ terms},\end{aligned} \tag{2.2}$$

$$\begin{aligned}[Q_I, \hat{P}_A] &= -\frac{i}{2}\epsilon_{IJ}Q_J\gamma_A,\\ [Q_I, \hat{J}_{AB}] &= -\frac{1}{2}Q_I\gamma_{AB},\end{aligned} \tag{2.3}$$

$$\begin{aligned}[Q_I, P_{A'}] &= \frac{1}{2}\epsilon_{IJ}Q_J\gamma_{A'},\\ [Q_I, J_{A'B'}] &= -\frac{1}{2}Q_I\gamma_{A'B'},\end{aligned} \tag{2.4}$$

$$\begin{aligned}&\{Q_{\alpha iI}, Q_{\beta jJ}\}\\ &= \delta_{IJ}[-2iC_{ij}(C\gamma^A)_{\alpha\beta}\hat{P}_A + 2C_{\alpha\beta}(C'\gamma^{A'})_{ij}P_{A'}]\\ &+ \epsilon_{IJ}[C_{ij}(C\gamma^{AB})_{\alpha\beta}\hat{J}_{AB} - C_{\alpha\beta}(C'\gamma^{A'B'})_{ij}J_{A'B'}],\end{aligned}$$

where

$$Q_{\alpha iI} \equiv Q^{J\beta j}\delta_{JI}C_{\beta\alpha}C_{ji}$$

and our conventions for the charge conjugation matrix C are given in (6.5). Hermitean conjugation rules in this basis are

$$\begin{aligned}\hat{P}^\dagger_A &= -\hat{P}_A, \quad P^\dagger_{A'} = -P_{A'},\\ \hat{J}^\dagger_{AB} &= -\hat{J}_{AB}, \quad J^\dagger_{A'B'} = -J_{A'B'},\end{aligned} \tag{2.5}$$

$$(Q^{I\beta i})^\dagger(\gamma^0)^\beta_\alpha = -Q^{I\beta j}C_{\beta\alpha}C_{ji}. \tag{2.6}$$

Unless otherwise specified, we use the notation Q^I for $Q^{I\alpha i}$ and Q_I for $Q_{I\alpha i}$.

We use the form of the basis of the so(4, 2) subalgebra implied by its interpretation as the conformal algebra in 4 dimensions. The generators are then called the translations P^a, the conformal boosts K^a, the dilatation D, and the Lorentz rotations J^{ab} and satisfy the standard commutation relations

$$\begin{aligned}[P^a, J^{bc}] &= \eta^{ab}P^c - \eta^{ac}P^b,\\ [K^a, J^{bc}] &= \eta^{ab}K^c - \eta^{ac}K^b,\\ [P^a, K^b] &= \eta^{ab}D - J^{ab},\end{aligned} \tag{2.7}$$

$$\begin{aligned}[D, P^a] &= -P^a, \quad [D, K^a] = K^a,\\ [J^{ab}, J^{cd}] &= \eta^{bc}J^{ad} + 3 \text{ terms},\end{aligned} \tag{2.8}$$

where η^{ab} = (–, +, +, +) and a, b, c, d = 0, 1, 2, 3. We first transform the bosonic generators into the conformal algebra basis. To this end, we introduce the

Poincaré translations P^a, the conformal boosts K^a, and the dilatation D by

$$P^a = \hat{P}^a + \hat{J}^{4a}, \quad K^a = \frac{1}{2}(-\hat{P}^a + \hat{J}^{4a}), \qquad D = -\hat{P}^4. \tag{2.9}$$

Using commutation relations (2.1) and (2.2), we can easily verify that these generators satisfy the commutation relations given in (2.7) and (2.8).

Next, we introduce the new "charged" supergenerators

$$Q^q \equiv \frac{1}{\sqrt{2}}(Q^1 + iQ^2), \quad Q^{\bar{q}} \equiv \frac{1}{\sqrt{2}}(Q^1 - iQ^2). \tag{2.10}$$

We use the simplified notation

$$Q^{\alpha i} \equiv -Q^{q\alpha i}, \quad Q_{\alpha i} \equiv Q_{q\alpha i}. \tag{2.11}$$

The nonvanishing values of δ_{IJ} are then replaced by $\delta_{q\bar{q}} = 1$, and the Majorana condition becomes

$$(Q^{\beta i})^\dagger (\gamma^0)_\alpha^\beta = Q_{\alpha i}.$$

The commutators have the form

$$[Q^{\alpha i}, \hat{P}^A] = -\frac{1}{2}(\gamma^A Q)^{\alpha i}, \tag{2.12}$$

$$[Q^{\alpha i}, \hat{J}^{AB}] = \frac{1}{2}(\gamma^{AB} Q)^{\alpha i}, \tag{2.12}$$

$$[Q_{\alpha i}, \hat{P}^A] = \frac{1}{2}(Q\gamma^A)_{\alpha i}, \tag{2.13}$$

$$[Q_{\alpha i}, \hat{J}^{AB}] = -\frac{1}{2}(Q\gamma^{AB})_{\alpha i}, \tag{2.13}$$

$$[Q^{\alpha i}, P^{A'}] = -\frac{i}{2}(\gamma^{A'} Q)^{\alpha i}, \tag{2.14}$$

$$[Q^{\alpha i}, J^{A'B'}] = \frac{1}{2}(\gamma^{A'B'} Q)^{\alpha i}, \tag{2.14}$$

$$[Q_{\alpha i}, P^{A'}] = \frac{i}{2}(Q\gamma^{A'})_{\alpha i}, \tag{2.15}$$

$$[Q_{\alpha i}, J^{A'B'}] = -\frac{1}{2}(Q\gamma^{A'B'})_{\alpha i}, \tag{2.15}$$

while the anticommutators become

$$\{Q^{\alpha i}, Q_{\beta j}\} = [2i(\gamma^A)_\beta^\alpha \hat{P}^A + (\gamma^{AB})_\beta^\alpha \hat{J}^{AB}]\delta_j^i - 4i\delta_\beta^\alpha J^i{}_j, \tag{2.16}$$

where we use the notation

$$J^i{}_j \equiv -\frac{i}{2}(\gamma^{A'})^i{}_j P^{A'} + \frac{1}{4}(\gamma^{A'B'})^i{}_j J^{A'B'}. \tag{2.17}$$

Starting with the commutation relations for $P^{A'}$ and $J^{A'B'}$ and applying various Fierz identities, we can prove that $J^i{}_j$ ($J^{i\dagger}{}_j = J^j{}_i$) satisfy the commutation relations of the su(4) algebra.

Using the commutators in Eqs. (2.14), (2.15), and (2.17) and the completeness relation for the Dirac matrices, we can prove that

$$[Q_{\alpha i}, J^j{}_k] = \delta_i^j Q_{\alpha k} - \frac{1}{4}\delta_k^j Q_{\alpha i}, \qquad [Q^{\alpha i}, J^j{}_k] = -\delta_k^i Q^{\alpha j} + \frac{1}{4}\delta_k^j Q^{\alpha i}. \tag{2.18}$$

This demonstrates that the supercharges transform in the fundamental representations of su(4).

In what follows, we decompose the so(4, 1) Dirac and charge conjugation matrices in the sl(2) basis as

$$(\gamma^a)^\alpha{}_\beta = \begin{pmatrix} 0 & (\sigma^a)^{a\dot{b}} \\ (\bar{\sigma}^a)_{\dot{a}b} & 0 \end{pmatrix}, \qquad \gamma^4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad C_{\alpha\beta} = \begin{pmatrix} \epsilon_{ab} & 0 \\ 0 & \epsilon^{\dot{a}\dot{b}} \end{pmatrix}, \tag{2.19}$$

where the matrices $(\sigma^a)^{a\dot{a}}$ and $(\bar{\sigma}^a)_{\dot{a}a}$ are related to the Pauli matrices in the standard way:

$$\sigma^a = (1, \sigma^1, \sigma^2, \sigma^3), \quad \bar{\sigma}^a = (-1, \sigma^1, \sigma^2, \sigma^3). \tag{2.20}$$

We note that

$$(\sigma^a)_{a\dot{a}} = (\bar{\sigma}^a)_{\dot{a}a}, \quad (\bar{\sigma}^a)_{\dot{a}a} = (\sigma^{a*})_{a\dot{a}},$$

where

$$(\sigma^a)_{a\dot{a}} \equiv (\sigma^a)^{b\dot{b}} \epsilon_{ba} \epsilon_{\dot{b}\dot{a}}.$$

We use the following conventions for the sl(2) indices:

$$\psi^a = \epsilon^{ab}\psi_b, \quad \psi_a = \psi^b \epsilon_{ba}, \qquad \psi^{\dot{a}} = \epsilon^{\dot{a}\dot{b}}\psi_{\dot{b}}, \quad \psi_{\dot{a}} = \psi^{\dot{b}} \epsilon_{\dot{b}\dot{a}}. \tag{2.21}$$

with

$$\epsilon_{12} = \epsilon^{12} = -\epsilon_{\dot{1}\dot{2}} = -\epsilon^{\dot{1}\dot{2}} = 1.$$

We then decompose the supercharges in the sl(2) $\oplus$ su(4) basis as

$$Q^{\alpha i} = \begin{pmatrix} 2iv^{-1}Q^{ai} \\ 2vS^i_{\dot{a}} \end{pmatrix}, \qquad Q_{\alpha i} = (2vS_{ai}, -2iv^{-1}Q_i^{\dot{a}}), \quad v \equiv 2^{1/4}. \tag{2.22}$$

In terms of these new supercharges, the commutation relations become

$$[D, Q^{ai}] = -\frac{1}{2}Q^{ai}, \quad [D, S_i^a] = \frac{1}{2}S_i^a, \tag{2.23}$$

$$[S_i^{\mathrm{a}}, P^a] = \frac{i}{\sqrt{2}}(\sigma^a)^{\mathrm{a}\dot{\mathrm{a}}} Q_{\dot{\mathrm{a}}i},$$
$$[S^i_{\dot{\mathrm{a}}}, P^a] = -\frac{i}{\sqrt{2}}(\bar{\sigma}^a)_{\dot{\mathrm{a}}\mathrm{a}} Q^{\mathrm{a}i}, \tag{2.24}$$

$$[Q^{\mathrm{a}i}, K^a] = -\frac{i}{\sqrt{2}}(\sigma^a)^{\mathrm{a}\dot{\mathrm{a}}} S^i_{\dot{\mathrm{a}}},$$
$$[Q_{\dot{\mathrm{a}}i}, K^a] = \frac{i}{\sqrt{2}}(\bar{\sigma}^a)_{\dot{\mathrm{a}}\mathrm{a}} S_i^{\mathrm{a}}, \tag{2.25}$$

$$\{Q^{\mathrm{a}i}, Q_j^{\dot{\mathrm{b}}}\} = \frac{i}{\sqrt{2}}(\sigma_a)^{\mathrm{a}\dot{\mathrm{b}}} P^a \delta^i_j,$$
$$\{S_j^{\mathrm{a}}, S^{\dot{\mathrm{b}}i}\} = -\frac{i}{\sqrt{2}}(\sigma_a)^{\mathrm{a}\dot{\mathrm{b}}} K^a \delta^i_j, \tag{2.26}$$

$$[Q^{\mathrm{a}i}, J^{ab}] = \frac{1}{2}(\sigma^{ab})^{\mathrm{a}}{}_{\mathrm{b}} Q^{\mathrm{b}i}, \tag{2.27}$$

$$\{Q^{\mathrm{a}i}, S_j^{\mathrm{b}}\} = \left(\frac{1}{2}\epsilon^{\mathrm{ab}} D + \frac{1}{4}(\sigma_{ab})^{\mathrm{ab}} J^{ab}\right)\delta^i_j + \epsilon^{\mathrm{ab}} J^i{}_j, \tag{2.28}$$

where

$$(\sigma^{ab})^{\mathrm{ab}} = \epsilon^{\mathrm{bc}}(\sigma^{ab})^{\mathrm{a}}{}_{\mathrm{c}},$$
$$(\sigma^{ab})^{\mathrm{a}}{}_{\mathrm{b}} \equiv \frac{1}{2}(\sigma^a)^{\mathrm{a}\dot{\mathrm{c}}}(\bar{\sigma}^b)_{\dot{\mathrm{c}}\mathrm{b}} - (a \longleftrightarrow b).$$

Hermitean conjugation rules of the supercharges are

$$Q^{i\mathrm{a}\dagger} = Q^{\dot{\mathrm{a}}}_i, \quad Q_{\mathrm{a}}^{i\dagger} = -Q_{\dot{\mathrm{a}}i}; \tag{2.29}$$

similarly for the S supercharges. The spinor sl(2) indices a, b are raised and lowered as in (2.21). From these commutation relations, we learn that $Q^{\mathrm{a}i}$ and $Q^{\dot{\mathrm{a}}}_i$ can be interpreted as the supercharges of the super Poincaré subalgebra and S_i^{a} and $S^{\dot{\mathrm{a}}i}$ are the conformal supercharges.

This completes the description of the so(3, 1) ⊕ su(4) basis. We are now ready to introduce the light-cone basis. In the light-cone decomposition

$$x^a = (x^+, x^-, x, \bar{x}), \quad x^\pm \equiv \frac{1}{\sqrt{2}}(x^3 \pm x^0),$$
$$x, \bar{x} = \frac{1}{\sqrt{2}}(x^1 \pm i x^2), \tag{2.30}$$

we have the following generators:

$$J^{+-},\ J^{\pm x},\ J^{\pm\bar{x}},\ J^{x\bar{x}},\ P^\pm,$$
$$P^x,\ P^{\bar{x}},\ K^\pm,\ K^x,\ K^{\bar{x}}. \tag{2.31}$$

To simplify the notation, we set

$$P \equiv P^x, \quad \bar{P} = P^{\bar{x}},$$
$$K \equiv K^x, \quad \bar{K} = K^{\bar{x}}. \tag{2.32}$$

The light-cone form of the so(4, 2) commutation relations can be obtained from (2.7) using the light-cone metric elements

$$\eta^{+-} = \eta^{-+} = 1, \quad \eta^{x\bar{x}} = \eta^{\bar{x}x} = 1.$$

In this paper, the *so*(6) algebra is interpreted as the *su*(4) one:

$$[J^i_j, J^k{}_n] = \delta^k_j J^i{}_n - \delta^i_n J^k{}_j, \quad i, j, k, n = 1, 2, 3, 4. \tag{2.33}$$

The transformation of the supercharges amounts to attaching the + and – signs that explicitly show their J^{+-} charges. The corresponding supercharges are defined by

$$Q^{1i} \equiv -Q^{-i}, \quad Q^{2i} \equiv Q^{+i},$$
$$Q^{\dot{1}}_i \equiv -Q^-_i, \quad Q^{\dot{2}}_i \equiv Q^+_i, \tag{2.34}$$

$$S^1_i \equiv S^-_i, \quad S^2_i \equiv -S^+_i,$$
$$S^{\dot{1}i} \equiv S^{-i}, \quad S^{\dot{2}i} \equiv -S^{+i}. \tag{2.35}$$

The choice of signs in these definitions is a matter of convention.

We describe the odd part of the psu(2, 2|4) superalgebra in terms of 32 supercharges $Q^{\pm i}$, $Q^\pm_i$, $S^{\pm i}$, and $S^\pm_i$. They carry the D, J^{+-}, and $J^{x\bar{x}}$ charges, as follows from the structure of the algebra. The commutation relations of the supercharges with the dilatation D,

$$[D, Q^{\pm i}] = -\frac{1}{2}Q^{\pm i}, \quad [D, Q^\pm_i] = -\frac{1}{2}Q^\pm_i,$$
$$[D, S^{\pm i}] = \frac{1}{2}S^{\pm i}, \quad [D, S^\pm_i] = \frac{1}{2}S^\pm_i, \tag{2.36}$$

allow us to interpret the Q generators as the standard supercharges of the super Poincaré subalgebra and the S generators as the conformal supercharges. The supercharges with the superscript + (–) have a positive (negative) J^{+-} charge:

$$[J^{+-}, Q^{\pm i}] = \pm\frac{1}{2}Q^{\pm i}, \quad [J^{+-}, Q^\pm_i] = \pm\frac{1}{2}Q^\pm_i,$$
$$[J^{+-}, S^{\pm i}] = \pm\frac{1}{2}S^{\pm i}, \quad [J^{+-}, S^\pm_i] = \pm\frac{1}{2}S^\pm_i.$$

The $J^{x\bar{x}}$ charges are fixed by the commutation relations

$$[J^{x\bar{x}}, Q^{\pm i}] = \pm\frac{1}{2}Q^{\pm i}, \quad [J^{x\bar{x}}, Q^\pm_i] = \mp\frac{1}{2}Q^\pm_i, \tag{2.37}$$

$$[J^{x\bar{x}}, S_i^{\pm}] = \pm\frac{1}{2}S_i^{\pm}, \quad [J^{x\bar{x}}, S^{\pm i}] = \mp\frac{1}{2}S^{\pm i}. \tag{2.38}$$

The transformation properties of the Q supercharges with respect to the su(4) subalgebra are determined by

$$[Q_i^{\pm}, J^j{}_k] = \delta_i^j Q_k^{\pm} - \frac{1}{4}\delta_k^j Q_i^{\pm},$$

$$[Q^{\pm i}, J^j{}_k] = -\delta_k^i Q^{\pm j} + \frac{1}{4}\delta_k^j Q^{\pm i},$$

with similar relations for the S supercharges. Anticommutation relations between the supercharges are

$$\{Q^{\pm i}, Q_j^{\pm}\} = \mp iP^{\pm}\delta_j^i, \quad \{Q^{+i}, Q_j^{-}\} = -iP\delta_j^i, \tag{2.39}$$

$$\{S^{\pm i}, S_j^{\pm}\} = \pm iK^{\pm}\delta_j^i, \quad \{S^{-i}, S_j^{+}\} = iK\delta_j^i, \tag{2.40}$$

$$\{Q^{+i}, S_j^{+}\} = -J^{+x}\delta_j^i, \quad \{Q^{-i}, S_j^{-}\} = -J^{-\bar{x}}\delta_j^i,$$

$$\{Q^{\pm i}, S_j^{\mp}\} = \frac{1}{2}(J^{+-} + J^{x\bar{x}} \mp D)\delta_j^i \mp J^i{}_j.$$

The remaining commutation relations between odd and even generators have the form

$$[Q^{-i}, J^{+x}] = -Q^{+i}, \quad [S^{-i}, J^{+\bar{x}}] = -S^{+i},$$

$$[Q^{+i}, J^{-\bar{x}}] = Q^{-i}, \quad [S^{+i}, J^{-x}] = S^{-i},$$

$$[S_i^{\mp}, P^{\pm}] = iQ_i^{\pm}, \quad [S_i^{-}, P] = iQ_i^{-},$$

$$[S_i^{+}, \bar{P}] = -iQ_i^{+},$$

$$[Q^{\mp i}, K^{\pm}] = -iS^{\pm i}, \quad [Q^{-i}, K] = -iS^{-i},$$

$$[Q^{+i}, \bar{K}] = iS^{+i}.$$

The generators are subject to the hermitean conjugation conditions

$$(P^{\pm})^{\dagger} = -P^{\pm}, \quad P^{\dagger} = -\bar{P},$$

$$(K^{\pm})^{\dagger} = -K^{\pm}, \quad K^{\dagger} = -\bar{K},$$

$$(J^{\pm x})^{\dagger} = -J^{\pm\bar{x}}, \quad (J^{+-})^{\dagger} = -J^{+-},$$

$$(J^{x\bar{x}})^{\dagger} = J^{x\bar{x}}, \quad D^{\dagger} = -D, \quad (J^i{}_j)^{\dagger} = J^i{}_j.$$

$$(Q^{\pm i})^{\dagger} = Q_i^{\pm}, \quad (S^{\pm i})^{\dagger} = S_i^{\pm}. \tag{2.41}$$

All the remaining nontrivial (anti)commutation relations of the psu(2, 2|4) superalgebra can be obtained using these hermitean conjugation rules and the (anti)commutation relations given above.

3. THE GENERAL STRUCTURE OF THE $AdS_5 \times S^5$ STRING ACTION

The superstring action in $AdS_5 \times S^5$ [6] has the same structure as the flat-space GS action (1.4):

$$I = \int_{\partial M_3} \mathcal{L}_{\rm kin} + i\int_{M_3} \mathcal{H}. \tag{3.1}$$

The kinetic term of the $AdS_5 \times S^5$ GS action and the 3-form in its WZ term have the following form in the so(4, 1) $\oplus$ so(5) basis [6]:

$$\mathcal{L}_{\rm kin} = -\frac{1}{2}\sqrt{g}g^{\mu\nu}(\hat{L}_\mu^A\hat{L}_\nu^A + L_\mu^{A'}L_\nu^{A'}), \tag{3.2}$$

$$\mathcal{H} = s^{IJ}\hat{L}^A\bar{L}^I\gamma^A L^J + is^{IJ}L^{A'}\bar{L}^J\gamma^{A'}L^J. \tag{3.3}$$

They are expressed in terms of the Cartan 1-forms defined in the so(4, 1) $\oplus$ so(5) basis by

$$G^{-1}dG = (G^{-1}dG)_{\rm bos} + L^{I\alpha i}Q_{I\alpha i}, \tag{3.4}$$

where the restriction to the bosonic part is

$$(G^{-1}dG)_{\rm bos} = \hat{L}^A\hat{P}^A + \frac{1}{2}\hat{L}^{AB}\hat{J}^{AB} + L^{A'}P^{A'} + \frac{1}{2}L^{A'B'}J^{A'B'}. \tag{3.5}$$

In [6], the Cartan forms used in writing the action were given in the so(4, 1) $\oplus$ so(5) basis of psu(2, 2|4). This allows one to present the $AdS_5 \times S^5$ GS action in the form that is similar to the flat-space one. Our present goal is to rewrite the action in the light-cone basis discussed in the previous section and then to impose the κ-symmetry light-cone gauge.

We first consider the so(3, 1) $\oplus$ su(4) basis and define the bosonic (even) Cartan forms by

$$(G^{-1}dG)_{\rm bos} = L_P^a P^a + L_K^a K^a + L_D D + \frac{1}{2}L^{ab}J^{ab} + L^i{}_j J^j{}_i. \tag{3.6}$$

Comparing this with (3.5) and using (2.9) and (2.17), we obtain

$$\hat{L}^a = L_P^a - \frac{1}{2}L_K^a, \quad \hat{L}^{4a} = L_P^a + \frac{1}{2}L_K^a, \quad \hat{L}^4 = -L_D, \tag{3.7}$$

$$L^i{}_j = \frac{i}{2}(\gamma^{A'})^i{}_j L^{A'} - \frac{1}{4}(\gamma^{A'B'})^i{}_j L^{A'B'}. \tag{3.8}$$

Using these relations in the expression for the kinetic term (3.2) gives the action in Eq. (3.25).

We next consider the fermionic 1-forms. We define

$$L^q \equiv \frac{1}{\sqrt{2}}(L^1 + iL^2), \quad L^{\bar{q}} \equiv \frac{1}{\sqrt{2}}(L^1 - iL^2); \tag{3.9}$$

introduce the notation $L^{\alpha i} = L^{q\alpha i}$ and $L_{\alpha i} = L_{q\alpha i}$; and use the decomposition into sl(2) ⊕ su(4) Cartan 1-forms

$$L^{\alpha i} = \frac{1}{2}\begin{pmatrix} v^{-1}L_S^{\dot{a}i} \\ iv L_{Q\dot{a}}^{i} \end{pmatrix}, \qquad L_{\alpha i} = \frac{1}{2}(-iv L_{Qai},\ v^{-1}L_{Si}^{\dot{a}}). \tag{3.10}$$

Hermitean conjugation rules for the new Cartan 1-forms then take the same form as in (2.29). The light-cone frame Cartan 1-forms are defined by

$$L_{Qi}^{1} = -L_{Qi}^{-}, \quad L_{Qi}^{2} = -L_{Qi}^{+}, \qquad L_{Q}^{\dot{1}i} = -L_{Q}^{-i}, \quad L_{Q}^{\dot{2}i} = -L_{Q}^{+i}, \tag{3.11}$$

$$L_{S}^{1i} = L_{S}^{-i}, \quad L_{S}^{2i} = L_{S}^{+i}, \qquad L_{Si}^{\dot{1}} = L_{Si}^{-}, \quad L_{Si}^{\dot{2}} = L_{Si}^{+}. \tag{3.12}$$

These relations imply

$$L^{I\alpha i}Q_{I\alpha i} = L^{\alpha i}Q_{\alpha i} - L_{\alpha i}Q^{\alpha i} \tag{3.13}$$

$$= L_{Qi}^{a}Q_{a}^{i} - L_{Q}^{\dot{a}i}Q_{\dot{a}i} + L_{S}^{\dot{a}i}S_{ai} - L_{Si}^{\dot{a}}S_{\dot{a}}^{i} \tag{3.14}$$

$$= L_{Q}^{+i}Q_{i}^{-} + L_{Q}^{-i}Q_{i}^{+} + L_{Qi}^{+}Q^{-i} + L_{Qi}^{-}Q^{+i} + L_{S}^{-i}S_{i}^{+} + L_{S}^{+i}S_{i}^{-} + L_{Si}^{-}S^{+i} + L_{Si}^{+}S^{-i}. \tag{3.15}$$

Representation (3.14) corresponds to the sl(2) ⊕ su(4) basis and (3.15) to the light-cone basis.

With the relation between the Cartan 1-forms given by Eqs. (3.9)–(3.12), we are ready to consider the decomposition of the WZ 3-form (3.3). We start with the AdS_5 contribution given by the first term in right-hand side of (3.3). Taking Eq. (3.9) into account and using $\bar{L}^I = L^I CC'$, we can rewrite the AdS_5 contribution in terms of the "charged" Cartan forms L^q and $L^{\bar{q}}$ as

$$\mathcal{H}_{AdS_5} = \mathcal{H}^{q}_{AdS_5} + \mathcal{H}^{\bar{q}}_{AdS_5}, \tag{3.16}$$

$$\mathcal{H}^{q}_{AdS_5} \equiv \hat{L}^{A}L^{q\alpha i}(C\gamma^{A})_{\alpha\beta}C'_{ij}L^{q\beta j}, \qquad \mathcal{H}^{\bar{q}}_{AdS_5} \equiv \hat{L}^{A}L^{\bar{q}\alpha i}(C\gamma^{A})_{\alpha\beta}C'_{ij}L^{\bar{q}\beta j}. \tag{3.17}$$

Because $i\mathcal{H}^{\bar{q}}_{AdS_5}$ is hermitean conjugate to $i\mathcal{H}^{q}_{AdS_5}$, we restrict our attention to the decomposition of the first term. We obtain

$$\mathcal{H}^{q}_{AdS_5} = \hat{L}^{a}L^{q\alpha i}(C\gamma^{a})_{\alpha\beta}C'_{ij}L^{q\beta j}$$

$$-L_D L^{q\alpha i}(C\gamma^{4})_{\alpha\beta}C'_{ij}L^{q\beta j} = \frac{i}{2}\hat{L}^{a}L_{Sa}^{i}C'_{ij}(\sigma^{a})^{\dot{a}\dot{b}}L_{Q\dot{b}}^{j} \tag{3.18}$$

$$+\frac{1}{4}L_D\left(\frac{1}{\sqrt{2}}L_S^{\dot{a}i}C'_{ij}L_{Sa}^{j} + \sqrt{2}L_Q^{\dot{a}i}C'_{ij}L_{Q\dot{a}}^{j}\right)$$

$$= -\frac{i}{\sqrt{2}}(\hat{L}^{+}L_S^{-i}C'_{ij}L_Q^{-j} + \hat{L}^{-}L_Q^{+i}C'_{ij}L_S^{+j} + \hat{L}^{x}L_S^{-i}C'_{ij}L_Q^{+j} + \hat{L}^{\bar{x}}L_S^{+i}C'_{ij}L_Q^{-j}) \tag{3.19}$$

$$+\frac{1}{\sqrt{2}}L_D\left(\frac{1}{2}L_S^{+i}C'_{ij}L_S^{-j} + L_Q^{-i}C'_{ij}L_Q^{+j}\right).$$

Equation (3.19) provides the representation of the AdS_5 part of the 3-form in the sl(2) ⊕ su(4) basis and Eq. (3.19) in the light-cone basis.

We now consider the S^5 part of the WZ 3-form in (3.3), i.e., $is^{IJ}L^{A'}\bar{L}^{I}\gamma^{A'}L^{J}$. Representing it in terms of the charged Cartan forms as in (3.16),

$$\mathcal{H}_{S^5} = \mathcal{H}^{q}_{S^5} + \mathcal{H}^{\bar{q}}_{S^5},$$

we obtain

$$\mathcal{H}^{q}_{S^5} = iL^{A'}L^{q\alpha i}C_{\alpha\beta}(C'\gamma^{A'})_{ij}L^{q\beta j} = -2L^{q\alpha i}C_{\alpha\beta}C'_{ik}L^{k}{}_{j}L^{q\beta j} \tag{3.20}$$

$$= \frac{1}{2\sqrt{2}}L_S^{\dot{a}i}(C'L)_{ij}L_{Sa}^{j} - \frac{1}{\sqrt{2}}L_Q^{\dot{a}i}(C'L)_{ij}L_{Q\dot{a}}^{j} \tag{3.21}$$

$$= \frac{1}{2\sqrt{2}}[L_S^{+i}(C'L)_{ij}L_S^{-j} - L_S^{-i}(C'L)_{ij}L_S^{+j}] + \frac{1}{\sqrt{2}}[L_Q^{+i}(C'L)_{ij}L_Q^{-j} - L_Q^{-i}(C'L)_{ij}L_Q^{+j}]. \tag{3.22}$$

In (3.20), we exploited relation (3.8) and used the fact that $(C'\gamma^{A'B'})_{ij}$ is symmetric in i, j; the charge conjugation matrix $C_{\alpha\beta}$ is antisymmetric in α, β; and the fermionic Cartan 1-forms L^q commute with each other. Equation (3.21) provides the representation of the S^5 part of the WZ 3-form in the sl(2) ⊕ su(4) basis, and Eq. (3.22) in the light-cone basis.

To summarize, the Cartan 1-forms in the light-cone basis are defined by

$$G^{-1}dG = L_P^{a}P^{a} + L_K^{a}K^{a} + L_D D + \frac{1}{2}L^{ab}J^{ab} + L^{i}{}_{j}J^{j}{}_{i} + L_Q^{-i}Q_i^{+} + L_{Qi}^{-}Q^{+i} + L_Q^{+i}Q_i^{-} + L_{Qi}^{+}Q^{-i} + L_S^{-i}S_i^{+} + L_{Si}^{-}S^{+i} + L_S^{+i}S_i^{-} + L_{Si}^{+}S^{-i}, \tag{3.23}$$

where G is a coset representative in PSU(2, 2|4). We also define the combinations

$$\hat{L}^{\alpha} \equiv L_P^a - \frac{1}{2}L_K^a, \quad L^{A'} \equiv -\frac{i}{2}(\gamma^{A'})^i{}_j L^j{}_i, \qquad (C'L)_{ij} \equiv C'_{ik}L^k{}_j. \tag{3.24}$$

The kinetic term in (3.1) then takes the form

$$\mathcal{L}_{\rm kin} = -\frac{1}{2}\sqrt{g} \times g^{\mu\nu}(\hat{L}^a_\mu \hat{L}^a_\nu + L_{D\mu}L_{D\nu} + L^{A'}_\mu L^{A'}_\nu), \tag{3.25}$$

and the 3-form $\mathcal{H}$ in the WZ term can be written as (we suppress the signs of exterior products of 1-forms)

$$\mathcal{H} = \mathcal{H}^q_{AdS_5} + \mathcal{H}^q_{S^5} - \text{H.c.}, \tag{3.26}$$

$$\mathcal{H}^q_{AdS_5} = -\frac{i}{\sqrt{2}}(\hat{L}^+ L_S^{-i} C'_{ij} L_Q^{-j} + \hat{L}^- L_Q^{+i} C'_{ij} L_S^{+j} + \hat{L}^x L_S^{-i} C'_{ij} L_Q^{+j} + \hat{L}^{\bar{x}} L_S^{+i} C'_{ij} L_Q^{-j}) + \frac{1}{\sqrt{2}} L_D \left(\frac{1}{2} L_S^{+i} C'_{ij} L_S^{-j} + L_Q^{-i} C'_{ij} L_Q^{+j}\right), \tag{3.27}$$

$$\mathcal{H}^q_{S^5} = \frac{1}{2\sqrt{2}}[L_S^{+i}(C'L)_{ij}L_S^{-j} - L_S^{-i}(C'L)_{ij}L_S^{+j}] + \frac{1}{\sqrt{2}}[L_Q^{+i}(C'L)_{ij}L_Q^{-j} - L_Q^{-i}(C'L)_{ij}L_Q^{+j}]. \tag{3.28}$$

4. THE $AdS_5 \times S^5$ ACTION IN THE COVARIANT S-GAUGE

The results for the Cartan forms in the sl(2) ⊕ su(4) basis described above allow us to find another version of the κ-symmetry gauge-fixed action of the superstring in $AdS_5 \times S^5$. We start with the supercoset representative

$$G = g_{x,\theta} g_\eta g_y g_\phi, \tag{4.1}$$

$$g_{x,\theta} = \exp(x^a P^a + \theta^{\dot{a}}_i Q^i_{\dot{a}} - \theta^{\dot{a}i} Q_{\dot{a}i}), \tag{4.2}$$

$$g_\eta = \exp(\eta^{\dot{a}i} S_{\dot{a}i} - \eta^{\dot{a}}_i S^i_{\dot{a}}), \tag{4.3}$$

and impose the κ-symmetry gauge by

$$\eta^{\dot{a}i} = \eta^{\dot{a}}_i = 0, \tag{4.4}$$

i.e.,

$$G_{g.f.} = g_{x,\theta} g_y g_\phi. \tag{4.5}$$

Because we have set to zero the fermionic coordinates η that correspond to the conformal supercharges S, we call this the S gauge. The resulting gauge-fixed expressions for the Cartan 1-forms are given by

$$L^a_P = e^\phi \times \left[dx^a - \frac{i}{2\sqrt{2}}(\theta_{i\dot{a}}(\sigma^a)^{\dot{a}\dot{b}} d\theta^i_{\dot{b}} + \theta^{i\dot{a}}(\bar{\sigma}^a)_{\dot{a}\dot{b}} d\theta^{\dot{b}}_i)\right], \tag{4.6}$$

$$L^{\dot{a}}_{Qi} = e^{\phi/2}\tilde{d}\theta^{\dot{a}}_i, \quad L^{\dot{a}i}_Q = e^{\phi/2}\tilde{d}\theta^{\dot{a}i}, \tag{4.7}$$

$$L_D = d\phi, \quad L^i{}_j = (dUU^{-1})^i{}_j, \tag{4.8}$$

where $\tilde{d}\theta$ is defined as in (5.20) and the matrix U is defined by (7.3) and (7.4). All the remaining Cartan 1-forms vanish.

Using $L_S = 0$, we express the AdS_5 part of the 3-form $\mathcal{H}$ in (3.19) as

$$\mathcal{H}^q_{AdS_5} = \frac{1}{2\sqrt{2}} d\phi e^\phi \tilde{d}\theta^{\dot{a}i} C'_{ij} \tilde{d}\theta^j_{\dot{a}}. \tag{4.9}$$

At the same time, Eq. (3.21) gives

$$\mathcal{H}^q_{S^5} = -\frac{1}{\sqrt{2}} e^\phi \tilde{d}\theta^{\dot{a}i}(C'L)_{ij}\tilde{d}\theta^j_{\dot{a}}. \tag{4.10}$$

Thus, we conclude that

$$\mathcal{H}^q_{AdS_5} + \mathcal{H}^q_{S^5} = d\left(\frac{1}{2\sqrt{2}} e^\phi \tilde{d}\theta^{\dot{a}i} C'_{ij} \tilde{d}\theta^j_{\dot{a}}\right), \tag{4.11}$$

which allows us to find the two-dimensional form of the WZ term.

Using the above relations and Eqs. (3.25) and (3.24) and taking into account that $L^a_K = 0$, we finally obtain the kinetic and WZ parts of the $AdS_5 \times S^5$ string Lagrangian (cf. (6.13), (6.14))

$$\mathcal{L}_{\rm kin} = -\frac{1}{2}\sqrt{g} \times g^{\mu\nu}(L^a_{P\mu} L^a_{P\nu} + \partial_\mu\phi\partial_\nu\phi + e^{A'}_\mu e^{A'}_\nu), \tag{4.12}$$

$$\mathcal{L}_{\rm WZ} = \frac{i}{2\sqrt{2}}\epsilon^{\mu\nu} e^\phi \tilde{\partial}_\mu \theta^{\dot{a}i} C'_{ij} \tilde{\partial}_\nu \theta^j_{\dot{a}} + \text{H.c.}, \tag{4.13}$$

where $L^a_{P\mu}$ is given by (4.6). We note that in this S-gauge, the 1-form $L^{A'}$, which is given in terms of $L^i{}_j$ as in (3.24), is equal simply to the S^5 1-form $e^{A'}$. The reason is that in contrast to what happens in the light-cone gauge (5.15), the Cartan form $L^i{}_j$ does not contain fermionic contributions here [see (4.8)].

5. THE COORDINATE PARAMETRIZATION OF CARTAN FORMS AND FIXING THE LIGHT-CONE κ-SYMMETRY GAUGE

To express the Cartan 1-forms in terms of the even and odd coordinate fields, we start with the supercoset representative

$$G = g_{x,\theta} g_\eta g_y g_\phi, \tag{5.1}$$

where

$$\begin{aligned} g_{x,\theta} &= \exp(x^a P^a + \theta^{-i} Q_i^+ \\ &+ \theta_i^- Q^{+i} + \theta^{+i} Q_i^- + \theta_i^+ Q^{-i}), \end{aligned} \tag{5.2}$$

$$g_\eta = \exp(\eta^{-i} S_i^+ + \eta_i^- S^{+i} + \eta^{+i} S_i^- + \eta_i^+ S^{-i}); \tag{5.3}$$

and g_ϕ and g_y depend on the radial AdS_5 coordinate ϕ and the S^5 coordinates $y^{A'}$

$$g_\phi \equiv \exp(\phi D), \tag{5.4}$$

$$g_y \equiv \exp(y^i{}_j J^j{}_i), \quad y^i{}_j \equiv \frac{i}{2}(\gamma^{A'})^i{}_j y^{A'}. \tag{5.5}$$

The choice of the parametrization of the coset representative in (5.1) corresponds to what is usually referred to as the "Killing gauge" in superspace.

Because the supercharges transform in the fundamental representation of su(4), the corresponding fermionic coordinates θ and η also transform in the fundamental representation of su(4).

The above expressions provide the definition of the Cartan forms in the light-cone basis. We further specify these expressions by setting some of the fermionic coordinates to zero, which corresponds to fixing a particular κ-symmetry gauge. More specifically, we fix the κ-symmetry by setting to zero all the Grassmann coordinates that carry a positive J^{+-} charge [cf. (6.9)],

$$\theta^{+i} = \theta_i^+ = \eta^{+i} = \eta_i^+ = 0. \tag{5.6}$$

To simplify the notation, we set

$$\theta^i \equiv \theta^{-i}, \quad \theta_i \equiv \theta_i^-, \quad \eta^i \equiv \eta^{-i}, \quad \eta_i \equiv \eta_i^-. \tag{5.7}$$

As the result, the κ-symmetry fixed form of coset representative (5.1) is

$$G_{g.f.} = (g_{x,\theta})_{g.f.} (g_\eta)_{g.f.} g_y g_\phi, \tag{5.8}$$

$$(g_{x,\theta})_{g.f.} = \exp(x^a P^a + \theta^i Q_i^+ + \theta_i Q^{+i}), \tag{5.9}$$

$$(g_\eta)_{g.f.} = \exp(\eta^i S_i^+ + \eta_i S^{+i}). \tag{5.10}$$

Inserting this $G_{g.f.}$ in (3.23), we obtain the κ-symmetry gauge-fixed expressions for the Cartan 1-forms

$$\begin{aligned} L_P^+ &= e^\phi dx^+, \\ L_P^- &= e^\phi \Big(dx^- - \frac{i}{2}\tilde\theta^i \tilde d\theta_i - \frac{i}{2}\tilde\theta_i d\tilde\theta^i \Big), \end{aligned} \tag{5.11}$$

$$L_P^x = e^\phi dx, \quad L_P^{\bar x} = e^\phi d\bar x, \tag{5.12}$$

$$L_K^- = e^{-\phi}\Big[\frac{1}{4}(\tilde\eta^2)^2 dx^+ + \frac{i}{2}\tilde\eta^i \tilde d\eta_i + \frac{i}{2}\tilde\eta_i \tilde d\eta^i \Big], \tag{5.13}$$

$$L_D = d\phi, \tag{5.14}$$

$$L^i{}_j = (dUU^{-1})^i{}_j + i\Big(\tilde\eta^i \tilde\eta_j - \frac{1}{4}\tilde\eta^2 \delta_j^i \Big) dx^+, \tag{5.15}$$

$$\begin{aligned} L_Q^{-i} &= e^{\phi/2}(\tilde d\theta^i + i\tilde\eta^i dx), \\ L_{Qi}^- &= e^{\phi/2}(\tilde d\theta_i - i\tilde\eta_i d\bar x), \end{aligned} \tag{5.16}$$

$$L_Q^{+i} = -ie^{\phi/2}\tilde\eta^i dx^+, \quad L_{Qi}^+ = ie^{\phi/2}\tilde\eta_i dx^+, \tag{5.17}$$

$$\begin{aligned} L_S^{-i} &= e^{-\phi/2}\Big(\tilde d\eta^i + \frac{i}{2}\tilde\eta^2 \tilde\eta^i dx^+ \Big), \\ L_{Si}^- &= e^{-\phi/2}\Big(\tilde d\eta_i - \frac{i}{2}\tilde\eta^2 \tilde\eta_i dx^+ \Big). \end{aligned} \tag{5.18}$$

All the remaining forms are equal to zero. We have introduced the notation

$$\tilde\theta^i \equiv U^i{}_j \theta^j, \quad \tilde\theta_i \equiv \theta_j (U^{-1})^j{}_i, \tag{5.19}$$

$$\tilde d\theta^i \equiv U^i{}_j d\theta^j, \quad \tilde d\theta_i \equiv d\theta_j (U^{-1})^j{}_i, \tag{5.20}$$

and similarly for η. We note that $\tilde\theta^2 = \theta^2$ and $\tilde\theta\tilde d\theta = \theta d\theta$. The matrix $U \in \mathrm{SU}(4)$ is defined in terms of the S^5 coordinates $y^i{}_j$ or $y^{A'}$ by (7.3) and (7.4). It can be written explicitly as

$$\begin{aligned} U &= \cos\frac{|y|}{2} + i\gamma^{A'} n^{A'} \sin\frac{|y|}{2}, \\ |y| &\equiv \sqrt{y^{A'} y^{A'}}, \quad n^{A'} \equiv \frac{y^{A'}}{|y|}. \end{aligned} \tag{5.21}$$

6. THE $AdS_5 \times S^5$ STRING ACTION IN THE LIGHT-CONE GAUGE

We first review the fixing of the light-cone fermionic gauge in flat-space. One usually imposes the κ-symmetry light-cone gauge by starting with the component form of the action given by (1.5). It turns out to be cumbersome to generalize this procedure to the case of strings in $AdS_5 \times S^5$. It is more convenient to first impose the light-cone gauge at the level of the Cartan forms $L^{\hat A}$ and L^I and then use them in the action taken in its general form (1.4). The light-cone-gauge form of $L^{\hat A}$ is

$$\begin{aligned} L^+ &= dx^+, \quad L^- = dx^- - i\bar\theta^I \Gamma^- d\theta^I, \\ L^N &= dx^N, \quad N = 1, \ldots, 8, \end{aligned} \tag{6.1}$$

where θ^I are subject to the light-cone gauge condition $\Gamma^+\theta^I = 0$.[3] Inserting these expressions in action (1.4) we obtain

$$\mathcal{L} = \mathcal{L}_{\rm kin} + \mathcal{L}_{WZ}, \tag{6.2}$$

$$\mathcal{L}_{\rm kin} = \sqrt{g}g^{\mu\nu}\Big(-\partial_\mu x^+\partial_\nu x^- - \frac{1}{2}\partial_\mu x^N\partial_\nu x^N + i\partial_\mu x^+\bar\theta^I\Gamma^-\partial_\nu\theta^I\Big), \tag{6.3}$$

$$\mathcal{L}_{\rm WZ} = -i\epsilon^{\mu\nu}s^{IJ}\partial_\mu x^+\bar\theta^I\Gamma^-\partial_\nu\theta^J. \tag{6.4}$$

Next, we do the "5 + 5" splitting of the ten-dimensional Clifford algebra generators, the charge conjugation matrix $\mathcal{C}$, and the supercoordinates:

$$\Gamma^A = \gamma^A\times I\times\sigma_1,\quad \Gamma^{A'} = I\times\gamma^{A'}\times\sigma_2,$$
$$\mathcal{C} = C\times C'\times i\sigma_2,\quad \theta^I = \begin{pmatrix}\theta^{I\alpha i}\\ 0\end{pmatrix}, \tag{6.5}$$

where I is the 4×4 unit matrix; σ_n are the Pauli matrices; $\alpha = 1, 2, 3, 4$; and $i = 1, 2, 3, 4$. We also introduce the complex coordinates

$$\theta^q \equiv \frac{1}{\sqrt{2}}(\theta^1 + i\theta^2) \tag{6.6}$$

and use the parametrization

$$\theta^{q\alpha i} = \frac{v}{2}\begin{pmatrix}\eta^{-i}\\ \eta^{+i}\\ -i\theta^{+i}\\ i\theta^{-i}\end{pmatrix}. \tag{6.7}$$

The decompositions of the so(4, 1) gamma matrices are given in (2.19). The light-cone gauge

$$\Gamma^+\theta^I = 0,\quad \Gamma^+ \equiv \frac{1}{\sqrt{2}}(\Gamma^3 \pm \Gamma^0) \tag{6.8}$$

leads to

$$\theta^{+i} = \eta^{+i} = 0. \tag{6.9}$$

Using the notation

$$\theta^i \equiv \theta^{-i},\quad \eta^i \equiv \eta^{-i},\quad \theta_i = (\theta^i)^\dagger,\quad \eta_i = (\eta^i)^\dagger \tag{6.10}$$

and inserting the above decomposition into action (6.2), we finally obtain the expressions for the kinetic and WZ parts of the light-cone-gauge flat-space GS Lagrangian

$$\mathcal{L}_{\rm kin} = \sqrt{g}g^{\mu\nu}\Big[-\partial_\mu x^+\partial_\nu x^- - \frac{1}{2}\partial_\mu x^N\partial_\nu x^N + \partial_\mu x^+\Big(\frac{i}{2}\theta_i\partial_\nu\theta^i + \frac{i}{2}\eta_i\partial_\nu\eta^i + \text{H.c.}\Big)\Big], \tag{6.11}$$

$$\mathcal{L}_{\rm WZ} = -\epsilon^{\mu\nu}\partial_\mu x^+\eta^i C'_{ij}\partial_\nu\theta^j + \text{H.c.} \tag{6.12}$$

It is to this form of the flat GS action that our light-cone $AdS_5\times S^5$ action reduces in the flat-space limit. A characteristic feature of this parametrization is that, while the kinetic term is diagonal in θ and η, these variables are mixed in the WZ term.

We now turn to the $AdS_5\times S^5$ case. Inserting the above expressions (5.11)–(5.18) in action (3.1), we obtain the light-cone gauge-fixed superstring Lagrangian in terms of the light-cone supercoset coordinates:

$$\mathcal{L}_{\rm kin} = \sqrt{g}g^{\mu\nu}\Big(-e^{2\phi}(\partial_\mu x^+\partial_\nu x^- + \partial_\mu x\partial_\nu\bar x) - \frac{1}{2}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}e_\mu^{A'}e_\nu^{A'} + \partial_\mu x^+\Big[\frac{i}{2}e^{2\phi}(\theta^i\partial_\nu\theta_i + \theta_i\partial_\nu\theta^i) + \frac{i}{4}(\eta^i\partial_\nu\eta_i + \eta_i\partial_\nu\eta^i) + \frac{1}{2}\tilde\eta_i e^i_{\nu j}\tilde\eta^j\Big] + \frac{1}{8}\partial_\mu x^+\partial_\nu x^+[(\eta^2)^2 - (\tilde\eta_i(\gamma^{A'})^i{}_j\tilde\eta^j)^2]\Big), \tag{6.13}$$

$$\mathcal{L}_{\rm WZ} = -\frac{\epsilon^{\mu\nu}}{\sqrt{2}}e^\phi\partial_\mu x^+\tilde\eta^i C'_{ij}(\tilde\partial_\nu\theta^j + i\tilde\eta^j\partial_\nu x) + \text{H.c.} \tag{6.14}$$

Several remarks are in order.

(i) In the flat-space limit, this action reduces to the GS light-cone κ-symmetry gauge-fixed action represented in the form (6.11) and (6.12). In the particle theory limit as $\alpha' \to 0$ (corresponding to keeping only the τ dependence of fields and omitting the WZ term), this action reduces (after an appropriate bosonic light-cone gauge fixing and rescaling some of the fermionic variables) to the light-cone action of a superparticle propagating in $AdS_5\times S^5$ [24].[4]

(ii) The kinetic terms for the fermionic coordinates have a manifest linear su(4) invariance. In the remaining terms, this symmetry is not manifested and is not realized linearly.

(iii) Because the WZ term depends on θ through its derivative, it is invariant under a shift of θ. To maintain this invariance in the kinetic terms, the shift of θ must be supplemented, as usual, by an appropriate transformation of x^-. The action is invariant under the shifts of

[3] The transverse bosonic Cartan forms L^N in (6.1) must not be confused with the fermionic ones L^I.

[4] The Hamiltonian for the superparticle in $AdS_5\times S^5$ was found in [24] (see Eq. (12) there). The action is obtained from the Hamiltonian in the usual way.

the bosonic coordinates x^a along the boundary directions.

(iv) As in the superparticle case [24, 25], this action is quadratic in half of the fermionic coordinates (θ) but is of a higher (quartic) order in the other half (η). It was the intention to split the fermionic variables in such θ and η that motivated our choice of the supercoset parametrization in (5.8).

(v) The action contains the terms of the type $(\eta^2)^2$ and $\eta_i e^i_j \eta^j$ that played an important role [24] in establishing the AdS CFT correspondence in the superparticle case. These terms must also play a similar important role in formulating the AdS/CFT correspondence at the string theory level.

The fermionic variables θ and η defined in (5.1) have opposite conformal dimensions. It is convenient, however, to use the variables with the same conformal dimensions.[5] To achieve this, we rescale η as

$$\eta^i \longrightarrow \sqrt{2}e^{\phi}\eta^i, \quad \eta_i \longrightarrow \sqrt{2}e^{\phi}\eta_i. \tag{6.15}$$

The action in Eqs. (6.13) and (6.14) can then be written as

$$\begin{aligned}\mathcal{L}_{\rm kin} &= \sqrt{g}g^{\mu\nu}\Big[-e^{2\phi}(\partial_\mu x^+\partial_\nu x^- + \partial_\mu x\partial_\nu \bar{x}) \\ &\quad -\frac{1}{2}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}G_{\mathcal{A}\mathcal{B}}(y)D_\mu y^{\mathcal{A}}D_\mu y^{\mathcal{B}}\Big] \\ &\quad + \frac{i}{\sqrt{2}}\sqrt{g}g^{\mu\nu}e^{2\phi}\partial_\mu x^+[\theta^i\partial_\nu\theta_i + \theta_i\partial_\nu\theta^i \\ &\quad + \eta^i\partial_\nu\eta_i + \eta_i\partial_\nu\eta^i - ie^{2\phi}\partial_\nu x^+(\eta^2)^2],\end{aligned} \tag{6.16}$$

$$\begin{aligned}\mathcal{L}_{\rm WZ} &= -\epsilon^{\mu\nu}e^{2\phi}\partial_\mu x^+\tilde{\eta}^i C'_{ij} \\ &\quad \times(\tilde{\partial_\nu}\theta^j + i\sqrt{2}e^{\phi}\tilde{\eta}^j\partial_\nu x) + \text{H.c.},\end{aligned} \tag{6.17}$$

where $G_{\mathcal{A}\mathcal{B}}$ is the metric of the 5-sphere [6] and the differential $D_\mu y^{\mathcal{A}}$ is defined by

$$D_\mu y^{\mathcal{A}} = \partial_\mu y^{\mathcal{A}} + 2i\eta_i(V^{\mathcal{A}})^i{}_j\eta^j e^{2\phi}\partial_\mu x^+, \tag{6.18}$$

with $(V^{\mathcal{A}})^i_j$ being the components of the Killing vectors $(V^{\mathcal{A}})^i_j\partial_{\mathcal{A}}$ of S^5 (and $\partial_{\mathcal{A}} = \partial/\partial y^{\mathcal{A}}$).

We note that x^+ enters the action only through the combination $e^{2\phi}\partial_\mu x^+$. An attractive feature of this representation is that the terms in (6.13) involving $\tilde{\eta}_i(\gamma^{A'})^i_j\tilde{\eta}^j$ are now collected in the second term in the derivative (6.18) and, thus, have a natural geometrical interpretation, i.e., multiplying the Killing vectors.

The Killing vectors $(V^{\mathcal{A}})^i_j\partial_{\mathcal{A}}$ satisfy the so(6) $\simeq$ su(4) commutation relations (2.33) and can be written as

$$(V^{\mathcal{A}})^i{}_j\partial_{\mathcal{A}} = \frac{1}{4}(\gamma^{A'B'})^i{}_jV^{A'B'} + \frac{i}{2}(\gamma^{A'})^i{}_jV^{A'}, \tag{6.19}$$

where $V^{A'}$ and $V^{A'B'}$ correspond to the 5 translations and the $SO(5)$ rotations respectively and are given by (cf. (5.21))

$$V^{A'} = [|y|\cot|y|(\delta^{A'\mathcal{A}} - n^{A'}n^{\mathcal{A}}) + n^{A'}n^{\mathcal{A}}]\partial_{\mathcal{A}}, \tag{6.20}$$

$$V^{A'B'} = y^{A'}\partial_{B'} - y^{B'}\partial_{A'}, \tag{6.21}$$

where $\delta^{A'\mathcal{A}}$ is the Kronecker delta symbol and we use the conventions

$$y^{\mathcal{A}} = \delta^{\mathcal{A}}_{A'}y^{A'}, \quad n^{\mathcal{A}} = \delta^{\mathcal{A}}_{A'}n^{A'}, \quad n^{\mathcal{A}} = n_{\mathcal{A}}.$$

In these coordinates, the S^5 metric tensor has the form

$$\begin{aligned}G_{\mathcal{A}\mathcal{B}} &= e^{A'}_{\mathcal{A}}e^{A'}_{\mathcal{B}}, \\ e^{A'}_{\mathcal{A}} &= \frac{\sin|y|}{|y|}(\delta^{A'}_{\mathcal{A}} - n_{\mathcal{A}}n^{A'}) + n_{\mathcal{A}}n^{A'}.\end{aligned} \tag{6.22}$$

In this section, we have discussed the light-cone action in the Killing parametrization of superspace. To obtain the light-cone-gauge action in the Wess–Zumino parametrization, one must make the redefinitions [cf. (5.19) and (5.20)]

$$\theta^i \longrightarrow (U^{-1})^i{}_j\theta^i, \quad \theta_i \longrightarrow \theta_jU^j{}_i, \tag{6.23}$$

$$\eta^i \longrightarrow \sqrt{2}e^{\phi}(U^{-1})^i{}_j\eta^j, \quad \eta_i \longrightarrow \sqrt{2}e^{\phi}\eta_jU^j{}_i. \tag{6.24}$$

The fermionic derivatives ∂_μ then acquire the generalized connection $\Omega_\mu = \partial_\mu UU^{-1}$ (Eq. (7.7)); i.e., they become the covariant derivatives $\mathcal{D}_\mu$ (see (7.6)). The action is given in terms of these new variables in the next section.

Finally, we note that our results for the $AdS_5 \times S^5$ space can be generalized to $AdS_3 \times S^3$ in a rather straightforward way. To obtain the light-cone-gauge action for this case, one could use the κ-invariant action of [39] and then apply the same procedure of the light-cone splitting and gauge fixing as developed in this paper. However, our light-cone-gauge action is already written in the form that allows a straightforward generalization to $AdS_3 \times S^3$: only the dimensional reduction remains to be done. In discussing the $AdS_3 \times S^3$ Lagrangian, we use the WZ parametrization where the action has the form given by (7.1) for definiteness. To

[5] Similar Grassmann variables with the same conformal dimensions were used in the light-cone formulation of the superparticle in $AdS_5 \times S^5$ [24, 25].

[6] We introduced the coordinate S^5 indices $\mathcal{A}, \mathcal{B} = 1, \ldots, 5$ (to be distinguished from the tangent space indices A', B') and set $y^{\mathcal{A}} = \delta^{\mathcal{A}}_{A'}y^{A'}$.

obtain the $\mathcal{L}_B$ and $\mathcal{L}_F^{(2)}$ terms in the $AdS_3 \times S^3$ case, we must set

$$x = \bar{x} = 0$$

in (7.2) and (7.5) and also assume the fermionic variables θ and η to transform in the fundamental representation of $\widetilde{\ }$(2) and SU(2) (with the indices i, j taking the values 1, 2). The matrix C'_{ij} is then given by

$$C' = h\sigma_2, \quad |h| = 1.$$

The matrices $(\gamma^{A'})^i{}_j$ where $A' = 1, 2, 3$, are now the SO(3) Dirac gamma matrices. The quartic part of the Lagrangian $\mathcal{L}_F^{(4)}$ simplifies to

$$\mathcal{L}_F^{(4)} = 2\sqrt{g}g^{\mu\nu}e^{4\phi}\partial_\mu x^+ \partial_\nu x^+ (\eta^i\eta_i)^2. \qquad (6.25)$$

7. DIFFERENT FORMS AND SOME PROPERTIES OF THE LIGHT-CONE-GAUGE ACTION

Choosing the light-cone gauge in the parametrization of the supercoset

$$\frac{\mathrm{PSU}(2,2|4)}{\mathrm{SO}(4,1)\times\mathrm{SO}(5)}$$

described above, we can write the $AdS_5 \times S^5$ superstring Lagrangian of [6] (in the Wess–Zumino superspace parametrization) as

$$\mathcal{L} = \mathcal{L}_B + \mathcal{L}_F^{(2)} + \mathcal{L}_F^{(4)}, \qquad (7.1)$$

where

$$\mathcal{L}_B = -\frac{1}{2}\sqrt{g}g^{\mu\nu}G_{MN}(X)\partial_\nu X^M \partial_\nu X^N$$

is the standard bosonic sigma-model action with $AdS_5 \times S^5$ as the target space,[7]

$$\mathcal{L}_B = -\sqrt{g}g^{\mu\nu} \times \left[e^{2\phi}(\partial_\mu x^+ \partial_\nu x^- + \partial_\mu x \partial_\nu \bar{x}) + \frac{1}{2}\partial_\mu\phi\partial_\nu\phi + \frac{1}{2}e_\mu^{A'}e_\nu^{A'}\right]. \qquad (7.2)$$

In this formula, $e_\mu^{A'}$ is the projection of the S^5 vielbein that in our special parametrization is given by

$$e_\mu^{A'} = -\frac{i}{2}\mathrm{Tr}(\gamma^{A'}\partial_\mu U U^{-1}), \qquad U^i{}_j \equiv (e^y)^i{}_j, \quad U^\dagger U = I, \qquad (7.3)$$

where Tr is taken over i, j. The matrix $U \in SU(4)$ depends on five independent coordinates $y^{A'}$,

$$y^i{}_j \equiv \frac{i}{2}y^{A'}(\gamma^{A'})^i{}_j, \quad (y^i{}_j)^* = -y^j{}_i, \quad y^i{}_i = 0, \qquad (7.4)$$

where $y^{A'}$ are the SO(5) Dirac matrices. In Eq. (7.1), $\mathcal{L}_F^{(2)}$ is the quadratic part of the fermionic action,

$$\mathcal{L}_F^{(2)} = e^{2\phi}\partial_\mu x^+ \times \left[\frac{i}{2}\sqrt{g}g^{\mu\nu}(\theta_i\mathcal{D}_\nu\theta^i + \eta_i\mathcal{D}_\nu\eta^i - i\eta_i e^i_{\nu j}\eta^j) - \epsilon^{\mu\nu}\eta^i C'_{ij}(\mathcal{D}_\nu\theta^j + i\sqrt{2}e^\phi\eta^j\partial_\nu x)\right] + \text{H.c.} \qquad (7.5)$$

Here, we used the notation for the generalized spinor derivative on S^5

$$\mathcal{D}\theta^i = d\theta^i - \Omega^i{}_j\theta^j, \quad \mathcal{D}\theta_i = d\theta_i + \theta_j\Omega^j{}_i, \qquad e^i{}_j \equiv (\gamma^{A'})^i{}_j e^{A'}, \qquad (7.6)$$

with

$$\mathcal{D} = \mathcal{D}_\mu d\sigma^\mu, \quad e^i{}_j = e^i_{\mu j}d\sigma^\mu,$$

where $\sigma^\mu = (\tau, \sigma)$ are two-dimensional coordinates.

This derivative has the general representation

$$\mathcal{D} = d + \Omega^i{}_j J^j{}_i$$

and satisfies the relation

$$\mathcal{D}^2 = 0,$$

with Ω^i_j given by

$$\Omega = dUU^{-1}, \quad d\Omega - \Omega\wedge\Omega = 0, \qquad (7.7)$$

and can be written in terms of the S^5 spin connection $\omega^{A'B'}$ and the 5-bein $e^{A'}$ as

$$\Omega^i{}_j = -\frac{1}{4}(\gamma^{A'B'})^i{}_j\omega^{A'B'} + \frac{i}{2}(\gamma^{A'})^i{}_j e^{A'}. \qquad (7.8)$$

C'_{ij} is the constant charge conjugation matrix of the SO(5) Dirac matrix algebra:

$$C'^\dagger C' = I, \quad C'^T = -C'.$$

The hermitean conjugation rules are

$$\theta_i^\dagger = \theta^i, \quad \eta_i^\dagger = \eta^i.$$

The P-odd $\epsilon^{\mu\nu}$-dependent term in (7.5) comes from the WZ term in the original supercoset GS action [6].

The quartic fermionic term in (7.1) depends only on η, not on θ:

$$\mathcal{L}_F^{(4)} = \frac{1}{2}\sqrt{g}g^{\mu\nu}e^{4\phi}\partial_\mu x^+ \times \partial_\nu x^+[(\eta^i\eta_i)^2 - (\eta_i(\gamma^{A'})^i{}_j\eta^j)^2]. \qquad (7.9)$$

[7] Our index notation differs from [6]: here, we use $\mu, \nu = 0, 1$ for two-dimensional indices; i, j for $SU(4)$ indices; $A = 0, 1, \ldots, 4$ for AdS_5; and $A' = 1, \ldots, 5$ for S^5 tangent space indices (repeated indices are contracted with flat metric). We set $\epsilon^{01} = 1$.

The action in Eqs. (7.1), (7.2), (7.5), and (7.9) has several important properties.

(i) The dependence on x^- is only linear through the bosonic $\partial x^+ \partial x^-$ term in (7.2). The bosonic factor in the fermion kinetic term is simply $e^{2\phi}\partial x^+$. It is the crucial property of this light-cone κ-symmetry gauge-fixed form of the action that the fermion kinetic term involves the derivative of only one space-time direction, x^+; this implies that the (non)degeneracy of the action does not depend on the transverse string profile.[8]

(ii) The fact that the action has only quadratic and quartic fermionic terms is related to special symmetries of the $AdS_5 \times S^5$ background (a covariantly constant curvature and the 5-form field strength). The presence of the η^4 term (7.9) reflects the curvature of the background.[9] As follows from the discussion in [6], the "extra" terms in (7.5) similar to $\eta_i e^i_j \eta^j$ and $\eta C' \eta \partial x$ must have the interpretation of the couplings to the RR 5-form background. The gauge that we considered treats the AdS_5 and S^5 factors asymmetrically. In particular, the action contains only SO(5) but not SO(4, 1) gamma matrices, and θ_i and η_i are not spinors under SO(4, 1).[10]

(iii) The $AdS_5 \times S^5$ superstring action depends on two parameters: the scale (equal radii) R of $AdS_5 \times S^5$ and the inverse string tension or α'. Restoring the dependence on R that was set equal to 1 in (7.1), one finds that in the flat-space limit as $R \longrightarrow \infty$, the quartic term (7.9) goes away, while the kinetic term (7.5) reduces to the standard one with $\mathcal{D}_\mu \longrightarrow \partial_\mu$. The resulting action is equivalent to the flat-space light-cone GS action [3] after expressing each of the two SO(8) spinors in terms of the two SU(4) spinors. The action takes the "diagonal" form in terms of the combinations $\psi^i_{1,2}$ of our two fermionic variables (see (7.19) in what follows). As $\alpha' \longrightarrow 0$, the action has the correct particle limit; i.e., it reduces to the light-cone-gauge superparticle action in $AdS_5 \times S^5$ [24].

(iv) A special feature of this action is that the SU(4) ≈ SO(6) symmetry is realized linearly on fermions but not on bosons. This is a consequence of the purely bosonic factor SO(4, 1) × SO(5) in the underlying supercoset

$$\frac{\mathrm{PSU}(2,2|4)}{\mathrm{SO}(4,1)\times \mathrm{SO}(5)}.$$

The S^5 = SO(6)/SO(5) part of the bosonic action can be represented as a special case of the two-dimensional G/H coset sigma model with the Lagrangian

$$L = \mathrm{Tr}(\partial_\mu \mathcal{U} \mathcal{U}^{-1} + A_\mu)^2, \quad \mathcal{U} \in G = \mathrm{SO}(6),$$

where the two-dimensional gauge field A_μ is in the algebra of H = SO(5). This action does not have the manifested SO(6) symmetry after A_μ is integrated out and $\mathcal{U}$ is restricted to the coset as a gauge choice.

(v) The action is symmetric under the shift $\theta \longrightarrow \theta + \epsilon$ supplemented by an appropriate transformation of x^-. Here, ϵ is a Killing spinor on S^5 satisfying the equation $\mathcal{D}\epsilon^i = 0$. It is because of this symmetry that the theory is linear in θ (i.e., no quartic interactions in θ occur).

To proceed further with quantizing the theory, one would like, as in the flat case, to eliminate the ∂x^+ factors from the fermion kinetic terms in (7.5). In the flat-space, this was possible in the bosonic light-cone gauge. In the BDHP formulation [30, 31] that we are using, this can be done by fixing the conformal gauge

$$\gamma^{\mu\nu} = \eta^{\mu\nu}, \quad \gamma^{\mu\nu} \equiv \sqrt{g} g^{\mu\nu}, \quad \det\gamma^{\mu\nu} = -1 \qquad (7.10)$$

and then noting that, because the resulting equation

$$\partial^2 x^+ = 0$$

has a general solution

$$x^+(\tau, \sigma) = f(\tau - \sigma) + h(\tau + \sigma),$$

we can fix the residual conformal diffeomorphism symmetry on the plane by setting

$$x^+(\tau, \sigma) = p^+ \tau.$$

An alternative (equivalent) approach is to use the original GGRT [32] formulation based on writing the Nambu action in the canonical first-order form (with constraints added with Lagrange multipliers) and fixing the diffeomorphism invariance by two conditions imposed on one coordinate and one canonical momentum:

$$x^+ = p^+ \tau, \quad P^+ = \mathrm{const}.$$

The first approach based on the conformal gauge does not generally apply to curved spaces with null Killing vectors that are not the direct products $R^{1,1} \times M^8$

[8] The structure of the action is therefore similar to that of the light-cone-gauge action for the GS string in the curved magnetic RR background constructed in [26].

[9] We note that the light-cone-gauge GS action in a curved space of the form $R^{1,1} \times M^8$ with generic NSNS and RR backgrounds [27] (reconstructed from the light-cone flat-space GS vertex operators [28]) generally contains, terms higher than quartic fermionic terms multiplied be higher derivatives of the background fields. This light-cone GS action has a quartic fermionic term [27, 29] involving the curvature tensor

$$R_{..}\partial x^+ \partial x^+ (\bar\theta \Gamma^{-\cdot\cdot} \theta)(\bar\theta \Gamma^{-\cdot\cdot} \theta) \sim R_{..}(p^+)^2 (\bar\theta \Gamma^{-\cdot\cdot} \theta)(\bar\theta \Gamma^{-\cdot\cdot} \theta)$$

that is similar to the one present in the NSR string action (i.e., in the standard two-dimensional supersymmetric sigma model).

[10] θ and η are not scalars with respect to SO(4, 1). Combined with fermions eliminated by the κ-symmetry gauge, they transform in the spinor representation of SO(4, 1) × SO(5). However, after the gauge fixing based on gamma matrices from the AdS_5 part ($\gamma^+\theta = 0$), the SO(4, 1) group, with the exception of its SO(2) subgroup generated by $J^{x\bar{x}}$ [23] (which is a part of the little group for the AdS_5 case), is realized nonlinearly. Thus (modulo subtleties of a nonlinear realization of su(4) on bosons), the algebra so(2) ⊕ su(4) is a counterpart of the so(8) algebra in the flat-space case.

(the gauge conditions is inconsistent in general with the classical equations of motion). It does apply, however, if the null Killing vector is covariantly constant [33]. There is no need, in principle, to insist on fixing the standard conformal gauge (7.10). Instead, one can fix the diffeomorphism gauge by imposing two conditions:

$$\gamma^{00} = -1, \quad x^+ = p^+\tau.$$

This choice is consistent provided the background metric satisfies [34]

$$G_{+-} = 1, \quad G_{--} = G_{-i} = 0, \quad \partial_- G_{MN} = 0.$$

This approach is essentially equivalent to the GGRT approach applied to the curved space case.

The above conditions do not apply to the AdS case: the null Killing vectors are not covariantly constant and $G_{+-} = e^{2\phi} \neq 1$.[11] It is easy to see, however, that a slight modification of the above conditions on γ^{00} and x^+ represents a consistent gauge choice:

$$e^{2\phi}\gamma^{00} = -1, \quad x^+ = p^+\tau. \tag{7.11}$$

In what follows, we do not discuss in detail the consequences of fixing the bosonic light-cone gauge (7.11) in superstring action (7.1) but follow a simplified approach based on using a particular classical solution.

We first do not make any explicit gauge choice but consider the superstring path integral assuming that there are no sources for x^-. The linear dependence of the action in Eqs. (7.1) and (7.2) on x^- allows us to integrate over x^- explicitly. This produces the δ-function constraint imposing the equation of motion for x^+, which is formally solved by setting

$$\sqrt{g}g^{\mu\nu}e^{2\phi}\partial_\nu x^+ = \epsilon^{\mu\nu}\partial_\nu f, \tag{7.12}$$

where $f(\tau, \sigma)$ is an arbitrary function. Because our action (7.1) depends on x^+ only through $e^{2\phi}\partial x^+$, we are then able to integrate over x^+ as well, eliminating it in favor of the function f. The action then contains the fermionic terms (7.5) and (7.9) with

$$e^{2\phi}\partial_\mu x^+ \longrightarrow f_\mu \equiv g_{\mu\nu}\frac{\epsilon^{\nu\lambda}}{\sqrt{g}}\partial_\lambda f. \tag{7.13}$$

The resulting fermion kinetic term is then nondegenerate (for a properly chosen f) and can be interpreted as the action of two-dimensional fermions in a curved two-dimensional geometry determined by f and $g_{\mu\nu}$ (cf. [4, 37, 38]).

We can further simplify the action by making a special choice of f and fixing a diffeomorphism gauge on $g_{\mu\nu}$ in a consistent way. One possibility is to choose a diagonal gauge on $g_{\mu\nu}$ [36] and $f \sim \sigma$ in accordance with (7.12), which implies that $x^+ \sim \tau$; i.e.,[12]

$$f = \sigma, \quad x^+ = \tau, \quad \sqrt{g}g^{\mu\nu} = \mathrm{diag}(-e^{-2\phi}, e^{2\phi}). \tag{7.14}$$

Similarly to the flat-space case [3] and the "long" string cases discussed in [4], the resulting action can then be put into the "two-dimensional spinor" form. Indeed, the "8 + 8" fermionic degrees of freedom can be organized into four Dirac two-dimensional spinors defined in the curved two-dimensional geometry. Using (7.14), we can write kinetic term (7.5) as

$$\begin{aligned}\mathcal{L}_F^{(2)} &= -\frac{i}{2}(\theta_i\mathcal{D}_0\theta^i + \eta_i\mathcal{D}_0\eta^i - i\eta_i e_{0j}^i\eta^j) \\ &- e^{2\phi}\eta^i C'_{ij}(\mathcal{D}_1\theta^j + i\sqrt{2}e^\phi\eta^j\partial_1 x) + \mathrm{H.c.}\end{aligned} \tag{7.15}$$

Introducing a two-dimensional zweibein corresponding to the metric in (7.14),

$$e_\mu^m = \mathrm{diag}(e^{2\phi}, 1), \quad g_{\mu\nu} = -e_\mu^0 e_\nu^0 + e_\mu^1 e_\nu^1, \tag{7.16}$$

we can put (7.15) in the two-dimensional form as

$$\begin{aligned}e^{-1}\mathcal{L}_F^{(2)} &= \frac{i}{2}\bar\psi\rho^m e_m^\mu\mathcal{D}_\mu\psi - \frac{i}{2}\bar\psi\psi\partial_1\phi \\ &+ \frac{1}{\sqrt{2}}\bar\psi_i e_{0j}^i\rho^-\psi^j + i\sqrt{2}e^\phi(\psi^i)^T\pi^- C'_{ij}\psi^j\partial_1\bar{x} + \mathrm{H.c.},\end{aligned} \tag{7.17}$$

where ρ^m are two-dimensional Dirac matrices,

$$\begin{aligned}&\rho^0 = i\sigma_2, \quad \rho^1 = \sigma_1, \quad \rho^3 = \rho^0\rho^1 = \sigma_3, \\ &\rho^\pm \equiv \frac{1}{\sqrt{2}}(\rho^3 \pm \rho^0), \quad \pi^- \equiv \frac{1}{2}(1 - \rho^1),\end{aligned} \tag{7.18}$$

$\bar\psi_i = (\psi^i)^\dagger\rho^0$, $\bar\psi\psi$ stands for $\bar\psi_i\psi^i$, ψ^T denotes the transposition of a two-dimensional spinor, and ψ are related to the original (two-dimensional scalar) fermionic variables θ and η by[13]

[11] In fact, there is no globally well-defined null Killing vector in the AdS space because its norm, which is proportional to $e^{2\phi}$, vanishes at the horizon $\phi = -\infty$ (this point and the possibility to fix a global diffeomorphism gauge for the AdS string were discussed in [35]). In this paper, we use a formal approach to this issue: since the boundary SYM theory in $R^{1,3}$ has a well-defined light-cone description, it must be possible to fix some analogue of the light-cone gauge for the dual string as well (assuming it is defined on the Poincaré patch of the AdS space). A potential problem of that approach, which will be reflected in the degeneracy of the resulting light-cone gauge-fixed action near the horizon region, should then be addressed at a later stage.

[12] We note that the standard conformal gauge $\sqrt{g}g^{\mu\nu} = \mathrm{diag}(-1, 1)$ leads to an inconsistency for generic ϕ if one insists on the simplest $f = \sigma$ choice. Consistency for generic ϕ is achieved only if f (and x^+) is nontrivial. However, the structure of the resulting action is complicated.

[13] In our notation $i\bar\psi\rho^m\nabla_m\psi = -i\psi_1^\dagger(\nabla_0 - \nabla_1)\psi_1 - i\psi_2^\dagger(\nabla_0 + \nabla_1)\psi_2$, $\nabla_m = e_m^\mu\partial_\mu$.

$$\psi^i = \begin{pmatrix} \psi_1^i \\ \psi_2^i \end{pmatrix}, \quad \psi_1^i = \frac{1}{\sqrt{2}}[\theta^i - i(C'^{-1})^{ij}\eta_j], \qquad (7.19)$$
$$\psi_2^i = \frac{1}{\sqrt{2}}[\theta^i + i(C'^{-1})^{ij}\eta_j].$$

The quartic interaction term (7.9) takes the form

$$e^{-1}\mathcal{L}_F^{(4)} = \frac{1}{4}[(\bar\psi_i(\gamma^{A'})^i{}_j\rho^-\psi^j)^2 - (\bar\psi_i\rho^-\psi^j)^2]. \quad (7.20)$$

The total action is thus a certain G/H bosonic sigma model coupled to a Thirring-type two-dimensional fermionic model in curved two-dimensional geometry (7.16) (determined by the profile of the radial function of the AdS space) and coupled to some two-dimensional vector fields. The interactions are such that they ensure the quantum two-dimensional conformal invariance of the entire model [6].

Properties of the resulting action and the possibility of putting it into a simpler and more useful form remain to be studied. It is clear of course that the action has a rather complicated structure and is not solvable in terms of free fields in any obvious way. A hope is that the light-cone form of the action that we have found (or its first-order phase-space analog) may suggest a choice of more adequate variables that may allow further progress.

We finish this discussion with several remarks.

The mass term $\bar\psi\psi\partial_1\phi$ in (7.17) is similar to the one in [4] (where the background string configuration was nonconstant only in the radial ϕ direction) and has its origin in the $\epsilon^{\mu\nu}e^{2\phi}\partial_\mu x^+\partial_\nu\phi\eta^i C'_{ij}\theta^j$ term appearing after the $\eta \longleftrightarrow \theta$ symmetrization of the $\epsilon^{\mu\nu}$ term in (7.5) (its "noncovariance" is thus a consequence of the choice $x^+ = \tau$).

The action is symmetric under shifting

$$\psi^i \longrightarrow \psi^i + \rho^-\epsilon^i,$$

where ϵ^i is the two-dimensional Killing spinor. This symmetry reflects the fact that our original action is symmetric under shifting θ^i by a Killing spinor on S^5.

The two-dimensional Lorentz invariance is preserved by the fermionic light-cone gauge (the original GS fermions θ are two-dimensional scalars) but is broken by our special choice of bosonic gauge (7.14). The special form of $g_{\mu\nu}$ in (7.14) implies a "noncovariant" dependence on ϕ in the bosonic part of the action: the action (7.2) for the three fields ϕ, x, $\bar x$ and the 5-sphere coordinates $y^{A'}$ is given by

$$\mathcal{L}_B = \partial_0 x\partial_0\bar x - e^{4\phi}\partial_1 x\partial_1\bar x + \frac{1}{2}e^{-2\phi}\partial_0\phi\partial_0\phi - \frac{1}{2}e^{2\phi}\partial_1\phi\partial_1\phi + \frac{1}{2}G_{\mathcal{AB}}(y)(e^{-2\phi}\partial_0 y^{\mathcal{A}}\partial_0 y^{\mathcal{B}} - e^{2\phi}\partial_1 y^{\mathcal{A}}\partial_1 y^{\mathcal{B}}), \qquad (7.21)$$

where $G_{\mathcal{AB}}$ is the metric of the 5-sphere.[14] A peculiarity of the $g_{\mu\nu}$ gauge choice in (7.14) compared to the usual conformal gauge is that here the S^5 part of the action is no longer decoupled from the radial AdS_5 direction ϕ.

The form of the quadratic fermionic part of the $AdS_5 \times S^5$ superstring action expanded near a straight long string configuration along the ϕ direction of AdS_5 was discussed in [4] using the "covariant" κ-symmetry gauge condition $\theta^1 = \theta^2$ (an equivalent result was also found in the $\theta^1 = i\gamma_4\theta^2$ gauge used in [14, 15]). It is easy to show that an equivalent fermionic action is also found in the present light-cone κ-symmetry gauge. Expanding near the configuration $x^0 = \tau$, $\phi = \sigma$, $x = 0$, $y = 0$ (which is a classical string solution, as is easy to check) and choosing the bosonic gauge such that the two-dimensional metric $g_{\mu\nu}$ is equal to the induced (AdS_2) metric

$$ds^2 = \frac{1}{\sigma^2}(d\tau^2 + d\sigma^2).$$

We find that the corresponding function f in (7.12) is then equal to σ^{-2}. The quadratic part of the fermion action (7.17) becomes (with the η fermions redefined by the constant unitary matrix C' in (7.15))

$$\int d\tau d\sigma\sigma^{-2}(\theta\partial_0\theta + \eta\partial_0\eta - \eta\partial_1\theta). \qquad (7.22)$$

Rescaling the fields as $\theta = \sigma\theta'$ and $\eta = \sigma\eta'$ such that their normalization is σ-independent,

$$\int d\tau d\sigma\sqrt{g}\theta\theta = \int d\tau d\sigma\theta'\theta',$$

and integrating by parts, we find

$$\int d\tau d\sigma(\theta'\partial_0\theta' + \eta'\partial_0\eta' - \eta'\partial_1\theta' - \sigma^{-1}\eta'\theta'). \quad (7.23)$$

The first three terms here are similar to those in the flat GS action, while the last term represents the AdS_2 fermion mass term and is the same as found in [4]. Indeed, diagonalizing the action as in (7.19), we obtain

$$\int d\tau d\sigma(\psi_+\partial_+\psi_+ + \psi_-\partial_-\psi_- - \sigma^{-1}\psi_+\psi_-), \qquad (7.24)$$

[14]Here, we renamed the (tangent space) indices A', B' into the coordinate space ones $\mathcal{A}$, $\mathcal{B}$ for consistency with the notation used later in Section 6 ($y^{\mathcal{A}} \equiv y^{A'}$).

which is the special case of the general form of the quadratic action (7.17) with $\partial_\sigma\phi$ in the mass term computed for $\phi = \ln\sigma$.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 99-02-17916), the INTAS (grant no. 991590), and the DOE (grant no. DOE/ER/01545).

REFERENCES

1. A. M. Polyakov, hep-th/0006132; Int. J. Mod. Phys. A **14**, 645 (1999); hep-th/9809057; Nucl. Phys. B (Proc. Suppl.) **68**, 1 (1998); hep-th/9711002.
2. J. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); hep-th/9711200; S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, Phys. Lett. B **428**, 105 (1998); hep-th/9802109; E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998); hep-th/9802150.
3. M. B. Green and J. H. Schwarz, Phys. Lett. B **136B**, 367 (1984); Nucl. Phys. B **243**, 285 (1984).
4. N. Drukker, D. J. Gross, and A. A. Tseytlin, JHEP **0004**, 021 (2000); hep-th/0001204.
5. M. Henneaux and L. Mezincescu, Phys. Lett. B **152B**, 340 (1985).
6. R. R. Metsaev and A. A. Tseytlin, Nucl. Phys. B **533**, 109 (1998); hep-th/9805028.
7. J. H. Schwarz, Nucl. Phys. B **226**, 269 (1983).
8. M. Günaydin and N. Marcus, Class. Quantum Grav. **2**, L19 (1985).
9. M. Günaydin, L. J. Romans, and N. P. Warner, Phys. Lett. B **154B**, 268 (1985); M. Pernici, K. Pilch, and P. van Nieuwenhuizen, Nucl. Phys. B **259**, 460 (1985).
10. R. Haag, J. T. Lopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975).
11. R. Kallosh, J. Rahmfeld, and A. Rajaraman, JHEP **9809**, 002 (1998); hep-th/9805217.
12. R. R. Metsaev and A. A. Tseytlin, Phys. Lett. B **436**, 281 (1998); hep-th/9806095.
13. I. Pesando, JHEP **9811**, 002 (1998); hep-th/9808020; Mod. Phys. Lett. A **14**, 343 (1999); hep-th/9808146.
14. R. Kallosh and J. Rahmfeld, Phys. Lett. B **443**, 143 (1998); hep-th/9808038.
15. R. Kallosh and A. A. Tseytlin, JHEP **9810**, 016 (1998); hep-th/9808088.
16. J. Maldacena, Phys. Rev. Lett. **80**, 4859 (1998); hep-th/9803002; S.-J. Rey and J. Yee, hep-th/9803001.
17. S. Forste, D. Ghoshal, and S. Theisen, JHEP **9808**, 013 (1999); hep-th/9903042.
18. M. B. Green and J. H. Schwarz, Phys. Lett. B **109B**, 444 (1982).
19. S. Carlip, Nucl. Phys. B **284**, 365 (1987); R. E. Kallosh, Phys. Lett. B **195**, 369 (1987).
20. R. Kallosh and A. Yu. Morozov, Int. J. Mod. Phys. A **3**, 1943 (1988).
21. I. Pesando, hep-th/9912284.
22. L. Brink, O. Lindgren, and B. E. Nilsson, Nucl. Phys. B **212**, 401 (1983); Phys. Lett. B **123B**, 323 (1983); S. Mandelstam, Nucl. Phys. B **213**, 149 (1983); L. Brink, M. B. Green, and J. H. Schwarz, Nucl. Phys. B **223**, 125 (1983).
23. R. R. Metsaev, Nucl. Phys. B **563**, 295 (1999); hep-th/9906217.
24. R. R. Metsaev, Phys. Lett. B **468**, 65 (1999); hep-th/9908114.
25. R. R. Metsaev, hep-th/0002008.
26. J. Russo and A. A. Tseytlin, JHEP **9804**, 014 (1998); hep-th/9804076.
27. E. S. Fradkin and A. A. Tseytlin, Phys. Lett. B **160B**, 69 (1985).
28. M. B. Green and J. H. Schwarz, Nucl. Phys. B **198**, 252 (1982); J. H. Schwarz, Phys. Rep. **8**, 269 (1973).
29. P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985).
30. L. Brink, P. Di Vecchia, and P. Howe, Phys. Lett. B **65B**, 471 (1976).
31. A. M. Polyakov, Phys. Lett. B **103B**, 207 (1981).
32. P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, Nucl. Phys. B **56**, 109 (1973).
33. G. T. Horowitz and A. R. Steif, Phys. Rev. D **42**, 1950 (1990).
34. R. E. Rudd, Nucl. Phys. B **427**, 81 (1994); hep-th/9402106.
35. A. Rajaraman and M. Rozali, Phys. Lett. B **468**, 58 (1999); hep-th/9902046.
36. A. M. Polyakov, private communication (2000).
37. P. B. Wiegmann, Nucl. Phys. B **323**, 330 (1989).
38. A. R. Kavalov, I. K. Kostov, and A. G. Sedrakian, Phys. Lett. B **175**, 331 (1986); A. G. Sedrakian and R. Stora, Phys. Lett. B **188**, 442 (1987); D. R. Karakhanian and A. G. Sedrakian, Preprint YERPHI-1127(4)-89 (1989); D. R. Karakhanian, Preprint YERPHI-1246-32-90 (1990).
39. I. Pesando, JHEP **9902**, 007 (1999); hep-th/9809145; J. Rahmfeld and A. Rajaraman, Phys. Rev. D **60**, 064014 (1999); hep-th/9809164; J. Park and S. Rey, JHEP **9901**, 001 (1999); hep-th/9812062.

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1115–1129.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1291–1307.

ATOMS, SPECTRA, RADIATION

Theory of Multiple Scattering

V. I. Yurchenko

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
e-mail: yurchenko@vxinpz.inp.nsk.su

Received June 26, 2000

Abstract—Consideration is given to problems of obtaining exact and approximate solutions of kinetic equations in the multiple scattering problem. For cross sections which are rational functions of χ^2 ($\chi = 2\sin(\delta/2)$, δ is the scattering angle) exact solutions are obtained as a series in terms of Legendre polynomials. The limits of validity of the kinetic equation for the distribution function in terms of the variable $q = 2\sin(\vartheta/2)$ are refined [1] and the solutions of this equation are compared with the exact solutions of the Rutherford and Mott cross sections. The problem of convergence of approximate solutions in the form of a series in terms of Legendre polynomials and a series in powers of $1/B$ is solved. These approximations are obtained and their limits of validity are determined. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

At present the problem of multiple scattering is considered using the kinetic equation for the angular distribution function

$$\frac{\partial f(\mathbf{n}\cdot\mathbf{i},\, f)}{\partial t} = N\int\sigma(\mathbf{n}\cdot\mathbf{n}')[f(\mathbf{n}'\cdot\mathbf{i},\, f) - f(\mathbf{n}\cdot\mathbf{i},\, f)]\frac{d\Omega'}{2\pi}, \quad (1)$$

which for

$$f(\cos\vartheta, 0) = \delta(1-\cos\vartheta)$$

has the solution (P_l are Legendre polynomials)

$$f(\cos\vartheta, t) = \sum_{l=0}^{\infty}\frac{2l+1}{2}P_l(\cos\vartheta)\exp(-Q_l(t)), \quad (2)$$

$$Q_l(t) = Nt\int_0^{\pi}\sigma(\cos\delta)\sin\delta d\delta[1-P_l(\cos\delta)], \quad (3)$$

and also the approximate kinetic equation derived in [1] (the notation is the same as in [1])

$$\frac{\partial f(q,t)}{\partial t} = N\int\sigma(\chi)[f(|\mathbf{q}-\boldsymbol{\chi}|, t) - f(q,t)]\frac{d\boldsymbol{\chi}}{2\pi}, \quad (4)$$

where integration is performed over the planar region

$$q = 2\sin(\vartheta/2), \quad \chi = 2\sin(\delta/2).$$

For

$$f(q, 0) = \delta(1-\cos\vartheta) = \delta(q)/q$$

its solution is

$$f(q,t) = \int_0^{\infty}\eta d\eta J_0(\eta q)\exp(-Q(\eta, t)), \quad (5)$$

$$Q(\eta, t) = Nt\int_0^{\infty}\sigma(\chi)\chi d\chi\{1 - J_0(\eta\chi)\}. \quad (6)$$

In accordance with its conditions of derivation [1], this approximation can be called the small transferred momentum (small transfer) approximation although within the limits of validity of the method determined in [1] the solution (5), (6) is fairly exact for $q \sim 1$. The small angle approximation is not used, i.e., the solution is obtained for the general case of the cross section. For the Rutherford cross section it can be expressed as the series [1–4]

$$f(q) = \frac{1}{\lambda^2}[f^{(0)}(X) + B^{-1}f^{(1)}(X) + B^{-2}f^{(2)}(X) + \ldots],$$

$$X = \frac{q}{\lambda}, \quad \frac{1}{B} = \frac{\lambda_a^2}{\lambda^2}, \quad \lambda_a^2 = \bar{n}\chi_a^2,$$

where χ_a is the cutoff angle [2], $\bar{n}$ is the average number of collisions along the path t, and λ is the characteristic size of the multiple scattering region.

It was shown in [1] that Eq. (4) can be applied over the entire angular range and the approximate solutions of Eqs. (1) and (4) for the Rutherford and Mott cross sections were compared. This analysis raised various questions, which are listed as follows.

(1) In view of the lack of exact solutions of Eq. (1), the author used approximations for which the series (2) converges, which led to some procedural difficulties

since the results depend to some extent on the method of truncating the series (2).

(2) Over a wide range of λ values the solution (5) expressed as a series in powers of $1/B$ allowing for three terms of the expansion is much more exact in the single region ($q \gg \lambda$) than in the plural ($q \sim (2\text{–}4)\lambda$) and multiple ($0 \leq q \sim \lambda$) scattering regions. Consequently, we need to allow for a larger number of terms of the expansion.

(3) To this should be added the fact that series in powers of $1/B$ diverge (see Section 4 of the present study), i.e., questions arise in connection with estimating its accuracy and limits of validity.

(4) For these reasons it has not been possible to completely compare the solutions of Eqs. (1) and (4) for one particular cross section which is fundamental or at least desirable for any approximate method. In [1] the Monte Carlo method was used to monitor the approximations.

The aim of the present study is to solve these problems. Exact solutions of Eq. (1) are compared with the solutions of Eq. (4) for the Rutherford [2]

$$\sigma_R(\chi) = \frac{2s^2}{(\chi^2 + \chi_a^2)^2} \tag{7}$$

and Mott cross sections

$$\sigma_M(\chi) = \sigma_R(\chi)\left(1 - \frac{1}{4}\beta^2\chi^2\right) \tag{8}$$

allowing for the atomic form factor (the cutoff angle χ_a). Solutions of Eq. (1) for these cases are obtained in the present study. The approximate solutions are analyzed, including a converging series in powers of $1/B$, from the point of view of their suitability for a small average number of collisions. We first consider the general case of the cross section

$$\sigma(\chi) = \sigma_R(\chi)\kappa(\chi), \tag{9}$$

where the factor $\kappa(\chi)$ allows for the difference between the real cross section and the cross section (7).

2. GENERAL CROSS-SECTION CASE

We shall show that for the cross section (9) the general formula for Q_l in (2) has the form

$$Q_l = \chi_c^2 \sum_{k=1}^{l} \varepsilon^{k-1}\{C_{lk}[kA(\varepsilon) + \varepsilon A'(\varepsilon)] - ka_{lk}\}, \tag{10}$$

where

$$\chi_c^2 = Nts^2, \quad \varepsilon = \chi_a^2,$$

$$A(\varepsilon) = \int_0^2 \frac{\kappa(\chi)}{\chi^2 + \varepsilon} 2\chi d\chi,$$
$$A'(\varepsilon) = -\int_0^2 \frac{\kappa(\chi)}{(\chi^2 + \varepsilon)^2} 2\chi d\chi, \tag{11}$$

$$a_{lk} = (-1)^k \frac{1}{2^k} \int_{-1}^{1} \frac{\gamma_{lk}(x)\kappa(x)dx}{(1-x)^{k+1}}, \tag{12}$$

$$\gamma_{lk}(x) = 1 - P_l(x) - \sum_{m=1}^{k} (-1)^{m+1} C_{lm}\chi^{2m}. \tag{13}$$

In these formulas we have

$$x = \cos\delta, \quad \chi^2 = 2(1-x),$$

C_{lk} are the coefficients in the representation of $P_l(x)$ as a hypergeometric series:

$$P_l(x) = \sum_{k=0}^{l} (-1)^k C_{lk}\chi^{2k},$$
$$C_{lk} = \frac{1}{k!^2 4^k} \prod_{j=0}^{k-1} [l(l+1) - j(j+1)], \quad C_{l0} = 1. \tag{14}$$

With the representation (14) in the integral (3) we have

$$1 - P_l(x) = \sum_{k=1}^{l} (-1)^{k+1} C_{lk}\chi^{2k}. \tag{15}$$

For the cross section (9) the contribution of the kth term $(-1)^{k+1}C_{lk}\chi^{2k}$ in this sum to the integral (we assume $\chi^{2k} = [(\chi^2 + \chi_a^2) - \chi_a^2]^k$ and use the Newton binomial formula) has the form

$$\chi_c^2 \left\{ \varepsilon^{k-1} C_{lk}[kA(\varepsilon) + \varepsilon A'(\varepsilon)] + \sum_{j=0}^{k-2} b_j \varepsilon^j \right\}, \tag{16}$$

where b_j are various coefficients. Thus, Q_l can be expressed in the form (10) with as yet unknown (apart from $a_{ll} = 0$) values of a_{lk}.

In order to determine these we isolate the first k terms in the sum (15) and write Q_l as the sum of two integrals:

$$\frac{Q_l}{\chi_c^2} = \int_0^2 2\chi d\chi g(\chi) \sum_{m=1}^{k} (-1)^{m+1} C_{lm}\chi^{2m} + \int_0^2 2\chi d\chi g(\chi)\gamma_{lk}(\chi),$$

where

$$g(\chi) = \frac{\kappa(\chi)}{(\chi^2+\chi_a^2)^2},$$

the value of γ_{lk} is given by Eq. (13), and

$$\gamma_{l0}(\chi) = 1 - P_l(x), \quad \gamma_{ll}(\chi) = 0.$$

The first integral, in accordance with (16), gives power terms proportional to ε^m with $m \le k-2$ and terms of the type $\varepsilon^{m-1}[mA(\varepsilon) + \varepsilon A'(\varepsilon)]$ with $m \le k$. Taking into account (10) we obtain

$$\int_0^2 2\chi d\chi g(\chi)\gamma_{lk}(\chi) = -ka_{lk}\varepsilon^{k-1} + \sum_{m=k+1}^{l} \varepsilon^{m-1}\{C_{lm}[mA(\varepsilon) + \varepsilon A'(\varepsilon)] - ma_{lm}\} + (\text{terms} \propto \varepsilon^m, m \le k-2).$$

It then follows that

$$\lim_{\varepsilon\to 0}\frac{d^{k-1}}{d\varepsilon^{k-1}}\int_0^2 2\chi d\chi g(\chi)\gamma_{lk}(\chi) = -k!a_{lk},$$

which yields Eq. (12).

We note another case when

$$\sigma(\chi) = \frac{2s^2\kappa(\chi)}{\chi^2+\varepsilon}.$$

Since the negative derivative of this quantity with respect to ε is equal to the cross section (9), integrating (10) with respect to ε, we obtain for this case

$$Q_l = \chi_c^2\left\{a_{l0} - \sum_{k=1}^{l}\varepsilon^k[C_{lk}A(\varepsilon) - a_{lk}]\right\}. \tag{17}$$

The integration constant for $\varepsilon = 0$ is $a_{l0} = Q_l/\chi_c^2$ which allowing for (3) yields Eq. (12) for a_{l0}. Formula (17) can be used to write a solution for the cross section $\sigma(\chi) = \sigma_M(\chi)\kappa(\chi)$, taking into account that

$$\sigma_M(\chi) = \left(1 + \frac{1}{4}\varepsilon\beta^2\right)\sigma_R(\chi) - \frac{1}{4}\beta^2\frac{2s^2}{\chi^2+\chi_a^2}. \tag{18}$$

The same transformations can also be made for the solution (5), (6). Using instead of (14) the expansion

$$J_0(\eta\chi) = \sum_{k=0}^{\infty}(-1)^k C_k(\eta)\chi^{2k}, \qquad C_k(\eta) = \frac{1}{k!^2 4^k}\eta^{2k} \tag{19}$$

and repeating the previous reasoning, we obtain

$$Q(\eta) = \chi_c^2 \times \sum_{k=1}^{\infty}\varepsilon^{k-1}\{C_k(\eta)[kA(\varepsilon) + \varepsilon A'(\varepsilon)] - ka_k(\eta)\},$$

$$a_k(\eta) = (-1)^k\int_0^2\frac{\gamma_k(\eta\chi)}{\chi^{2(k+1)}}\kappa(\chi)2\chi d\chi,$$

$$\gamma_k(\eta\chi) = 1 - J_0(\eta\chi) - \sum_{m=1}^{k}(-1)^{m+1}C_m(\eta)\chi^{2m}.$$

As before, $A(\varepsilon)$ and $A'(\varepsilon)$ are given by Eq. (11). Note that in the case of the cross section (7), if we go over to an infinite upper limit in the integrals in these formulas, divergences occur in the integrals for $A(\varepsilon)$ and $a_k(\eta)$ which however cancel each other out.

The formulas (10)–(14) give a general solution of the problem. They are structural formulas which show, for example, that the terms in (2) depend on two constants, $A(\varepsilon)$ and $A'(\varepsilon)$, the values of a_{lk} do not depend on the cutoff angle χ_a, and so on. These formulas can be used to obtain exact or, bearing in mind the smallness of ε, approximate solutions.

3. RUTHERFORD AND MOTT CROSS SECTIONS

For $\kappa(x) = 1$ the integral in Eq. (12) is obtained in Appendix A and

$$a_{lk} = 2C_{lk}(S_l - S_k), \quad S_l = \sum_{k=1}^{l}\frac{1}{k}, \quad S_0 = 0,$$

$$A(\varepsilon) = \ln\frac{4+\varepsilon}{\varepsilon}, \quad \varepsilon A'(\varepsilon) = -\frac{4}{4+\varepsilon}.$$

For the Rutherford cross section (7), we obtain

$$Q_l^R = \chi_c^2 \times \sum_{k=1}^{l}\varepsilon^{k-1}kC_{lk}\left[\ln\frac{4+\varepsilon}{\varepsilon} - 2(S_l - S_k) - \frac{1}{k}\frac{4}{4+\varepsilon}\right]. \tag{20}$$

For the Mott cross section (8), taking into account (17) and (18), we find

$$Q_l^M = \left(1 + \frac{\varepsilon\beta^2}{4}\right)Q_l^R - \frac{\chi_c^2\beta^2}{4} \times\left\{\ln\frac{4+\varepsilon}{\varepsilon} - \sum_{k=0}^{l}\varepsilon^k C_{lk}\left[\ln\left(\frac{4+\varepsilon}{\varepsilon} - 2(S_l - S_k)\right)\right]\right\}. \tag{21}$$

Thus, the series (2), (20) is a solution of Eq. (1) for the cross section (7) and the series (2), (21) is a solution for the

cross section (8). The Goudsmit–Sanderson approximation used in [1] for the Rutherford cross section

$$Q_l^R \approx \chi_c^2 C_{l1}\left[\ln\frac{4}{\varepsilon} - 2(S_l - S_1) - 1\right]$$

is, with insignificant differences ($\varepsilon \ll 1$), the first term of the expansion in Eq. (20) which includes terms to the power zero of the factor ε^k while the approximation for the Mott cross section obtained in [1]

$$Q_l^M \approx \chi_c^2\left\{C_{l1}\left[\ln\frac{4}{\varepsilon} - 2(S_l - S_1) - 1\right] - \frac{1}{2}\beta^2 S_l\right\}$$

comprises terms to the power zero of the factor ε^k in Eq. (21).

These solutions can be used to check and refine the various approximations. It is also possible to compare the solutions of Eqs. (1) and (4).

For the cross section (7), Eq. (6) is reduced to the form

$$Q^R(\eta) = \frac{\chi_c^2}{\chi_a^2}\left\{1 + \zeta\frac{d}{d\zeta}\int_0^\infty \frac{J_0(x)x\,dx}{x^2 + \zeta^2}\right\}, \quad \zeta = \chi_a\eta.$$

The integral is equal to $K_0(\zeta)$ and allowing for $K_0'(\zeta) = -K_1(\zeta)$ we obtain

$$Q^R(\eta) = \frac{\chi_c^2}{\chi_a^2}\{1 - \xi_a\eta K_1(\chi_a\eta)\}. \tag{22}$$

Expressing K_1 as a series we obtain

$$Q^R(\eta) = \chi_c^2\sum_{k=1}^\infty \varepsilon^{k-1}kC_k(\eta) \times\left[\ln\frac{4}{\varepsilon} - 2(S(\eta) - S_k) - \frac{1}{k}\right], \tag{23}$$

$$S(\eta) = \ln\eta + C, \quad C_k(\eta) = \frac{1}{k!^2 4^k}\eta^{2k},$$

where C is the Euler constant and $C_k(\eta)$ are the coefficients in Eq. (19). Formula (20) yields (23) when

$$C_{lk} \longrightarrow C_k(\eta), \quad S_l \longrightarrow S(\eta), \quad \ln\frac{4+\varepsilon}{\varepsilon} \longrightarrow \ln\frac{4}{\varepsilon}, \quad \frac{4}{4+\varepsilon} \longrightarrow 1. \tag{24}$$

For the Mott cross section the integral (6) diverges unless the cross section is truncated at $\chi = 2$. A formula of the type (21) may be obtained using various approximations. In (18) we represent

$$\frac{1}{\chi^2 + \chi_a^2} = \left[\frac{1}{\chi^2 + \chi_a^2} - \frac{1}{\chi^2}\right] + \frac{1}{\chi^2}.$$

The first term here gives a converging integral given by

$$\chi_c^2\{\ln(4/\varepsilon) - 2K_0(\chi_a\eta) - 2\ln(\eta + C)\},$$

and the integral of the second is approximately [1]

$$2\chi_c^2\int_0^2 \frac{1 - J_0(\eta\chi)}{\chi^2}\chi\,d\chi \approx 2\chi_c^2\lim_{\epsilon\to 0}\left\{\int_\epsilon^2\frac{d\chi}{\chi} - \int_\epsilon^\infty \frac{J_0(\eta\chi)}{\chi}d\chi\right\} = 2\chi_c^2(\ln\eta + C).$$

Taking this into account we have

$$Q^M(\eta) \approx \left(1 + \frac{\varepsilon\beta^2}{4}\right)Q^R(\eta) - \frac{\chi_c^2\beta^2}{4}\left\{\ln\frac{4}{\varepsilon} - 2K_0(\chi_a\eta)\right\}.$$

Expressing K_0 as a series, we obtain

$$Q_l^M \approx \left(1 + \frac{\varepsilon\beta^2}{4}\right)Q_l^R - \frac{\chi_c^2\beta^2}{4} \times\left\{\ln\frac{4}{\varepsilon} - \sum_{k=0}^l \varepsilon^k C_k(\eta)\left[\ln\frac{4}{\varepsilon} - 2(S(\eta) - S_k)\right]\right\}. \tag{25}$$

Thus, using these approximations, as for the Rutherford cross section we obtain agreement with the solution (21) allowing for the formulas (24).

Using these formulas we can find solutions of Eqs. (1) and (4) for the cross sections $\sigma(\chi)$ which are rational functions of χ^2. For example, for $k = 1, 2$ we obtain

$$\frac{1}{(\chi^2 + \varepsilon)^k(\chi^2 + a)^m} = \frac{(-1)^{m+k-2}}{(m-1)!} \times\frac{\partial^{k-1}}{\partial\varepsilon^{k-1}}\frac{\partial^{m-1}}{\partial a^{m-1}}\frac{1}{a - \varepsilon}\left(\frac{1}{\chi^2 + \varepsilon} - \frac{1}{\chi^2 + a}\right),$$

and taking this into account we obtain solutions for the cross sections $\sigma_R(\chi)\kappa(\chi)$ and $\sigma_M(\chi)\kappa(\chi)$ where

$$\kappa(\chi) = \sum_m \frac{b_m}{(\chi^2 + a_m)^m}.$$

This type of factor truncates the cross section in the range $\chi^2 > a_m$ and this allows us to obtain a solution making approximate allowance for the nuclear form factor. It can be seen that solutions are in fact required for the cross sections

$$\sigma \sim \frac{1}{\chi^2 + \varepsilon} - \frac{1}{\chi^2 + a}.$$

For a series in terms of Legendre polynomials these are obtained using formula (17) where

$$a_{lk} = 2C_{lk}(S_l - S_k).$$

The possibilities of series for obtaining Q_l, $Q(\eta)$ are generally limited. For $Q(\eta)$ for example, asymptotic expansions are required for $\chi_a\eta \gg 1$. For Q_l with large l we need to use the substitution [1][1]

$$P_l(\cos(\delta)) \longrightarrow J_0(\eta\chi), \quad \eta = \sqrt{l(l+1)}$$

to go over to the integral

$$Q_l = Nt\int_0^2 \sigma(\chi)\chi d\chi\{1 - J_0(\eta\chi)\}$$

$$\longrightarrow \bar{n} - Nt\int_0^\infty J_0(\eta\chi)\sigma(\chi)\chi d\chi,$$

where $\bar{n} = Nt\sigma_t$ is the average number of collisions for a cross section truncated at $\chi = 2$. Thus, if summation in (2) is performed as far as $l \sim 1/\chi_a$, analytic results of the problem (4)–(6) must be used to obtain exact solutions.

In the following particular attention will be paid to the Rutherford cross section. Results of a comparison between the approximate and the exact solutions are discussed in Section 6.

4. SERIES IN POWERS OF 1/*B*. THE FUNCTIONS $f^{(n)}(X)$

The solution (5), (22) depends on the parameters χ_a^2 and χ_c^2/χ_a^2. The second of these is equal to the average number of collisions

$$\bar{n} = \chi_c^2/\chi_a^2,$$

since the total cross section is $\sigma_t = \sigma^2/\chi_a^2$. Subsequently we use the parameters λ and B. In addition to B we also use the clearer small parameter

$$æ = \chi_a^2/\lambda^2.$$

We define B and λ by means of [1, 3, 4]

$$B - \ln B = \ln\bar{n} - 2C + 1, \quad \lambda^2 = \bar{n}B\chi_a^2, \qquad (26)$$

where

$$\bar{n}æB = 1,$$

and in (5) and (22) we convert to the variables

$$u = \lambda\eta, \quad X = q/\lambda.$$

The series (23) will have the form (we shall subsequently drop the index R and we use the value of p as the argument)

$$Q(p) = \sum_{k=1}^{\infty} \frac{k}{k!^2} æ^{k-1} p^k \left\{1 - \frac{1}{B}\left(\ln p - 2S_k + \frac{k+1}{k}\right)\right\},$$

$$p = \frac{u^2}{4}. \qquad (27)$$

Using these transformations a transition is in fact made to the parameters λ, $\bar{n}$ [in accordance with (26) $B = B(\bar{n})$ and $æ = 1/\bar{n}B$]. The distribution function for the Rutherford cross section has the simple form

$$f(q) = \psi(X, \bar{n})/\lambda^2 \text{ or } f(q)qdq = \psi(X, \bar{n})XdX, \qquad (28)$$

where the function ψ only depends on λ in terms of the relative quantity $X = q/\lambda$.

We shall assume that the average number of collisions $\bar{n}$ is fairly high so that $1/B \ll 1$ and $æ = 1/\bar{n}B \ll 1$. Bearing this in mind we only retain the first term of the expansion in (27). Then

$$Q(p) \approx p\left(1 - \frac{\ln p}{B}\right), \qquad (29)$$

and expressing $\exp(-Q)$ and the solution (5) as a series in powers of $1/B$ we obtain [3, 4]

$$f(q) = \frac{1}{\lambda^2}\sum_{n=0}^{n_m} B^{-n} f^{(n)}(X), \quad X = q/\lambda, \qquad (30)$$

$$f^{(n)}(X) = \frac{2}{n!}\int_0^\infty e^{-p}[p\ln p]^n J_0(2\sqrt{p}X)dp. \qquad (31)$$

We note that the parameter λ is interpreted as the characteristic size of the region of multiple scattering (diffusion region) in which the particle distribution is close to Gaussian ($f^{(0)}(X) = 2\exp(-X^2)$) and most of the scattered particles are concentrated.

For the functions $f^{(n)}(X)$ we find representations in the form of series and their asymptotic expansions. In the intermediate formulas we make the substitution

$$f^{(n)}(X) \longrightarrow f^{(n)}(z), \quad z = X^2.$$

Substituting into (31) the expansion (19), we obtain the auxiliary series

$$f^{(n)} = 2\sum_{k=0}^{\infty}(-1)^k \frac{(k+1)\ldots(k+n)}{k!} h_n(k+n+1)z^k,$$

[1] This can be justified in the semiclassical approximation if the initial equation for P_l is written in terms of the variable χ. This equation has the form

$$\frac{1}{\chi}\frac{d}{d\chi}\chi\left(1 - \frac{\chi^2}{4}\right)\frac{d}{d\chi}P_l = -l(l+1)P_l$$

and in the semiclassical approximation gives $P_l(\cos\delta) \approx J_0(\eta\chi)$. With this relationship there is no need to use the small-angle approximation for the cross section when reducing the series (2) to the integral (5) or when deriving Eq. (4) (see [1] on this topic). The cross section remains exact which in our view ensures that the solution (5), (6) is valid over the entire range of angles.

$$h_n(\nu+1) = \frac{1}{n!\Gamma(\nu+1)}\int_0^\infty e^{-p}p^\nu(\ln p)^n dp$$

$$= \frac{1}{n!\Gamma(\nu+1)}\frac{d^n}{d\nu^n}\Gamma(\nu+1).$$

The formulas for h_n are given in Appendix B.

The series is only suitable in calculations as far as $X \sim 3$. Using the relationship

$$h_n(\nu+1) = h_{n-1}(\nu)/\nu + h_n(\nu),$$

obtained from the nth-order derivative of $\Gamma(\nu+1) = \nu\Gamma(\nu)$, we transform this to give

$$f^{(n)}(z) = 2\sum_{k=0}^{\infty}(-1)^k\frac{(k+1)\ldots(k+n-1)}{k!}h_{n-1}(k+n)z^k$$

$$+ 2n!h_n(n) - 2\sum_{k=0}^{\infty}(-1)^k\frac{(k+1)\ldots(k+n)}{k!}$$

$$\times\left(\frac{z^{k+1}}{k+1} + n\frac{z^{k+1}}{(k+1)^2}\right)h_n(k+n+1).$$

It is easy to see that we have written the relationship

$$f^{(n)}(z) = f^{(n-1)}(z) + 2n!h_n(n)$$

$$-\int_0^z f^{(n)}(\eta)d\eta - n\int_0^z\frac{d\eta}{\eta}\int_0^\eta f^{(n)}(\xi)d\xi.$$

Allowing for

$$\frac{1}{2}\int_0^\infty f^{(n)}(\xi)d\xi = \delta_{0n}$$

(see Appendix C) for the function

$$F_n = \frac{1}{2}\int_z^\infty f^{(n)}(\xi)d\xi$$

we obtain the equation

$$F_n'' + F_n' + \frac{n}{2}F_n = F_{n-1}''$$

with the initial conditions

$$F_n(0) = 0, \quad F_n'(0) = -n!h_n(n+1).$$

This allows us to systematically seek functions $F_n(z)$, $f^{(n)}(z)$ having increasingly higher values of n (see Appendix D).

A similar approach may be applied to obtain functions of a more general form with $m > n$

$$f_m^{(n)}(X) = \frac{2}{n!}\int_0^\infty e^{-p}p^m(\ln p)^n J_0(2\sqrt{p}X)dp.$$

The need for these functions arises when constructing approximate solutions. The formulas for these are in fact the same as those in the case $m = n$. The equation for

$$F_{nm}(z) = \frac{1}{2}\int_z^\infty f_m^{(n)}(\xi)d\xi,$$

for example, has the form

$$F_{nm}'' + F_{nm}' + \frac{m}{z}F_{nm} = F_{n-1,m-1}''.$$

For small m we can use the following relationship to obtain $f_m^{(n)}(X)$

$$f_{m+1}^{(n)}(z) = -\frac{d}{dz}z\frac{d}{dz}f_m^{(n)}(z).$$

For $m = 0, 1, 2$ we then require the functions

$$f_m^{(0)}(z) = 2m!e^{-z}L_m(z),$$

$$L_0(z) = 1, \quad L_1(z) = 1-z, \quad L_2(z) = 1-2z+z^2/2.$$

For the functions F_n we can obtain an analytic recurrence formula. We write

$$F_n(z) = n!h_n(n+1)e^{-z}L_n^{-1}(z) + \Delta(z),$$

where L_n^{-1} is the Laguerre polynomial L_n^j for $j = -1$ and the first term is the solution of the homogeneous equation for F_n. For $\Delta(z)$ we obtain an equation with zero initial conditions and by integrating this consistently k times between the limits 0 and z, we obtain

$$z\Delta_k'' + (z-k)\Delta_k' + (n-k)\Delta_k = s_k,$$

where Δ_k, s_k are k-fold integrals of Δ in the limits between 0 and z, and $s = zF_{n-1}''$. Solving this equation for $k = n$, we obtain

$$F_n(z) = n!h_n(n+1)e^{-z}L_n^{-1}(z)$$

$$+\frac{d^{n-1}}{dz^{n-1}}\left\{z^n e^{-z}\int_0^z e^t t^{-n-1}s_n(t)dt\right\}.$$

Thus, the function F_n can be expressed in terms of specific integrals of F_{n-1}. In the case $n = 1$, for example,

$$L_1^{-1}(z) = -z, \quad s_1 = ze^{-z}, \quad h_1(2) = 1-C,$$

$$F_1(z) = e^{-z} - 1 + ze^{-z}[\mathrm{Ei}(z) - \ln z],$$

and for $f^{(1)}$ we obtain the well-known formula [3, 4] (we go over to the argument X):

$$f^{(1)}(X) = 2e^{-z}(z-1)[\mathrm{Ei}(z) - \ln z] - 2(1-2e^{-z}).$$

The series obtained from the equation for F_n (see Appendix D)

$$F_n(X) = n!e^{-z}\sum_{k=1}^{\infty}\frac{a_k}{k!}z^k,$$

$$f^{(n)}(X) = 2n!e^{-z}\left[-a_1 + \sum_{k=1}^{\infty}(a_k - a_{k+1})\frac{z^k}{k!}\right]$$

can be applied[2] as far as $z \sim 100$ and are matched with the asymptotic series

$$F_n(X) = \sum_{k=0}^{\infty}\frac{(k+n)!}{k!z^{k+n}}\sum_{j=0}^{n-1}\alpha_{n-j}(k+n+1)\frac{\ln^j z}{j!},$$

$$f^{(n)}(X) = 2\sum_{k=0}^{\infty}\frac{(k+n)!}{k!z^{k+n+1}}\sum_{j=0}^{n-1}\beta_{n-j}(k+n+1)\frac{\ln^j z}{j!}.$$

The functions $f^{(n)}(X)$ are plotted in Fig. 1. For $n \geq 2$ the properties of these functions vary slowly as the number n increases. The functions $f^{(0)}$ and $f^{(1)}$ are distinguished by the fact that $f^{(0)}$ determines the behavior of the distribution function for small angles and $f^{(1)}$ determines the behavior for large angles.

For large n we can obtain fairly exact approximations for these functions. The Fourier transforms of these functions

$$f_n(p) = \frac{1}{n!}e^{-p}[p\ln p]^n \tag{32}$$

have maxima for $p > 1$ (see Fig. 2). The region of this maximum for fairly large n makes the main contribution to the Fourier integral. We write approximately

$$f_n(p) \approx f_{n,\max}[(1-\beta)g_m(p) + \beta g_{m+1}(p)],$$
$$g_m(p) = t^m e^{-m(t-1)}, \quad t = p/p_{n0}, \tag{33}$$

where $f_{n,\max}$ is the value of f_n at the maximum and the maximum point p_{n0} is determined by solving the transcendental equation

$$p_{n0} = n(1 + 1/\ln p_{n0}). \tag{34}$$

In formula (33) the maximum points of the functions on the left and right and their values at the maxima are the same. The condition for matching of the second derivatives for $p = p_{n0}$ has the form

$$\mu = m + \beta = n(1 + \ln^{-1}p_{n0} + \ln^{-2}p_{n0}).$$

[2] In calculations of the terms of the series, for example for F_n using the formula

$$\exp\{ln|a_k| - \ln k! + k\ln z - z + \ln n!\}.$$

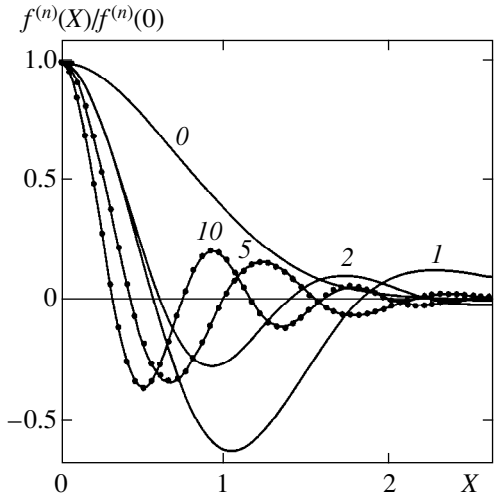


Fig.1. The functions $f^{(n)}(X)$. The numbers on the curves give the values of n.

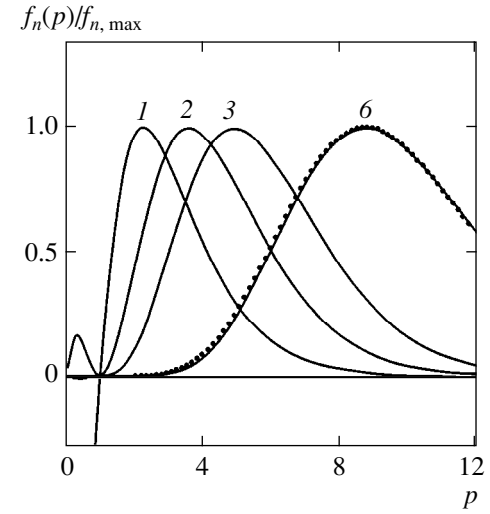


Fig. 2. The functions $f_n(p)$. The numbers on the curves give the values of n.

The values of m and β are then determined by the condition

$$m < \mu < m + 1.$$

Taking this into account we have

$$f^{(n)}(X) \approx f_{n,\max}\{a_1 e^{-\alpha_1 z}L_m(\alpha_1 z) + a_2 e^{-\alpha_2 z}L_{m+1}(\alpha_2 z)\},$$

$$z = X^2, \quad \alpha_1 = \frac{p_{n0}}{m}, \quad \alpha_2 = \frac{p_{n0}}{m+1}, \tag{35}$$

$$a_1 = \frac{2(1-\beta)m!\alpha_1}{e^{-m}m^m}, \quad a_2 = \frac{2\beta(m+1)!\alpha_2}{e^{-m-1}(m+1)^{m+1}}.$$

The parameters a_1, a_2, α_1, and α_2 depend on n, the index n being omitted for conciseness. The approximate val-

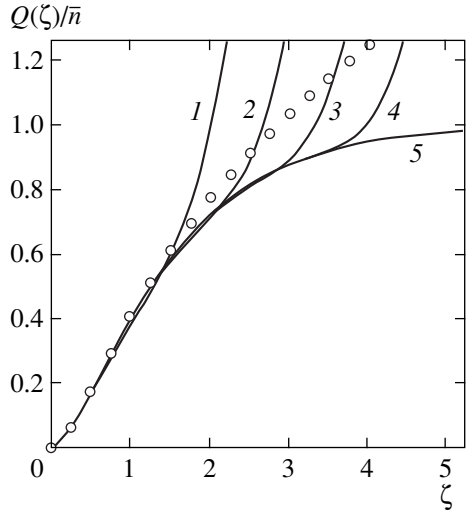


Fig. 3. Various approximations for $Q(\zeta)$, $\zeta = \chi_a\eta$, for $\bar{n} = 5$: (*1–4*) formula (36) for $m = 1$–4; (*5*) formula (22); symbols—values of $\tilde{Q}$ using formulas (40), (22).

ues of the functions f_n and $f^{(n)}$ are shown by the symbols in Figs. 1 and 2.

To conclude this section we show that the series (30) diverges. We take $q = 0$ when

$$f^{(n)}(0) = \frac{2}{n!}\frac{d^n}{d\nu^n}\Gamma(\nu+1)\bigg|_{\nu=n}.$$

For fairly large ν we have

$$\Gamma'(\nu+1) = \Gamma(\nu+1)\psi(\nu+1) \approx \Gamma(\nu+1)\ln\nu.$$

Retaining in each differentiation the largest term in the factor at $\Gamma(\nu+1)$ we obtain

$$(d^n/d\nu^n)\Gamma(\nu+1) \sim \Gamma(\nu+1)\ln^n\nu,$$

i.e., $f^{(n)}(0) \sim 2\ln^n n$, the general term in the expansion in (30) is proportional to $[\ln n/B]^n$ and the series clearly diverges.

Thus, we either need to allow for a finite (which?) number of terms in the expansion or find a converging series in powers of $1/B$. For this we require convergence of the initial integral (5) which diverges with the approximation (29).

5. TRUNCATION OF INTEGRALS AND SERIES. CONVERGING SERIES IN POWERS OF $1/B$

In formula (27) the expression in braces becomes negative for fairly large p so that when allowance is made for m terms of the expansion we have

$$\lim_{p\to\infty} Q(p) \longrightarrow -\infty$$

and the integral (5) diverges. We take into account the next $(m+1)$th term of the expansion and, assuming that $1/B$ is small, we drop the term $\sim 1/B$. We obtain

$$Q(p) \approx \sum_{k=1}^{m} \ldots + æ^m \frac{m+1}{(m+1)!^2} p^{m+1}, \qquad (36)$$

where the sign of the sum denotes the series (27) truncated at the mth term. This improves the accuracy of the approximation in the range of typical values $p \sim 1$ ($\eta \sim 1/\lambda$) and also ensures convergence of the integral (5).

A similar procedure involves combining the terms in (27) having the same total power of the products of the small quantities æ and $1/B$, i.e., terms proportional to $æ^i$ and $æ^{i-1}/B$, and then allowing for a finite number of terms of the expansion. The terms of this expansion tend to infinity as p increases, i.e., high values of u are truncated (Fig. 3) and the integral converges. The approximation with $m = 0$, for example, corresponds to the first term of this expansion (with $i = 0$) equal to p and describes a Gaussian distribution. For Q_l a similar approximation is obtained by going over to the initial variables ε and η in (36) and using the correspondence formulas (24). We obtain the truncated series (20):

$$Q_l \approx \chi_c^2 \left\{ \sum_{k=1}^{m} \ldots + (m+1) C_{l,m+1} B\varepsilon^m \right\}. \qquad (37)$$

For $m > l$ the series is in fact truncated at $m = l$ whereas for $m > l$ we have $C_{lm} = 0$, i.e., Q_l has exact values for $l \le m$. Thus, for $m \ge 1$ the normalization and the value of $\overline{q^2}$ which are determined by the terms with $l = 0, 1$ in (2) will be exact.

The question of interpreting these approximations then arises since they neglect nonscattered particles (it is in fact assumed that these particles make a negligible contribution to the solution).

In this context it is useful to note that equations of the type (1), (4) are obeyed by a certain class of systems in which transitions between states form a Markov circuit. Let us assume that the system is exposed to a certain random sequence of actions and $\rho_n(\lambda)d\lambda$ is the probability of observing the system in the range of states $d\lambda$ after the nth action. We then have

$$\rho_n(\lambda) = \int W(\lambda, \lambda')\rho_{n-1}(\lambda')d\lambda'$$

(Smoluchowski equation), where $W(\lambda, \lambda')d\lambda$ is the probability of a transition into the range $d\lambda$ from the initial state λ' as a result of a single action. The probability of n events taking place by the time t (or over the path t) obeys a Poisson distribution with the average number

of events $\bar{n}$ over time t. After this time has elapsed, the average distribution function is given by

$$f(\lambda, \bar{n}) = \sum_{n=0}^{\infty} e^{-\bar{n}} \frac{\bar{n}^n}{n!} \rho_n(\lambda).$$

Differentiating this relationship with respect to $\bar{n}$, we obtain after elementary calculations

$$\frac{\partial f(\lambda, \bar{n})}{\partial \bar{n}} = -f(\lambda, \bar{n}) + \int W(\lambda, \lambda') f(\lambda', \bar{n}) d\lambda'.$$

This equation is of the same type as (1) and (4) (in the equations it is sufficient to convert from the path covered t and the cross section σ to $\bar{n} = Nt\sigma_t$ and $W = \sigma/\sigma_t$). It can be seen that the solutions of these equations are averages over the Poisson distribution and for a given value of $\bar{n}$ we need the expression

$$f(q, t) = e^{-\bar{n}} f_\delta(q) + (1 - e^{-\bar{n}}) f_s(q, t),$$

where $f_\delta(q) = \delta(q)/q$ and $f_s(q, t)$ are the distribution functions of the nonscattered particles with $n = 0$ and scattered particles with $n \geq 1$, and $1 - e^{-\bar{n}}$ is the scattering probability.

The nonscattered particles can be isolated in the solutions (2), (5). For example, in (22) for $\zeta \gg 1$ we have

$$K_1(\zeta) \approx \sqrt{\pi/2\zeta} \exp(-\zeta) \quad \text{and} \quad \lim_{\eta \to \infty} Q(\eta) = \bar{n}.$$

In general, if the cross section contains no δ-type singularities, its Fourier transform for $\eta \longrightarrow \infty$ tends to zero, i.e., formulas (3) and (6) yield the limit

$$\lim_{\eta, l \to \infty} Q = \bar{n}.$$

This limit corresponds to nonscattered particles which can be isolated using the substitution

$$e^{-Q} \longrightarrow e^{-Q} - e^{-\bar{n}}$$

and taking into account the term $e^{-\bar{n}} f_\delta(q)$ in the solution. The formulas obtained for the function f_s, unlike the solutions (2), (3) and (5), (6) are correct in practical calculations since they contain no δ-type singularity.

For scattered particles we have

$$f_s(\cos\vartheta, t) = \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l(\cos\vartheta) \exp(-\tilde{Q}_l(t)), \quad (38)$$

$$f_s(q, t) = \int_0^{\infty} \eta d\eta J_0(\eta q) \exp(-\tilde{Q}(\eta, t)), \quad (39)$$

$$e^{-\tilde{Q}} = \frac{e^{-Q} - e^{-\bar{n}}}{1 - e^{-\bar{n}}}. \quad (40)$$

Note that for $\bar{n} \ll 1$ when $Q \ll 1$ and $e^{-\tilde{Q}} \approx 1 - Q/\bar{n}$ allowing for (3) and (6) we then obtain the same fairly understandable result for Eqs. (1) and (4):

$$f_s(q, t) \approx \sigma(q)/\sigma_t.$$

For $q \gg l$ when small η, $l \ll 1/\lambda$, and $Q \ll 1$ are important in the Fourier expansion, we thus obtain [neglecting the terms $\sim\delta(q)$] which are not required here]

$$f_s(q, t) \approx \frac{\bar{n}\sigma(q)}{\sigma_t(1 - e^{-\bar{n}})}.$$

In the single region this gives

$$f(q, t) \approx \frac{\bar{n}\sigma(q)}{\sigma_t}.$$

We shall now return to the approximations (36), (37). The factor $e^{-\tilde{Q}}$ in (38), (39) truncates high values of η, l ($e^{-\tilde{Q}} \longrightarrow 0$ for $\eta, l \longrightarrow \infty$) and in this sense the approximations (36), (37) may be used for scattered particles, i.e., we can assume $\tilde{Q} \approx Q$. It is useful to make the refinement

$$\tilde{Q} \approx \frac{Q}{1 - e^{-\bar{n}}}. \quad (41)$$

This relationship valid for small η, $l \ll 1/\lambda$ when $\tilde{Q}$, $Q \ll 1$ refines the large-angle distribution function if this is taken into account to determine the values of B and λ. For this we assume

$$Q = \gamma Q_s, \quad B = \frac{1}{\bar{n}_s æ},$$
$$\bar{n}_s = \frac{\bar{n}}{\gamma}, \quad \gamma = 1 - e^{-\bar{n}}. \quad (42)$$

The series for Q_s has the form

$$Q_s(p) = \sum_{k=1}^{\infty} \frac{k}{k!^2} æ^{k-1} p^k$$
$$= \left\{1 - a_s - \frac{1}{B}\left(\ln p - 2S_k + \frac{k+1}{k}\right)\right\}, \quad (43)$$

$$a_s = \frac{1}{B}(B + 2C + \ln æ - 1). \quad (44)$$

We determine the values of B and λ using the relationships

$$B - \ln B = \ln \bar{n}_s - 2C + 1,$$
$$\lambda^2 = \bar{n}_s B \chi_a^2. \quad (45)$$

We then have $a_s = 0$ and the series for Q_s is the same as (27).

Table 1. Rutherford cross section for the scattering of π-mesons on gold. E_{kin} = 50 MeV, $\chi_a = 2.77 \times 10^{-4}$

No.	ρt, g/cm^2	λ	B	$\bar{n}$	$\xi(0)$, %	ξ_{max}, %	$\xi(2)$, %	$\xi_{culc}(2)$, %
1	0.00375	3.51×10^{-3}	4.93	32.7	−0.096	0.61	0.002	0.002
2	0.015	8.14×10^{-3}	6.61	131	0.003	0.13	0.007	0.008
3	0.060	1.81×10^{-2}	8.21	523	−0.002	0.05	0.030	0.032
4	0.24	3.96×10^{-2}	9.77	2.09×10^3	−0.013	0.11	0.12	0.13
5	0.4	5.26×10^{-2}	10.34	3.49×10^3	−0.024	0.20	0.20	0.22
6	2	0.127	12.11	1.74×10^4	−0.14	1.28	1.01	1.07
7	4	0.185	12.86	3.49×10^4	−0.30	2.88	2.07	2.15
8	10	0.304	13.85	8.71×10^4	−0.80	8.46	5.59	5.37
9	40	0.640	15.34	3.49×10^5	–3.48	50.4	39.1	21.5

Finally, we have the approximation $\tilde{Q} \approx Q_s$ which is more accurate for small η and approximately truncates high values of η.

For $m = 1$ in (36) we obtain instead of (29)

$$\tilde{Q}(p) \approx p\left(1 - \frac{\ln p}{B}\right) + \frac{1}{2}æ p^2.$$

As previously, we can expand in powers of $1/B$ and obtain the series (30) with the functions $\tilde{f}^{(n)}(æ, X)$ which are determined by formula (31) with a truncating factor after the integral equal to $\exp(-æp^2/2)$. The normalization of these functions is the same as in the case $æ = 0$ (see Appendix C). However, these functions depend on two variables and are inconvenient for applications.

We write the expansion in the form

$$e^{-\tilde{Q}(p)} = \sum_{n=0}^{n_m} \frac{1}{n!} e^{-\alpha p} v^n(p),$$

$$v(p) = -ap(p-1) + \frac{p \ln p}{B},$$

$$a = æ/2, \quad \alpha = 1 + a.$$

Using this representation for $p \sim 1$ we can set

$$p - 1 \approx \ln p,$$

and thereby arrive at the approximation for $n \geq 2$:

$$\frac{1}{n!} e^{-\alpha p} v^n(p) \approx B^{-n} a_n f_n(c_n p), \quad c_n = \frac{p_{n0}}{p_n},$$

where $c_n \approx 1$, p_{n0} is the maximum point of the function $f_n(p)$ [see (32), (34)], and p_n is the maximum point of the nth term of the expansion. The functions on the left and right have maxima for $p = p_n$. We determine the value of a_n from the condition that their maxima agree. Converting to the inverse transform we obtain

$$f_s(q) = \frac{1}{\lambda^2}\left\{ a_{01} f_1^{(0)}(k_0 X) + a_{02} f_2^{(0)}(k_0 X) + \sum_{n=0}^{n_m} B^{-n} a_n c_n^{-1} f^{(n)}(X/\sqrt{c_n}) \right\},$$

$$a_{01} = \frac{a - B^{-1} \ln \alpha}{\alpha^2}, \quad a_{02} = -\frac{a}{\alpha^3}, \quad k_0 = \frac{1}{\sqrt{\alpha}}, \tag{46}$$

$$a_0 = 1, \quad a_1 = \frac{1}{\alpha}, \quad c_0 = c_1 = \alpha.$$

The functions $f_m^{(0)}$ for $m = 1, 2$ are given in the previous section and are essentially elementary. In the limits of validity of this approximate series it is sufficient to take $n_m = 15$, using the approximation (33), (35) for $n > 5$. The normalization of the function $f_s(q)$ is determined by the term with $n = 0$ and is 1. The asymptotic behavior is determined by the term with $n = 1$. For $q \gg \lambda$ we have

$$f_s(q) \approx \frac{1}{1 - e^{-\bar{n}}} \frac{2\chi_c^2}{q^4}, \quad f(q) \approx \frac{2\chi_c^2}{q^4}.$$

The factor $1/(1 - e^{-\bar{n}})$ appears here because of the refinements (41)–(45) made above.

6. DISCUSSION

Using the results presented in Section 3, we can compare the solutions (38), (39) of the kinetic Eqs. (1) and (4). For the Rutherford cross section (7) such a comparison is made in Fig. 4 for the data from Table 1 taken from [1]. Some details of the calculations are described in [6]. In the range of q values where the characteristic dimension can be considered to be λ, the

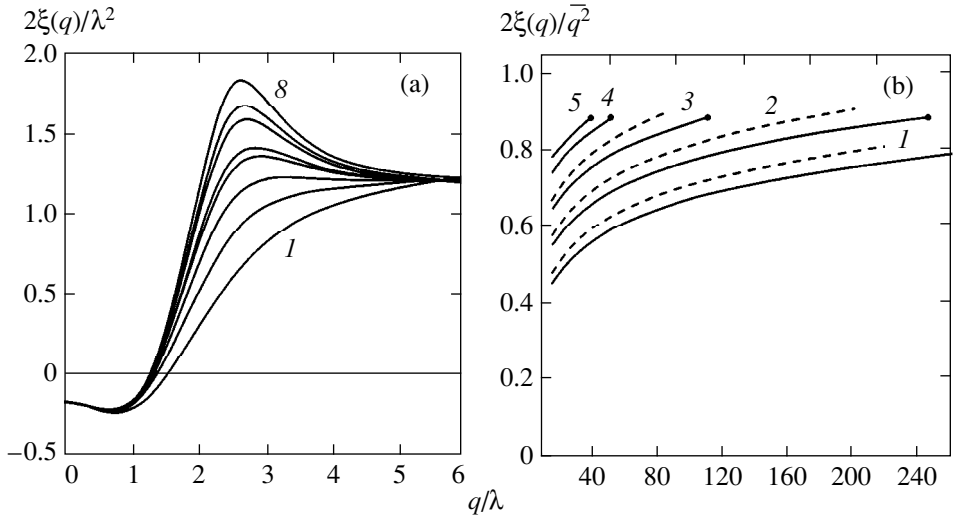


Fig. 4. The dependences $\xi = \xi(q)$, the relative error of the solution (39), (40), (22) compared with (38), (40), (20). The numbers on the curves correspond to the number of the variant in Table 1. The symbols on the curves in Fig. 4b correspond to $q = 2$, the dashed curves give the estimate $\xi_{\text{culc}}(q)$ (Section 6) for variants 1–3 in Table 1.

estimates of the relative error $\xi(q)$ of the solution of Eq. (4) obtained in [1] are fairly accurate. The error arises as a result of a transition from the spherical region of integration in Eq. (1) to a planar region of integration in Eq. (2) (see [1]). This region of q/λ values is smaller than or of the order of a few units (regions of plural and multiple scattering, Fig. 4a). Here we have

$$\xi(q) \sim \begin{cases} -0.3\lambda^2/2, & q < q_0, \\ 0, & q \approx q_0 = \sqrt{2}\lambda, \\ \lambda^2/2, & q > q_0. \end{cases}$$

In the region $q \gg \lambda$ the error is estimated using the formula

$$\xi(q) \sim \frac{a^2}{4\sigma(q)} \frac{1}{q} \frac{\partial}{\partial q} \frac{q^3}{4} \frac{\partial}{\partial q} \sigma(q),$$

where

$$a^2 \sim \overline{q^2} = Nt \int_0^2 \chi^2 \sigma(\chi) \chi d\chi.$$

In order to obtain this estimate in [1] we considered two important regions of integration: the vicinity of the cross section maximum centered at the point **n** and the vicinity of the distribution function maximum centered at the point **i** (Fig. 1 in [1]) and we integrated over these regions, integration being performed over the entire spherical surface in both cases. With a more accurate approach the limits of integration must be constrained since integration over Ω' should be single. It can be seen from Fig. 1 in [1] that when integration is performed near the point **n** ($d\Omega' = \chi d\chi d\varphi$) and near **i** ($d\Omega' = q' dq' d\varphi$), the limits in terms of the variables χ, q' can be taken for the estimates to be

$$\chi_m = \chi_m(q) = 2\sin(\vartheta/4),$$

which corresponds to the vector $\mathbf{n}'$ directed along the bisector of the angle ϑ. We then have

$$a^2(q) = Nt \int_0^{\chi_m} \chi^2 \sigma(\chi) \chi d\chi.$$

Taking this into account, we obtain a more accurate estimate for the Rutherford cross section for large angles, as can be seen from Fig. 4b (dashed curves)

$$\xi_{\text{culc}}(q) = a^2(q)/2.$$

In practical cases the corrections to the Rutherford cross section (Mott cross section, nuclear form factor) are only significant in the single region so that in the range of q/λ values shown in Fig. 4a the main influence remains for the Rutherford part of the cross section and the error estimates are as before. In the single region the error of the solution is estimated in accordance with the dependence $\sigma = \sigma(\chi)$. For the Mott cross section we obtain

$$\xi_{\text{culc}}(q) = \frac{1}{2} \frac{a^2(q)}{1 - \beta^2 q^2/4}.$$

Table 1 gives values of $\xi(q)$ for the series (46) compared with (38), (30), and (20) ($\xi_{\max}$ is the maximum error in the range $0 < q < 6$). For approximately $\lambda > 4 \times 10^{-2}$ the error of the series is associated with the approximate nature of Eq. (4) and is consistent with Fig. 4 whereas for lower values it is associated with the error of the

Table 2. Mott cross section for the scattering of electrons on aluminum. E_{kin} = 15 MeV, $\chi_a = 6.89 \times 10^{-4}$

No.	ρt, g/cm^2	λ	B	$\bar{n}$	$\xi(0)$, %	ξ_{max}, %	$\xi(q_e)$, %	$\xi_{culc}(q_e)$, %
1	0.00375	9.10×10^{-3}	5.01	34.8	–0.079	0.56	0.20	0.17
2	0.015	2.10×10^{-2}	6.68	139	–0.002	0.12	0.81	0.66
3	0.030	3.15×10^{-2}	7.49	278	–0.01	0.08	1.59	1.32
4	0.06	4.68×10^{-2}	8.28	556	–0.02	0.11	3.08	2.64
5	0.12	6.93×10^{-2}	9.06	1.11×10^3	–0.05	0.26	5.95	5.28
6	0.24	0.102	9.84	2.23×10^3	–0.10	0.63	11.3	10.6
7	0.4	0.135	10.41	3.71×10^3	–0.18	1.24	18.0	17.6
8	1	0.224	11.42	9.28×10^3	–0.47	5.23	39.1	44.0
9	2	0.328	12.17	1.86×10^4	–0.99	67.3	66.5	76.0

Table 3. Relative error $\xi(q)$ of the series in terms of Legendre polynomials (38), (41), (37) with different values of m as a function of the average number of collisions $\bar{n}$. $\lambda = 10^{-2}$. On the right-hand side of the table the first column gives the data for series (46) and the second column gives the data for the series (30) with $n_m = 2$

$\bar{n}$	$æ \times 10^2$	B	m	$\xi(0)$	ξ_{max}	$\xi(0)$	ξ_{max}, %
130	0.117	6.60	1	3.0×10^{-3}	3.2×10^{-2}	0.001	0.24
						–1.4	2.8
30	0.692	4.82	1	–0.12	–0.24	–0.13	0.72
			2	-2.0×10^{-4}	1.4×10^{-3}	–3.3	6.3
20	1.16	4.30	1	–0.48	–0.48	–0.50	1.0
			2	-1.4×10^{-2}	-1.4×10^{-2}	–4.6	8.3
15	1.70	3.92	1	–1.3	–1.3	–1.3	1.4
			2	–0.12	–0.12	–5.9	10.4
			3	-2.4×10^{-2}	-2.4×10^{-2}		
10	2.98	3.36	1	–4.7	–4.7	–4.7	–4.7
			2	–1.4	–1.4	–9.7	14.7
			3	–0.44	–0.44		
			5	0.29	0.29		
			7	0.80	0.80		
5	8.67	2.29	1	–14.1	–14.1	–13.5	–13.9
			2	–2.3	7.2	–14.6	–37.5
			3	7.0	8.4		
			4	16.1	16.1		

approximation (36) and the series (46) derived from it for $m = 1$.

Data for the Mott cross section [1] allowing for (21) are given in Table 2 where q_e is the value of q for $\vartheta = 150°$. It can be seen that the estimates for $q \gg \lambda$ are fairly reliable. We also note that $\xi \sim 10^{-3}$% for curve *1* in Fig. 4, i.e., any further analysis (for lower values of λ) has no practical meaning and the topic can be considered to be exhausted.

Data for the approximation (37) are given in Table 3. In view of the representation (28), it is sufficient to find the error as a function of the average number of collisions $\bar{n}$. Thus, the value of λ is fixed and all the parameters required in the problem are defined in terms of $\bar{n}$, λ. As m increases, the accuracy of the approximation increases but for low values of $\bar{n}$ it then begins to decrease (variant with $\bar{n} = 10$). It should be noted that the value of m should not be too large to describe particle scattering since in the limit $m \longrightarrow \infty$ the approximation (37) leads to a δ-type singularity in the solution. The associated distortion of the solution in the small angle range will be larger, the smaller $\bar{n}$. For $\bar{n} = 5$ the accuracy does not improve with increasing m. Conse-

quently, the approximation (37), like the approximation (36) together with the series (46) derived from this for $m = 1$ has a natural limit of validity at $\bar{n} \approx 10$. We stress that the parameter æ is still fairly small and this limit is not determined by any physical constraints but is associated with the fact that Eqs. (41), (36), and (37) are an approximation for Q/γ and not for $\tilde{Q}$ as can be seen from Fig. 3 and for $\bar{n} = 5$ the approximation becomes rough in the range $\eta > 1/\chi_a$.

Results for the series (30) and (46) are given on the right-hand side of Table 3. Assuming that $\xi \sim 5\%$ is the limiting value, the series (30) with $n_m = 2$ is suitable as far as $\bar{n} \approx 100$ [here by selecting n_m for each value of $\bar{n}$, we can improve the accuracy for $\xi(0)$ but not for $\xi_{\max}$] and the series (46) is suitable as far as $\bar{n} \approx 10$.

The approximation (37) is not much more complex than the Goudsmit–Sanderson solution in the same way that the series (46) is not much more complex than (30). These approximations are valid for $\bar{n} \sim 10$–1000 which corresponds to a matter density ~1–10 mg/cm^2. For lower values of $\bar{n}$ we need a more accurate approximation for the function $f_s(q)$ than (46).

7. CONCLUSIONS

In the first part of this study we obtained structural formulas for the solutions for the cross sections taking into account the cutoff angle χ_a

$$\sigma(\chi) = \sigma_R(\chi)\kappa(\chi), \quad \sigma(\chi) = \sigma_M(\chi)\kappa(\chi),$$

where $\chi = 2\sin(\delta/2)$, δ is the scattering angle, and the factor $\kappa(\chi)$ allows for differences between the real cross sections and the Rutherford σ_R and Mott σ_M cross sections. This means that both exact and approximate solutions can be obtained.

For the Rutherford and Mott cross sections we obtained solutions as a series in terms of Legendre polynomials which can be used to obtain solutions which are rational functions of χ^2.

These solutions were compared with the solutions of the approximate kinetic equation obtained in [1], without using any additional approximations for the Rutherford cross section. Taking these results into account we determined the limits of validity of this equation for which estimates were given in [1].

The results of the second part relate to the solution of fundamental problems: obtaining a converging series in powers of $1/B$ and functions of this series $f^{(n)}$. We confirmed that even when a considerable number of expansion terms are summed, a series in powers of $1/B$ is suitable as far as the minimum average number of collisions $\bar{n} \sim 10$. In order to obtain a converging series in powers of $1/B$ and a truncated series in terms of Legendre polynomials we need to allow for another term of the expansion $\sim\chi_a^2$ compared with [3, 4].

APPENDIX A

We shall consider the integral

$$I_k = \int_{-1}^{1} \frac{\gamma_k(x)}{(1-x)^{k+1}} dx$$

with a certain initial function $\gamma_k(x)$ [the index l is omitted from $\gamma_{lk}(\chi)$ for conciseness] which satisfies the following conditions in accordance with the definition (13)

$$\gamma_k(x) = O[(1-x)^{k+1}], \tag{A.1}$$

$$\gamma_k(x) = \sum_{i=0}^{l} A_i P_i(x), \tag{A.2}$$

$$A_i = \begin{cases} C_{lk}/C_{kk}, & i = k, \\ 0 \quad i = k+1, \ldots, l-1, \\ -1 \quad i = l. \end{cases} \tag{A.3}$$

Formula (A.3) is derived from the fact that in the expansion of χ^{2k} in terms of Legendre polynomials P_l the highest value of the index l is equal to k and the expansion coefficient for P_k is equal to $(-1)^k/C_{kk}$, as follows from (14).

Integrating by parts we obtain

$$\frac{\gamma_k(x)}{k(1-x)^k}\bigg|_{-1}^{1} - \frac{1}{2k}\int_{-1}^{1} \frac{2\gamma_k'(x)}{(1-x)^k} dx.$$

The first term, in accordance with (A.1) is given by

$$-\frac{1}{2^k k}\gamma_k(-1). \tag{A.4}$$

In the integral we replace

$$\frac{2}{(1-x)^k} = \frac{1-x^2}{(1-x)^{k+1}} + \frac{1}{(1-x)^{k-1}},$$

and after integrating by parts we obtain

$$-\frac{(1-x^2)\gamma_k'(x)}{2k^2(1-x)^k}\bigg|_{-1}^{1} + \frac{1}{2k^2}\int_{-1}^{1} \frac{[(1-x^2)\gamma_k'(x)]'}{(1-x)^k} dx$$

$$-\frac{\gamma_k(x)}{2k(1-x)^{k-1}}\bigg|_{-1}^{1} + \frac{k-1}{2k}\int_{-1}^{1} \frac{\gamma_k(x)}{(1-x)^k} dx.$$

The first term is zero since

$$\gamma_k'(x) = O[(1-x)^k]. \tag{A.5}$$

The third term is reduced with (A.4). Finally we arrive at the formulas:

$$I_k = \frac{1}{2k^2} I_{k-1}, \quad I_{k-1} = \int_{-1}^{1} \frac{\gamma_{k-1}(x)}{(1-x)^k} dx,$$

$$\gamma_{k-1}(x) = O[(1-x)^k],$$

$$\gamma_{k-1}(x) = -\sum_{i=0}^{0} [i(i+1) - (k-1)k] A_i P_i(x),$$

where the initial expression for $\gamma_{k-1}(x)$ has the form

$$\gamma_{k-1}(x) = k(k-1)\gamma_k(x) + [(1-x^2)\gamma_k'(x)]'$$

and allowance is made for formulas (A.1), (A.5), (A.2) and the relationship

$$[(1-x^2)P_i'(x)]' = -i(i+1)P_i(x).$$

Systematically using these formulas gives

$$I_k = \frac{1}{2^k k!^2} \int_{-1}^{1} \frac{\gamma_0(x)}{1-x} dx,$$

$$\gamma_0(x) = (-1)^{k+1} \times \sum_{i=0}^{l} \prod_{j=0}^{k-1} [i(i-1) - j(j-1)] A_i [1 - P_i(x)].$$

Here in the formula for $\gamma_0(x)$ we made the substitution

$$P_i(x) \longrightarrow P_i(x) - 1,$$

which is possible because $\gamma_0(1) = 0$ and $P_i(1) = 1$. Bearing in mind that terms with $i < k$ are zero in the sum, and also formulas (A.3), (14), and [4]

$$\int_{-1}^{1} \frac{1 - P_l(x)}{1-x} dx = 2S_l, \quad S_l = \sum_{k=1}^{l} \frac{1}{k},$$

we finally obtain

$$I_k = (-1)^k 2^{k+1} C_{lk} (S_l - S_k).$$

APPENDIX B

In order to determine

$$h_n(m, \nu + 1) = \frac{1}{n![\Gamma(\nu+1)]^m} \frac{d^n}{d\nu^n} [\Gamma(\nu+1)]^m$$

we take the $(n-1)$th order derivative of

$$\frac{d[\Gamma(\nu+1)]^m}{d\nu} = m[\Gamma(\nu+1)]^m \psi(\nu+1).$$

Here and subsequently $\Gamma(\nu+1)$ is a gamma function, $\psi(\nu+1)$ is a psi function, and $\zeta(m, \nu+1)$ is a Riemann zeta function. (Information on the functions used in this study can be found in [5].) Taking into account

$$\frac{1}{(i-1)!} \psi^{(i-1)}(\nu+1) = (-1)^i \zeta(i, \nu+1), \quad i = 2, 3\ldots$$

and supplementing the definition

$$\zeta(1, \nu+1) = -\psi(\nu+1),$$

we obtain the recurrence formula:

$$h_n(m, \nu+1) = \frac{m}{n} \sum_{i=1}^{n} (-1)^i h_{n-i}(m, \nu+1) \zeta(i, \nu+1),$$

$$n = 1, 2, \ldots, \quad h_0(m, \nu+1) = 1.$$

For the case $m = 1$ we have

$$h_n(\nu+1) \equiv h_n(1, \nu+1).$$

For small n (ν is an integer)

$$h_1(\nu+1) = \psi(\nu+1), \quad \psi(\nu+1) = -C + \sum_{k=1}^{\nu} \frac{1}{k},$$

$$h_2(\nu+1) = \frac{1}{2}[\psi^2(\nu+1) + \zeta(2, \nu+1)],$$

$$\zeta(2, \nu+1) = \frac{\pi^2}{6} - \sum_{k=1}^{\nu} \frac{1}{k^2},$$

and so on. As n increases, the explicit formulas become increasingly complex and a recurrence formula must be used.

For $m = 2$ the recurrence formula determines the value of $h_n(2, \nu+1)$ required in the asymptotic formulas for F_n, $f^{(n)}$.

APPENDIX C

Using the expansion

$$f(q) = \int_0^\infty f_\eta J_0(q\eta) \eta d\eta,$$

the normalization integral for the function $f(q)$ can be obtained immediately if we take into account

$$f_\eta = \int_0^\infty f(q) J_0(q\eta) q dq.$$

Assuming $\eta = 0$, we obtain

$$\int_0^\infty f(q) q dq = f_\eta|_{\eta=0} \text{ and } \int_0^\infty f^{(n)}(X) X dX = \delta_{0n}.$$

Acting on f_η with the operator

$$\hat{L} = \frac{1}{\eta}\frac{d}{d\eta}\eta\frac{d}{d\eta}$$

and bearing in mind that

$$\hat{L}J_0(q\eta) = -q^2 J_0(q\eta),$$

we obtain the formula for the second moment:

$$\overline{q^2} = \int_0^\infty f(q)q^3 dq = -\hat{L}f_\eta|_{\eta=0}.$$

From this it follows that

$$\int_0^\infty f^{(n)}(X)X^3 dX = \begin{cases} 1, & n = 0 \\ \infty, & n = 1 \\ 0, & n \geq 2, \end{cases}$$

i.e., the value of $\overline{q^2}$ in the range $0 \leq q \leq 2$ is mainly determined by the functions with $n = 0, 1$.

APPENDIX D

In the equation for $F_n(z)$ we use the representation

$$F_n(z) = e^{-1}\varphi_n(z), \quad \varphi_n(z) = n!\sum_{k=1}^\infty \frac{a_k}{k!}z^k.$$

For φ_n we obtain

$$\varphi_n'' - \varphi_n' + \frac{n}{z}\varphi_n = \varphi_{n-1}'' - 2\varphi_{n-1}' + \varphi_{n-1}$$

with the initial conditions

$$\varphi_n(0) = 0, \quad \varphi_n'(0) = -f^{(n)}(0)/2.$$

For the coefficients a_k we obtain

$$a_1 = -h_n(n+1), \quad a_{k+1} = \frac{k-n}{k}a_k + c_k,$$

$$c_k = \frac{1}{n}[a_{k+1} - 2a_k + a_{k-1}]_{n-1}, \quad k = 1, 2, \ldots,$$

where c_k are defined in terms of the values of a_k for φ_{n-1}.

For $n = 0$ we have

$$f^{(0)}(z) = 2e^{-z}, \quad F_0(z) = e^{-z}, \quad \varphi_0 = 1.$$

Thus, we have $c_k = \delta_{k1}$ for $n = 1$ and assuming $h_1(2) = \psi(2) = 1 - C$, we have $a_1 = -(1 - C)$, $a_2 = 1$, and so on. Finally we obtain

$$F_1(z) = e^{-z}\left[(C-1)z + \sum_{k=2}^\infty \frac{z^k}{(k-1)k!}\right].$$

Further application of the formulas presents no difficulties: in the formula for c_k we have $a_0 = 0$ and the formulas for $h_n(n+1)$ are given in Appendix B.

In order to obtain asymptotic formulas we use the representations

$$f^{(n)}(X) = (d^n/d\nu^n)I(\nu, z)|_{\nu=n},$$

$$I(\nu, z) = \frac{2}{n!}\int_0^\infty e^{-t}J_0(2\sqrt{zt})t^\nu dt$$

$$= \frac{2}{n!}\Gamma(\nu+1)^{-z}{}_1F_1(-\nu, 1; z).$$

For large z we have

$$I(\nu, z) = -\frac{2\sin(\nu\pi)}{\pi n!}z^{-(\nu+1)}\sum_{k=0}^\infty \frac{\Gamma^2(k+\nu+1)}{k!z^k}.$$

After differentiating we obtain an asymptotic formula for $f^{(n)}(X)$ (see Section 4) where

$$\beta_l(k+n+1) = (-1)^l\sum_{i=1}^l h_{l-i}(2, k+n+1)b_i,$$

$$b_i = (-1)^{(i+1)/2}\pi^{i-1}/i!,$$

where summation is performed over odd values of i and the values of $h_n(2, \nu+1)$ are determined in accordance with Appendix B. For small l (ν is an integer) we have

$$\beta_1(\nu+1) = 1, \quad \beta_2(\nu+1) = -2\psi(\nu+1),$$

$$\beta_3(\nu+1) = 2\psi^2(\nu+1) - \sum_{k=1}^\nu \frac{1}{k^2}.$$

The coefficients α_l is the asymptotic expression for F_n can be obtained using the simple recurrence formula

$$\alpha_l = \frac{1}{k+n}(\beta_l - \alpha_{l-1}), \quad l = 1, 2, \ldots, n, \quad \alpha_0 = 0,$$

with the result that

$$dF_n(z)/dz = -f^{(n)}(z)/2.$$

REFERENCES

1. V. I. Yurchenko, Zh. Éksp. Teor. Fiz. **116**, 418 (1999) [JETP **89**, 223 (1999)].
2. G. Moliere, Z. Naturforsch. A **2**, 133 (1947).
3. G. Moliere, Z. Naturforsch. A **3**, 78 (1948).
4. H. A. Bethe, Phys. Rev. **89**, 256 (1953).
5. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Fizmatgiz, Moscow, 1963; Academic, New York, 1980).
6. V. I. Yurchenko, Preprint No. 2000-69, IYaF SO RAN (Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1130–1140.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1307–1318.

ATOMS, SPECTRA, RADIATION

Stability of Weakly Nonparaxial Spatial Optical Solitons in a Medium with a Kerr Nonlinearity

N. N. Rozanov[a, *], N. V. Vysotina[a], and A. G. Vladimirov[b]

[a] *Research Institute of Laser Physics, St. Petersburg, 199034 Russia*
[b] *Institute of Physics, St. Petersburg State University, Petrodvorets, St. Petersburg, 198904 Russia*
* *e-mail: rosanov@ilph.spb.su*

Received June 29, 2000

Abstract—A variant of perturbation theory is developed to determine the characteristics and stability of transversely two-dimensional spatial solitons in a Kerr medium under conditions of small deviations from paraxial. Distributions of the transverse and longitudinal components of the soliton electric and magnetic fields are obtained. It is shown that the power of a nonparaxial soliton in a Kerr medium increases as the propagation constant increases. A linear analysis is made of soliton stability. In addition to confirming stability, this analysis revealed "internal modes" of nonparaxial solitons and their characteristics were determined. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The final stage of the self-focusing of high-intensity radiation in a transparent medium with a Kerr nonlinearity is of considerable interest and has been studied on many occasions (see the book [1] and the literature cited therein). Using the standard approximation of slowly varying amplitudes (quasi-optic or paraxial equation for the electric field envelope), the theory predicts the absence of stable spatial solitons (beams having a constant transverse field profile) and collapse of radiation beams having powers exceeding the critical self-focusing power. The limitation of collapse and formation of spatial solitons as predicted in [2] may occur for various reasons. The most common of these is the nonparaxial nature of narrow (width comparable to the wavelength) beams [1] which precludes us from using the approximation of the quasi-optic equation.

Although nonparaxial effects also arise for the scalar nonlinear wave equation [3–6], for electromagnetic radiation it is important to allow for its polarization. Thus, nonparaxial self-focusing theory should be based on a complete system of Maxwell nonlinear vector equations. Previously, vector self-focusing theory was preferentially developed for special cases of the polarization of radiation having an axisymmetric intensity distribution and unit nonzero electric field component [7–10], i.e., in fact for a scalar variant. The spatial solitons which may appear in this case have a power considerably higher than the critical self-focusing power, which serves as an indication of their instability [11]. For a transversely one-dimensional geometry it is possible to construct a fairly comprehensive classification of an infinite set of localized structures [12, 13] although in a continuous nonlinear medium all these structures are unstable with respect to decay along the other transverse coordinate. As far as we are aware, the existence of stable spatial solitons of electromagnetic radiation in a medium with a Kerr nonlinearity has not yet been proven. As will be shown subsequently, using numerical calculations of the type [14, 15] to solve this problem may not yield the correct result because establishment is extremely slow under weakly nonparaxial conditions.

The task for the present study is to make an analytic investigation of the characteristics and properties of weakly nonparaxial spatial transversely two-dimensional solitons of electromagnetic radiation in a medium with a Kerr nonlinearity. The analysis is based on perturbation theory with a small nonparaxial parameter which is used to find nonparaxial corrections to the soliton shape and to determine its stability. The initial (zeroth) approximation is the well-studied nonlinear Schrödinger equation and beams having an axisymmetric intensity distribution and linearly polarized radiation ("Townes mode"). Following [16] (see also [14]), in Section 2 we give a derivation of the control equation for the envelope of a weakly nonparaxial soliton field. We then obtain its approximate solution in Section 3, i.e., we determine the transverse distribution of the electron and magnetic field intensities. In Sections 4 and 5 we analyze soliton stability using a method proposed in [17] (see also [18]) for paraxial solitons in a medium with saturation of the nonlinearity. For this we use a linearized control equation whose properties are analyzed in Section 4. The final conclusion on the stability of a weakly nonparaxial soliton is formulated in Section 5 and calculations of various matrix elements are presented in the Appendix. The results are discussed in the Conclusions.

1063-7761/00/9106-1130$20.00

2. EVOLUTION EQUATION

The initial equations are the Maxwell equations for monochromatic radiation of frequency ω [in complex notation the factor $\exp(-i\omega t)$ is omitted, t is the time) in a nonmagnetic medium (with unit magnetic permeability):

$$\mathrm{rot}\mathbf{E} = i\frac{\omega}{c}\mathbf{H}, \quad \mathrm{rot}\mathbf{H} = -i\frac{\omega}{c}\mathbf{D}, \qquad \mathrm{div}\mathbf{H} = 0, \quad \mathrm{div}\mathbf{D} = 0. \tag{2.1}$$

Here $\mathbf{E}$ and $\mathbf{H}$ are the intensities of the electric and magnetic fields, c is the speed of light in vacuum, and $\mathbf{D}$ is the electric induction which has the form (Kerr striction nonlinearity)

$$\mathbf{D} = (\varepsilon_0 + \varepsilon_{nl})\mathbf{E}, \quad \varepsilon_{nl} = \varepsilon_2|\mathbf{E}|^2, \quad \varepsilon_2 > 0. \tag{2.2}$$

Here ε_0 is the linear permittivity. The form (2.2) allows only for self-interaction effects whereas the generation of third and high-order harmonics is considered to be ineffective (phase matching conditions are not satisfied for these).

Eliminating the magnetic field intensity from the Maxwell equation, we obtain the generalized Helmholtz equation

$$\frac{\partial^2\mathbf{E}}{\partial z^2} + \Delta_\perp\mathbf{E} + \frac{\omega^2}{c^2}\mathbf{D} - \mathrm{grad}\,\mathrm{div}\mathbf{E} = 0,$$

$$\Delta_\perp = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

From this vector equation it follows that

$$\frac{\partial^2\mathbf{E}_\perp}{\partial z^2} + \Delta_\perp\mathbf{E}_\perp + \frac{\omega^2}{c^2}\mathbf{D}_\perp - \mathrm{grad}_\perp\mathrm{div}\mathbf{E} = 0. \tag{2.3}$$

Here we introduce the transverse components of the electric field $\mathbf{E}_\perp = (E_x, E_y)$ and the induction $\mathbf{D}_\perp = (D_x, D_y)$. We transform the last term on the left-hand side of Eq. (2.3) as follows. From the final Maxwell equation (2.1) it follows that

$$\mathrm{div}\mathbf{E} = -\frac{1}{\varepsilon_0 + \varepsilon_{nl}}\mathbf{E}\cdot\mathrm{grad}\varepsilon_{nl} = -\frac{1}{\varepsilon_0 + \varepsilon_{nl}}\left(\mathbf{E}_\perp\cdot\mathrm{grad}_\perp\varepsilon_{nl} + E_z\frac{\partial\varepsilon_{nl}}{\partial z}\right). \tag{2.4}$$

In the lowest approximation (weak and continuously varying nonlinearity) we have $\mathrm{div}\mathbf{E} = 0$ whence

$$E_z \approx \frac{i}{k}\mathrm{div}_\perp\mathbf{E}_\perp, \tag{2.5}$$

where $k = (\omega/c)\sqrt{\varepsilon_0}$ is the wave number in the linear medium. For a more accurate estimate we have

$$\mathrm{div}\mathbf{E} \approx -\frac{1}{\varepsilon_0}\mathbf{E}_\perp\cdot\mathrm{grad}_\perp\varepsilon_{nl}.$$

We then arrive at a closed equation for the transverse field components:

$$\frac{\partial^2\mathbf{E}}{\partial z^2} + \Delta_\perp\mathbf{E}_\perp + \frac{\omega^2}{c^2}\varepsilon_0\mathbf{E}_\perp + \frac{\omega^2}{c^2}\varepsilon_2|\mathbf{E}_\perp|^2\mathbf{E}_\perp = \mathbf{Q}_s(\mathbf{E}_\perp), \tag{2.6}$$

where

$$\mathbf{Q}_s(\mathbf{E}_\perp) = -\frac{\varepsilon_2}{\varepsilon_0} \times [|\mathrm{div}_\perp\mathbf{E}_\perp|^2\mathbf{E}_\perp + \mathrm{grad}_\perp(\mathbf{E}_\perp\cdot\mathrm{grad}_\perp|\mathbf{E}_\perp|^2)]. \tag{2.7}$$

For a steady-state soliton we have

$$\mathbf{E}_\perp = \mathbf{A}_s(\mathbf{r}_\perp)e^{i\Gamma z}, \tag{2.8}$$

where Γ is the real propagation constant. The solitons have a continuous spectrum with respect to Γ and the condition $\Gamma^2 > k^2$ must be satisfied for the field to decrease (tend to zero) at the soliton edge, in accordance with (2.6). A measure of the nonparaxial property is given by

$$\mu^2 = \frac{\Gamma^2 - k^2}{k^2} \approx 2\frac{\Gamma - k}{k} \ll 1. \tag{2.9}$$

Inequality (2.9) implies that the propagation constant Γ for the soliton is close to the wave number k in the linear medium. This occurs if the soliton width is considerably greater than the wavelength of light, the maximum amplitude of the field is extremely small, and the power is close to the critical self-focusing power (see below). In this limit the order of the derivative with respect to z can be reduced [14]. We shall assume that the field is close to a steady-state soliton so that

$$\mathbf{E}_\perp = \mathbf{A}_\perp(\mathbf{r}_\perp, z)e^{i\Gamma z}, \tag{2.10}$$

where the dependence of the amplitude $\mathbf{A}_\perp(\mathbf{r}_\perp, z)$ on the longitudinal coordinate z is slow (on a scale of the order of the wavelength of light). Then, retaining terms of the lowest order of smallness in the transformations of $\partial^2\mathbf{E}/\partial r^2$, instead of (2.6) we obtain the evolution equation

$$2i\Gamma\frac{\partial\mathbf{A}_\perp}{\partial z} + \Delta_\perp\mathbf{A}_\perp - (\Gamma^2 - k^2)\mathbf{A}_\perp + k^2\frac{\varepsilon_2}{\varepsilon_0}|\mathbf{A}_\perp|^2\mathbf{A}_\perp = \mathbf{Q}_s(\mathbf{A}_\perp) + \mathbf{Q}_z(\mathbf{A}_\perp), \tag{2.11}$$

where

$$\mathbf{Q}_z(\mathbf{A}_\perp) = \frac{1}{4k^2}[\Delta_\perp - (\Gamma^2 - k^2)] \times \left[\Delta_\perp \mathbf{A}_\perp - (\Gamma^2 - k^2)\mathbf{A}_\perp + k^2\frac{\varepsilon_2}{\varepsilon_0}|\mathbf{A}_\perp|^2\mathbf{A}_\perp\right] + \frac{\varepsilon_2}{4\varepsilon_0}\left\{(\Delta_\perp \mathbf{A}_\perp \cdot \mathbf{A}_\perp^*)\mathbf{A}_\perp - (\Delta_\perp \mathbf{A}_\perp^* \cdot \mathbf{A}_\perp)\mathbf{A}_\perp + |\mathbf{A}_\perp|^2\left[\Delta_\perp \mathbf{A}_\perp - (\Gamma^2 - k^2)\mathbf{A}_\perp + k^2\frac{\varepsilon_2}{\varepsilon_0}|\mathbf{A}_\perp|^2\mathbf{A}_\perp\right]\right\}. \quad (2.12)$$

Note that for a stationary spatial soliton we have $\mathbf{Q}_z(\mathbf{A}_s) = 0$. Equation (2.11) can be used not only to find a weakly nonparaxial stationary soliton but also to investigate its stability. In order to isolate the nonparaxial parameter in explicit form in (2.11), we convert to dimensionless coordinates and amplitude:

$$z' = \frac{\Gamma^2 - k^2}{\Gamma}z, \quad (x', y') = \sqrt{\Gamma^2 - k^2}(x, y), \quad \mathbf{A}'_\perp = \frac{k}{\sqrt{\Gamma^2 - k^2}}\sqrt{\frac{\varepsilon_2}{\varepsilon_0}}\mathbf{A}_\perp, \quad \Delta'_\perp = \frac{1}{\Gamma^2 - k^2}\Delta_\perp. \quad (2.13)$$

Then (2.11) has the form

$$2i\frac{\partial \mathbf{A}'_\perp}{\partial z'} + \Delta'_\perp \mathbf{A}'_\perp - \mathbf{A}'_\perp + |\mathbf{A}'_\perp|^2\mathbf{A}'_\perp = \mu^2[\mathbf{Q}'_s(\mathbf{A}'_\perp) + \mathbf{Q}'_z(\mathbf{A}'_\perp)], \quad (2.14)$$

where

$$\mathbf{Q}'_s(\mathbf{A}'_\perp) = -[|\mathrm{div}'_\perp \mathbf{A}'_\perp|^2\mathbf{A}'_\perp + \mathrm{grad}'_\perp(A'_\perp \cdot \mathrm{grad}'_\perp|\mathbf{A}'_\perp|^2)],$$

$$\mathbf{Q}'_z(\mathbf{A}'_\perp) = \frac{1}{4}[\Delta'_\perp - 1][\Delta'_\perp \mathbf{A}'_\perp - \mathbf{A}'_\perp + |\mathbf{A}'_\perp|^2\mathbf{A}'_\perp] + \frac{1}{4}[(\Delta'_\perp \mathbf{A}'_\perp \cdot \mathbf{A}'^*_\perp)\mathbf{A}'_\perp - (\Delta'_\perp \mathbf{A}'^*_\perp \cdot \mathbf{A}'_\perp)\mathbf{A}'_\perp + |\mathbf{A}'_\perp|^2(\Delta'_\perp \mathbf{A}'_\perp - \mathbf{A}'_\perp + |\mathbf{A}'_\perp|^2\mathbf{A}'_\perp)]. \quad (2.15)$$

The dimensionless form (2.14) is convenient for determining the corrections to the shape of a steady-state soliton while the dimensional form (2.11) is convenient for analyzing its stability, containing derivatives of the amplitudes with respect to the propagation constant Γ.

3, NONPARAXIAL CORRECTIONS TO SOLITON SHAPE

The right-hand side of Eq. (2.14) serves as a correction (as a result of the nonparaxial property) to the nonlinear Schrödinger equation for which $\mu = 0$. Note that this correction is nonlocal since it not only depends on the transverse components of the field intensity but also on their derivatives in the transverse direction. We find the field distribution for a weakly nonparaxial stationary [of the form (2.8)] spatial soliton by solving Eq. (2.14) using perturbation theory with the small parameter (2.9). In the lowest approximation this equation gives the standard vector nonlinear Schrödinger equation (we omit the primes in this section):

$$\Delta_\perp A_{s0x} - A_{s0x} + (|A_{s0x}|^2 + |A_{s0y}|^2)A_{s0x} = 0, \quad \Delta_\perp A_{s0y} - A_{s0y} + (|A_{s0x}|^2 + |A_{s0y}|^2)A_{s0y} = 0. \quad (3.1)$$

For the main (fundamental) soliton the functions A_{s0x} and A_{s0y} can be considered to be real. Generally speaking, the Eqs. (3.1) are written for the particular case of a soliton with a common propagation constant for both polarizations but a difference between these values is only possible in the paraxial approximation [1].

We now introduce a small correction to the steady-state soliton:

$$\mathbf{A}_s(\mathbf{r}_\perp) = \mathbf{A}_{s0}(\mathbf{r}_\perp) + \mu^2\delta\mathbf{A}_s(\mathbf{r}_\perp). \quad (3.2)$$

Equation (2.14) linearized with respect to the perturbation $\delta\mathbf{A}_s$ is written in the form

$$\Delta_\perp \delta\mathbf{A}_s - \delta\mathbf{A}_s + [|A_{s0}|^2\delta\mathbf{A}_s + (\mathbf{A}^*_{s0} \cdot \delta\mathbf{A}_s)\mathbf{A}_{s0} + (\mathbf{A}_{s0} \cdot \delta\mathbf{A}^*_s)\mathbf{A}_{s0}] = \mathbf{Q}_s(\mathbf{A}_{s0}). \quad (3.3)$$

In terms of Cartesian components we have

$$\Delta_\perp \delta A_{sx} - \delta A_{sx} + [(A^2_{s0x} + A^2_{s0y})\delta A_{sx} + A^2_{s0x}(\delta A_{sx} + \delta A^*_{sx}) + A_{s0x}A_{s0y}(\delta A_{sy} + \delta A^*_{sy})] = Q_{sx}(A_{s0x}, A_{s0y}),$$
$$\Delta_\perp \delta A_{sy} - \delta A_{sy} + [(A^2_{s0x} + A^2_{s0y})\delta A_{sy} + A_{s0x}A_{s0y}(\delta A_{sx} + \delta A^*_{sx}) + A^2_{s0y}(\delta A_{sy} + \delta A^*_{sy})] = Q_{sy}(A_{s0x}, A_{s0y}), \quad (3.4)$$

$$Q_{sx}(A_{s0x}, A_{s0y}) = -\left\{\left|\frac{\partial A_{s0x}}{\partial x} + \frac{\partial A_{s0y}}{\partial x}\right|^2 A_{s0x} + \frac{\partial}{\partial x}\left[A_{s0x}\frac{\partial(A^2_{s0x} + A^2_{s0y})}{\partial x} + A_{s0y}\frac{\partial(A^2_{s0x} + A^2_{s0y})}{\partial y}\right]\right\}, \quad (3.5)$$

$$Q_{sy}(A_{s0x}, A_{s0y}) = -\left\{\left|\frac{\partial A_{s0x}}{\partial x} + \frac{\partial A_{s0y}}{\partial x}\right|^2 A_{s0y} + \frac{\partial}{\partial y}\left[A_{s0x}\frac{\partial(A^2_{s0x} + A^2_{s0y})}{\partial x} + A_{s0y}\frac{\partial(A^2_{s0x} + A^2_{s0y})}{\partial y}\right]\right\}.$$

As the zeroth approximation we take the linearly polarized Townes mode having an axisymmetric field distribution (polar coordinates r, φ, $r = \sqrt{x^2 + y^2}$):

$$A_{s0x} = F_0(r), \quad A_{s0y} = 0. \tag{3.6}$$

The function $F_0(r)$ is defined as finite over the entire range $0 < r < \infty$ and the solution of the equation which tends to zero as $r \longrightarrow \infty$ is

$$L_0 F_0 = 0, \quad L_0 = \Delta_\perp - 1 + F_0^2. \tag{3.7}$$

Bearing in mind the axial symmetry of the Townes mode, Eq. (3.7) has the form

$$\frac{d^2F_0}{dr^2} + \frac{1}{r}\frac{dF_0}{dr} - F_0 + F_0^3 = 0. \tag{3.8}$$

We write Eqs. (3.4) for the corrections to the soliton shape in the form

$$\begin{aligned} L_1 \delta A'_{sx} &= Q_{x0}(r) + Q_{x2}(r)\cos 2\varphi, \\ L_0 \delta A''_{sx} &= 0, \\ L_0 \delta A_{sy} &= Q_{y2}(r)\sin 2\varphi, \end{aligned} \tag{3.9}$$

where

$$L_1 = L_0 + 2F_0^2, \quad \delta A_{sx} = \delta A'_{sx} + i\delta A''_{sx}, \tag{3.10}$$

$$Q_{x0}(r) = -\left[\frac{1}{2}F_0\left(\frac{dF_0}{dr}\right)^2 + \left(\frac{d}{dr} + \frac{1}{r}\right)\left(F_0^2\frac{dF_0}{dr}\right)\right],$$

$$Q_{x2}(r) = -\left[\frac{1}{2}F_0\left(\frac{dF_0}{dr}\right)^2 + \left(\frac{d}{dr} - \frac{1}{r}\right)\left(F_0^2\frac{dF_0}{dr}\right)\right], \tag{3.11}$$

$$Q_{y2}(r) = -\left(\frac{d}{dr} - \frac{1}{r}\right)\left(F_0^2\frac{dF_0}{dr}\right).$$

The second of linear Eqs. (3.9) (for $\delta A''_{sx}$) has the solution [see (3.7)]

$$\delta A''_{sx} = C'' A_{s0}, \quad A_{s0} \equiv A_{s0x} = F_0, \tag{3.12}$$

which corresponds to a phase shift of the initial soliton. Since we are not interested in this shift, we can set $C'' = 0$ and accordingly $\delta A''_{sx} = 0$. The two remaining inhomogeneous Eqs. (3.9) can be solved provided that their right-hand sides are orthogonal to the solutions of the corresponding homogeneous equations with the boundary conditions specified above. For the last of Eqs. (3.9) orthogonality follows from the angular dependence of the right-hand side,

$$\int_0^{2\pi} d\varphi \sin 2\varphi = 0.$$

Thus, omitting the solution of the homogeneous equation (symmetry with respect to rotation of the axes x, y), we obtain the correction δA_{sy} in the form

$$\delta A_{sy} = F_y(r)\sin 2\varphi. \tag{3.13}$$

Then $F_y(r)$ is defined as the only finite axisymmetric solution of the equation

$$L_2 F_y = Q_{y2}(r), \quad L_2 = L_1 - \frac{4}{r^2}. \tag{3.14}$$

The solution of the homogeneous equation corresponding to the first of the equations (3.9) corresponds to a shift of the initial soliton along x and y:

$$\begin{aligned} \delta A'_{sx1} &= C_1\frac{\partial F_0}{\partial x} = C_1\frac{dF_0}{dr}\cos\varphi, \\ \delta A'_{sx2} &= C_2\frac{\partial F_0}{\partial y} = C_2\frac{dF_0}{dr}\sin\varphi. \end{aligned} \tag{3.15}$$

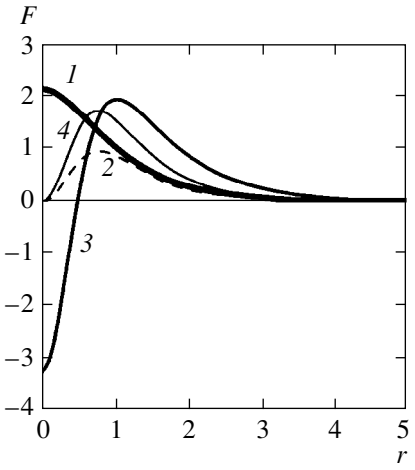
The orthogonality condition is again satisfied as a result of the angular dependence of these solutions and the right-hand side of this equation. Also omitting the solution of the homogeneous Eq. (3.14) ($C_1 = C_2 = 0$) we obtain $\delta A'_{sx}$ in the form

$$\delta A'_{sx} = F_1(r) + F_2(r)\cos 2\varphi. \tag{3.16}$$

The radial functions appearing in (3.16) are obtained as (unique) finite axisymmetric solutions of the equations

$$L_1 F_1 = Q_{x0}, \quad L_2 F_2 = Q_{x2}. \tag{3.17}$$

Graphs of these functions obtained by solving numerically axisymmetric variants of Eqs. (3.8), (3.14), and (3.17) with the conditions of finiteness for $r = 0$ and which decrease as $r \longrightarrow \infty$ specified above are plotted in the figure.



Radial profiles of the amplitudes F_0 of a fundamental soliton (Townes modes, curve *1*) and its nonparaxial distortions F_y (curve *2*), F_1 (curve *3*), and F_2 (curve *4*).

We shall now return to dimensional quantities. For the electric field intensity of a steady-state soliton we have

$$E_x = \sqrt{\frac{\varepsilon_0}{\varepsilon_2}}\{\mu F_0(r) + \mu^3[F_1(r) + F_2(r)\cos 2\varphi]\},$$
$$E_y = \sqrt{\frac{\varepsilon_0}{\varepsilon_2}}\mu^3 F_y(r)\sin 2\varphi, \qquad (3.18)$$
$$E_z = i\mu^2\sqrt{\frac{\varepsilon_0}{\varepsilon_2}}\frac{dF_0(r)}{dr}\cos\varphi,$$
$$r = \sqrt{\Gamma^2 - k^2}\rho, \quad \rho = \sqrt{x^2 + y^2}.$$

In accordance with (3.18), the field exhibits weak axial symmetry. The main correction to the initial linear polarization occurs as a result of the longitudinal component of the field $E_z \propto \mu^2$. For this correction the phase of the field is shifted relative to the main component by $\pi/2$ since the polarization becomes elliptic. Time oscillations of the electric intensity vector are described by a prolate ellipse in the *xz* plane. When additional allowance is made for corrections proportional to μ^3, it is found that the slope of the plane in which the ellipse is located varies over the beam cross section.

The components of the magnetic intensity are expressed in terms of the electric field intensity using the Maxwell equations (2.1):

$$H_x = -\frac{c}{\omega}\left(\Gamma E_y + i\frac{\partial E_z}{\partial y}\right) = -\mu^3\frac{\varepsilon_0}{2\sqrt{\varepsilon_2}}$$
$$\times\left(\frac{2}{r}\frac{dF_0}{dr} - F_0 + F_0^3 + 2F_y\right)\sin 2\varphi,$$
$$H_y = \frac{c}{\omega}\left(\Gamma E_x + i\frac{\partial E_z}{\partial x}\right) = H_{y0} + \mu^3\frac{\varepsilon_0}{2\sqrt{\varepsilon_2}} \qquad (3.19)$$
$$\times\left[(2F_1 + F_0^3) + \left(2F_2 + F_0^3 - F_0 + \frac{2}{r}\frac{dF_0}{dr}\right)\cos 2\varphi\right],$$
$$H_z = \frac{i}{k}\sqrt{\varepsilon_0}\frac{\partial E_x}{\partial y} = i\mu^2\frac{\varepsilon_0}{\sqrt{\varepsilon_2}}\frac{dF_0}{dr}\sin\varphi,$$
$$H_{y0} = \frac{\varepsilon_0}{\sqrt{\varepsilon_2}}\mu F_0(r).$$

The radiation power P is defined as the integral of the longitudinal component of the time-averaged Poynting vector over the transverse coordinates:

$$P = \frac{c}{8\pi}\iint \mathrm{Re}(E_x^* H_y - E_y^* H_x)\,dx\,dy. \qquad (3.20)$$

Substituting into (3.20) the approximate expressions for the soliton field intensities obtained above, we obtain the power

$$P = \frac{c^2}{8\pi\omega k}\frac{\varepsilon_0}{\varepsilon_2}(P_0 + \mu^2 p_1). \qquad (3.21)$$

The constants P_0 and p_1 are defined as follows:

$$P_0 = \int_0^\infty F_0^2(r) r\,dr = \frac{11.7}{2\pi} = 1.862,$$
$$p_1 = \int_0^\infty \left[2F_0(r)F_1(r) + \frac{1}{2}F_0^4(r)\right] r\,dr = 5.990. \qquad (3.22)$$

In (3.21) the term containing P_0 corresponds to the critical self-focusing power which agrees with that obtained in the paraxial limit ($\mu \longrightarrow 0$) where it does not depend on the propagation constant [1]. The term containing p_1 is the nonparaxial correction to the power which depends on the propagation constant. Note that the increase in soliton power with increasing propagation constant, which follows from Eqs. (3.2) and (3.22), is consistent with the Vakhitov–Kolokolov criterion [19] for soliton stability. However, this criterion was obtained in the paraxial approximation and thus we still need to demonstrate the stability of a nonparaxial soliton.

4. LINEARIZED EQUATIONS

For a linear analysis of stability, we shall set [see (3.2)]

$$\mathbf{A}_\perp(\mathbf{r}_\perp, z) = \mathbf{A}_s(\mathbf{r}_\perp) + \delta\mathbf{A}(\mathbf{r}_\perp, z). \qquad (4.1)$$

Substituting (4.1) into (2.11) and linearizing this with respect to the small perturbation $\delta\mathbf{A}$, we find

$$2i\Gamma\frac{\partial\delta\mathbf{A}}{\partial z} + \Delta_\perp\delta\mathbf{A} - (\Gamma^2 - k^2)\delta\mathbf{A} + k^2\frac{\varepsilon_2}{\varepsilon_0} \qquad (4.2)$$
$$\times[A_{s0}^2\delta\mathbf{A} + \mathbf{A}_{s0}(\mathbf{A}_{s0}\cdot\delta\mathbf{A}) + \mathbf{A}_{s0}(\mathbf{A}_{s0}\cdot\delta\mathbf{A}^*)] = \delta\mathbf{Q}.$$

Here we have

$$\delta\mathbf{Q} = \delta\mathbf{Q}_s + \delta\mathbf{Q}_z + \delta\mathbf{Q}_k,$$
$$\delta\mathbf{Q}_k = -k^2\frac{\varepsilon_2}{\varepsilon_0}[(\mathbf{A}_s^2 - \mathbf{A}_{s0}^2)\delta\mathbf{A} + \delta\mathbf{A}_s(\mathbf{A}_{s0}\cdot\delta\mathbf{A})$$
$$+ \mathbf{A}_{s0}(\delta\mathbf{A}_s\cdot\delta\mathbf{A}) + \delta\mathbf{A}_s(\delta\mathbf{A}_s\cdot\delta\mathbf{A}) + \delta\mathbf{A}_s(\mathbf{A}_{s0}\cdot\delta\mathbf{A}^*)$$
$$+ \mathbf{A}_{s0}(\delta\mathbf{A}_s\cdot\delta\mathbf{A}^*) + \delta\mathbf{A}_s(\delta\mathbf{A}_s\cdot\delta\mathbf{A}^*)],$$
$$\delta\mathbf{Q}_s = -\frac{\varepsilon_2}{\varepsilon_0}\{|\mathrm{div}_\perp\mathbf{A}_s|^2\delta\mathbf{A}$$
$$+ [\mathrm{div}_\perp\mathbf{A}_s^*\mathrm{div}_\perp\delta\mathbf{A} + \mathrm{div}_\perp\mathbf{A}_s\mathrm{div}_\perp\delta\mathbf{A}^*]\mathbf{A}_s$$
$$+ \mathrm{grad}_\perp[\mathbf{A}_s\cdot\mathrm{grad}_\perp(\mathbf{A}_s^*\cdot\delta\mathbf{A} + \mathbf{A}_s\cdot\delta\mathbf{A}^*)$$

$$+\delta\mathbf{A}\cdot\mathrm{grad}_\perp|\mathbf{A}_s|^2]\},$$
$$\delta\mathbf{Q}_z = \frac{1}{4k^2}[\Delta_\perp-(\Gamma^2-k^2)]^2\delta\mathbf{A} \tag{4.3}$$
$$+\frac{\varepsilon_2}{4\varepsilon_0}\{\Delta_\perp[|\mathbf{A}_s|^2\delta\mathbf{A}+\mathbf{A}_s(\mathbf{A}_s^*\cdot\delta\mathbf{A})+\mathbf{A}_s(\mathbf{A}_s\cdot\delta\mathbf{A}^*)]$$
$$+|\mathbf{A}_s|^2\Delta_\perp\delta\mathbf{A}+(\mathbf{A}_s^*\cdot\delta\mathbf{A}+\mathbf{A}_s\cdot\delta\mathbf{A}^*)\Delta_\perp\mathbf{A}_s$$
$$-2(\Gamma^2-k^2)[|\mathbf{A}_s|^2\delta\mathbf{A}+\mathbf{A}_s(\mathbf{A}_s^*\cdot\delta\mathbf{A})+\mathbf{A}_s(\mathbf{A}_s\cdot\delta\mathbf{A}^*)]$$
$$+k^2\frac{\varepsilon_2}{\varepsilon_0}[|\mathbf{A}_s|^4\delta\mathbf{A}+2\mathbf{A}_s|\mathbf{A}_s|^2(\mathbf{A}_s^*\cdot\delta\mathbf{A}+\mathbf{A}_s\cdot\delta\mathbf{A}^*)]$$
$$+(\Delta_\perp\delta\mathbf{A}\cdot\mathbf{A}_s^*)\mathbf{A}_s+(\Delta_\perp\mathbf{A}_s\cdot\delta\mathbf{A}^*)\mathbf{A}_s$$
$$+(\Delta_\perp\mathbf{A}_s\cdot\mathbf{A}_s^*)\delta\mathbf{A}-(\Delta_\perp\delta\mathbf{A}^*\cdot\mathbf{A}_s)\mathbf{A}_s$$
$$-(\Delta_\perp\mathbf{A}_s^*\cdot\delta\mathbf{A})\mathbf{A}_s-(\Delta_\perp\mathbf{A}_s^*\cdot\mathbf{A}_s)\delta\mathbf{A}\}.$$

Introducing the real and imaginary parts of the x-components of the perturbation $\delta A_x = \delta A_r + i\delta A_i$ and the inhomogeneity $\delta Q_x = \delta Q_r + i\delta Q_i$, we write the linear equation (4.2) in the form

$$2\Gamma\frac{\partial\delta A_r}{\partial z}+L_0\delta A_i = \delta Q_i,$$
$$-2\Gamma\frac{\partial\delta A_i}{\partial z}+L_1\delta A_r = \delta Q_r, \tag{4.4}$$
$$2i\Gamma\frac{\partial\delta A_y}{\partial z}+L_0\delta A_y = \delta Q_y,$$

and also in the matrix form

$$2\Gamma\frac{\partial\delta\mathbf{A}}{\partial z} = \mathbf{M}\delta\mathbf{A}. \tag{4.5}$$

Here

$$\delta\mathbf{A} = \begin{pmatrix}\delta A_r\\ \delta A_i\\ \delta A_y\end{pmatrix},$$

we have

$$\mathbf{M}\delta\mathbf{A} = \begin{pmatrix}-L_0\delta A_i+\delta Q_i\\ L_1\delta A_r-\delta Q_r\\ iL_0\delta A_y-i\delta Q_y\end{pmatrix},$$
$$A_{s0} = \mu\sqrt{\frac{\varepsilon_0}{\varepsilon_2}}F_0(\sqrt{\Gamma^2-k^2}\rho), \tag{4.6}$$
$$L_0 = \Delta_\perp-(\Gamma^2-k^2)+k^2\frac{\varepsilon_2}{\varepsilon_0}A_{s0}^2,$$
$$L_1 = L_0+2k^2\frac{\varepsilon_2}{\varepsilon_0}A_{s0}^2.$$

The form of the operators $L_{0,1}$ in (4.6) corresponds to the dimensional form of the relationships (3.7) and (3.10) with a choice of linearly polarized unperturbed soliton (3.6) when the control equations for the perturbations are simplified considerably. The properties of solutions of the equations corresponding to the linearized nonlinear Schrödinger equation ($\delta\mathbf{Q}\longrightarrow 0$) are also important:

$$2\Gamma\frac{\partial\delta A_{r0}}{\partial z}+L_0\delta A_{i0} = 0,$$
$$-2\Gamma\frac{\partial\delta A_{i0}}{\delta z}+L_1\delta A_{r0} = 0, \tag{4.7}$$
$$2i\Gamma\frac{\partial\delta A_{y0}}{\partial z}+L_0\delta A_{y0} = 0,$$

or in matrix form

$$2\Gamma\frac{\partial\delta\mathbf{A}_0}{\partial z} = \mathbf{M}_0\delta\mathbf{A}_0,$$
$$\mathbf{M}_0 = \begin{pmatrix}0 & -L_0 & 0\\ L_1 & 0 & 0\\ 0 & 0 & iL_0\end{pmatrix}. \tag{4.8}$$

We write the simplest solutions of the system (4.4) [or (4.5)] and (4.7) [or (4.8)]. First, it follows from the well-known symmetry with respect to the phase shift and propagation constant of a steady-state soliton that the linearized equations have two solutions. The first of these

$$\delta\mathbf{A}_\varphi = \begin{pmatrix}0\\ A_{sx}\\ iA_{sy}\end{pmatrix},\quad \delta\mathbf{A}_\varphi^{(0)} = \begin{pmatrix}0\\ A_{s0}\\ 0\end{pmatrix}, \tag{4.9}$$

corresponds to the eigenvector of the matrix $\mathbf{M}$ ($\mathbf{M}_0$) with zero eigenvalue

$$\mathbf{M}\delta\mathbf{A}_\varphi = 0,\quad \mathbf{M}_0\delta\mathbf{A}_\varphi^{(0)} = 0. \tag{4.10}$$

The second solution

$$\delta\mathbf{A}_\Gamma = \begin{pmatrix}\frac{1}{2\Gamma}\frac{\partial A_{sx}}{\partial\Gamma}\\ 0\\ \frac{i}{2\Gamma}\frac{\partial A_{sy}}{\partial\Gamma}\end{pmatrix},\quad \delta\mathbf{A}_\Gamma^{(0)} = \begin{pmatrix}\frac{1}{2\Gamma}\frac{\partial A_{s0}}{\partial\Gamma}\\ 0\\ 0\end{pmatrix}, \tag{4.11}$$

is not an eigenvalue but a root:

$$\mathbf{M}\delta\mathbf{A}_\Gamma = \delta\mathbf{A}_\varphi,\quad \mathbf{M}^2\delta\mathbf{A}_\Gamma = 0,$$
$$\mathbf{M}_0\delta\mathbf{A}_\Gamma^{(0)} = \delta\mathbf{A}_\varphi^{(0)},\quad \mathbf{M}_0^2\delta\mathbf{A}_\Gamma^{(0)} = 0. \tag{4.12}$$

Finally, the symmetry with respect to rotation of the *x*, *y* axes yields an eigenvector with zero eigenvalue:

$$\delta\mathbf{A}_{\text{rot}} = C_{\text{rot}}\begin{pmatrix} -A_{sy} \\ 0 \\ A_{sx} \end{pmatrix}, \quad \delta A_{\text{rot}}^{(0)} = C_{\text{rot}}\begin{pmatrix} 0 \\ 0 \\ A_{sx0} \end{pmatrix}, \tag{4.13}$$

$$\mathbf{M}\delta\mathbf{A}_{\text{rot}} = 0, \quad \mathbf{M}_0\delta\mathbf{A}_{\text{rot}}^{(0)} = 0.$$

Subsequently, disregarding the rotational transformation of the axes, we set $C_{\text{rot}} = 0$.

The following two solutions are specific to the linearized nonlinear Schrödinger equation [Eqs. (4.7) or (4.8)] whereas they are absent for the more general form of Eqs. (4.4) or (4.5). The invariance of the nonlinear Schrödinger equation to a focusing transformation determined by Talanov [20] thus yields the solution (4.7) [21]:

$$\delta A_{r0} = \frac{\partial A_{s0}}{\partial\Gamma}z, \quad \delta A_{i0} = \frac{1}{2\Gamma}\left(-\frac{\rho^2}{4} + \Gamma z^2\right)A_{s0}, \quad \delta A_{y0} = 0. \tag{4.14}$$

The corresponding vector

$$\delta\mathbf{A}_f = \begin{pmatrix} 0 \\ \dfrac{1}{8(k^2-\Gamma^2)}A_{s0}\rho^2 \\ 0 \end{pmatrix} \tag{4.15}$$

is also a root:

$$\mathbf{M}_0\delta\mathbf{A}_f = \delta\mathbf{A}_{\Gamma 0}, \quad \mathbf{M}_0^3\delta\mathbf{A}_f = 0. \tag{4.16}$$

The last of the solutions of linearized Eqs. (4.7) required for the following analysis has the form [18, 22]

$$\delta A_{r0} = a(\rho) + b(\rho)z^2, \quad \delta A_{i0} = c(\rho)z + d(\rho)z^3, \quad \delta A_{y0} = 0. \tag{4.17}$$

After substituting (4.17) into (4.7), we find

$$d = A_{s0}, \quad c = -\frac{3}{\Gamma^2-k^2}A_{s0}\rho^2, \quad b = 3\frac{\partial A_{s0}}{\partial\Gamma}. \tag{4.18}$$

The function $a(\rho)$ is defined as the only finite axisymmetric solution of the equation [the conditions for solubility are satisfied because of the axial symmetry of the right-hand side (4.19)]

$$L_1 a = -\frac{6\Gamma^3}{\Gamma^2-k^2}A_{s0}\rho^2, \tag{4.19}$$

or

$$\frac{d^2a}{d\rho^2} + \frac{1}{\rho}\frac{da}{d\rho} - (\Gamma^2-k^2)a + 3k^2\frac{\varepsilon_2}{\varepsilon_0}A_{s0}^2 a = -\frac{6\Gamma^3}{\Gamma^2-k^2}A_{s0}\rho^2. \tag{4.20}$$

We do not require the specific form of the function $a(\rho)$. The perturbation vector corresponding to this solution is also a root:

$$\mathbf{M}_0\delta\mathbf{A}_a = \delta\mathbf{A}_f, \quad \mathbf{M}_0^4\delta\mathbf{A}_a = 0,$$

$$\delta\mathbf{A}_a = \begin{pmatrix} \dfrac{1}{48\Gamma^3}a(\rho) \\ 0 \\ 0 \end{pmatrix}. \tag{4.21}$$

These solutions exhaust the family of localized axisymmetric solutions of the linearized nonlinear Schrödinger equation with a zero eigenvalue.

Using the small nonparaxial parameter (2.9), we can write expansions of the matrix $\mathbf{M}$, the eigenvector $\delta\mathbf{A}_\varphi$ and the root vector $\delta\mathbf{A}_\Gamma$ in the form

$$\mathbf{M} = \mathbf{M}_0 + \mu^2\mathbf{M}_2 + \mu^4\mathbf{M}_4 + \ldots,$$
$$\delta\mathbf{A}_\varphi = \delta\mathbf{A}_\varphi^{(0)} + \mu^2\delta\mathbf{A}_\varphi^{(2)} + \ldots, \tag{4.22}$$
$$\delta\mathbf{A}_\Gamma = \delta\mathbf{A}_\Gamma^{(0)} + \mu^2\delta\mathbf{A}_\Gamma^{(2)} + \ldots$$

Terms with a zero index are determined by the nonlinear Schrödinger equation and are given in Eqs. (4.8), (4.9), and (4.11). The remaining terms of the expansion are obtained using the expansion for the field of a steady-state soliton determined in Section 3.

5. STABILITY AND OSCILLATIONS OF PERTURBED SOLITONS

We shall now find the eigenvalue of the matrix operator $\mathbf{M}$ which goes to zero in the limit $\mu \longrightarrow 0$ (nonparaxial soliton limit). For this we shall seek the eigensolution of the linearized equation (4.5) in the form

$$\delta\mathbf{A} = \mathbf{\Psi}(x, y)e^{\mu\gamma z/2\Gamma}. \tag{5.1}$$

Here we introduce the unknown eigenvalue

$$\mu\gamma = \mu\gamma_1 + \mu^2\gamma_2 + \mu^3\gamma_3 + \ldots \tag{5.2}$$

and the eigenvector $\mathbf{\Psi}$ which obey the following equation derived from (4.5)

$$\mathbf{M}\mathbf{\Psi} = \mu\gamma\mathbf{\Psi}. \tag{5.3}$$

The expansion of the eigenvector $\mathbf{\Psi}$ can be conveniently expressed in the form

$$\mathbf{\Psi} = \delta\mathbf{A}_\varphi + \mu\gamma\delta\mathbf{A}_\Gamma + \mu^2\mathbf{\Psi}_2 + \mu^3\mathbf{\Psi}_3 + \mu^4\mathbf{\Psi}_4 + \ldots. \tag{5.4}$$

The first two terms on the right-hand side (5.4) are determined by Eqs. (4.9) and (4.11). Substituting into (5.3) the expansions of the corresponding quantities in powers of the small nonparaxial parameter μ and equating terms of the same order with respect to this parameter, we find in the second order in μ [equations of lower orders are automatically satisfied given the choice made in (5.4)]

$$\mathbf{M}_0\mathbf{\Psi}_2 = \gamma_1^2\delta\mathbf{A}_{\Gamma 0}. \tag{5.5}$$

Taking into account (4.16), we find that

$$\mathbf{\Psi}_2 = \gamma_1^2\delta\mathbf{A}_f. \tag{5.6}$$

In the third order, we have

$$\begin{aligned}\mathbf{M}_0\mathbf{\Psi}_3 &= 2\gamma_1\gamma_2\delta\mathbf{A}_{\Gamma 0} + \gamma_1\mathbf{\Psi}_2\\ &= 2\gamma_1\gamma_2\delta\mathbf{A}_{\Gamma 0} + \gamma_1^3\delta\mathbf{A}_f,\end{aligned} \tag{5.7}$$

so that, taking into account (4.16) and (4.21), we have

$$\mathbf{\Psi}_3 = 2\gamma_1\gamma_2\delta\mathbf{A}_f + \gamma_1^3\delta\mathbf{A}_a. \tag{5.8}$$

Finally, in the fourth order we have

$$\begin{aligned}\mathbf{M}_0\mathbf{\Psi}_4 + \mathbf{M}_2\mathbf{\Psi}_2 &= \gamma_1^2\delta\mathbf{A}_{\Gamma 2}\\ &+ (\gamma_2^2 + 2\gamma_1\gamma_3)\delta\mathbf{A}_{\Gamma 0} + \gamma_1\mathbf{\Psi}_3 + \gamma_2\mathbf{\Psi}_2,\end{aligned} \tag{5.9}$$

or allowing or (5.6) and (5.8)

$$\begin{aligned}\mathbf{M}_0\mathbf{\Psi}_4 + \gamma_1^2\mathbf{M}_2\delta\mathbf{A}_f &= \gamma_1^2\delta\mathbf{A}_{\Gamma 2}\\ &+ (\gamma_2^2 + 2\gamma_1\gamma_3)\delta\mathbf{A}_{\Gamma 0} + 3\gamma_1^2\gamma_2\delta\mathbf{A}_f + \gamma_1^4\delta\mathbf{A}_a.\end{aligned} \tag{5.10}$$

We now introduce the transposed operator matrix $\mathbf{M}_0^\dagger$ and the conjugate three-dimensional vectors $\mathbf{A}_{\varphi 0}^\dagger$ and $\mathbf{A}_{\Gamma 0}^\dagger$ using the relationships

$$\mathbf{M}_0^\dagger = \begin{pmatrix} 0 & L_1 & 0\\ -L_0 & 0 & 0\\ 0 & 0 & iL_0\end{pmatrix}, \tag{5.11}$$

$$\mathbf{M}_0^\dagger\delta\mathbf{A}_{\varphi 0}^\dagger = 0, \quad \mathbf{M}_0^\dagger\delta\mathbf{A}_{\Gamma 0}^\dagger = \delta\mathbf{A}_{\varphi 0}^\dagger.$$

In explicit form we have

$$\delta\mathbf{A}_{\varphi 0}^\dagger = \begin{pmatrix} A_{s0}\\ 0\\ 0\end{pmatrix}, \quad \delta\mathbf{A}_{\Gamma 0}^\dagger = \begin{pmatrix} 0\\ \dfrac{\partial A_{s0}}{\partial\Gamma}\\ 0\end{pmatrix}. \tag{5.12}$$

We introduce the scalar product of the three-dimensional vectors (4.6) using the relationships

$$\begin{aligned}\langle\mathbf{U},\mathbf{V}\rangle &= \frac{1}{2\pi}\\ &\times\int_0^{2\pi}d\varphi\int_0^\infty d\rho\rho(U_1V_1 + U_2V_2 + U_3V_3),\end{aligned} \tag{5.13}$$

where we have for the complex (third) components

$$U_3V_3 = \mathrm{Re}U_3\mathrm{Re}V_3 + \mathrm{Im}U_3\mathrm{Im}V_3. \tag{5.14}$$

Then, as a result of the self-adjoint property of the Laplace operator for arbitrary vectors $\mathbf{U}$ and $\mathbf{V}$ the following identity is satisfied

$$\langle\mathbf{U},\mathbf{M}_0\mathbf{V}\rangle = \langle\mathbf{M}_0^\dagger\mathbf{U},\mathbf{V}\rangle. \tag{5.15}$$

We now multiply the three-dimensional vector $\mathbf{A}_{\varphi 0}^\dagger$ by the left- and right-hand sides of Eq. (5.10) and equate these scalar products. Here we use the following relationships for the matrix elements:

$$\langle\delta\mathbf{A}_{\varphi 0}^\dagger,\mathbf{M}_0\mathbf{\Psi}_4\rangle = \langle\mathbf{M}_0^\dagger\delta\mathbf{A}_{\varphi 0}^\dagger,\mathbf{\Psi}_4\rangle = 0, \tag{5.16}$$

$$\begin{aligned}\langle\delta\mathbf{A}_{\varphi 0}^\dagger,\delta\mathbf{A}_{\Gamma 0}\rangle &= \left\langle\begin{pmatrix} A_{s0}\\ 0\\ 0\end{pmatrix},\begin{pmatrix} \dfrac{1}{2\Gamma}\dfrac{\delta A_{s0}}{\partial\Gamma}\\ 0\\ 0\end{pmatrix}\right\rangle\\ &= \frac{1}{2\Gamma}\int_0^\infty A_{s0}\frac{\partial A_{s0}}{\partial\Gamma}\rho d\rho = 0,\end{aligned} \tag{5.17}$$

$$\begin{aligned}&\langle\delta\mathbf{A}_{\varphi 0}^\dagger,\delta\mathbf{A}_f\rangle\\ &= \left\langle\begin{pmatrix} A_{s0}\\ 0\\ 0\end{pmatrix},\begin{pmatrix} 0\\ \dfrac{1}{8(k^2-\Gamma^2)}A_{s0}\rho^2\\ 0\end{pmatrix}\right\rangle = 0,\end{aligned} \tag{5.18}$$

$$\begin{aligned}m_{\varphi,a} &= \langle\delta\mathbf{A}_{\varphi 0}^\dagger,\delta\mathbf{A}_a\rangle = \langle\mathbf{M}_0^\dagger\delta\mathbf{A}_{\Gamma 0}^\dagger,\delta\mathbf{A}_a\rangle\\ &= \langle\delta\mathbf{A}_{\Gamma 0}^\dagger,\mathbf{M}_0\delta\mathbf{A}_a\rangle = \langle\delta\mathbf{A}_{\Gamma 0}^\dagger,\delta\mathbf{A}_f\rangle\\ &= -\frac{1}{32\Gamma(\Gamma^2-k^2)}\int_0^\infty\frac{\partial A_{s0}^2}{\partial\Gamma}\rho^3 d\rho,\end{aligned} \tag{5.19}$$

$$m_{\varphi,\Gamma} = \langle\delta\mathbf{A}_{\varphi 0}^\dagger,\delta\mathbf{A}_{\Gamma 2}\rangle = \frac{1}{2\Gamma}\int_0^\infty A_{s0}\frac{\partial\delta A_{s0}}{\partial\Gamma}\rho d\rho, \tag{5.20}$$

and we also introduce the notation

$$m_2 = \langle \delta\mathbf{A}_{\varphi 0}^{\dagger}, \mathbf{M}_2 \delta\mathbf{A}_f \rangle. \quad (5.21)$$

We then obtain an equation to determine the square of the eigenvalue in the lowest approximation (see also [17])

$$\gamma_1^2[(m_2 - m_{\varphi,\Gamma}) - m_{\varphi,a}\gamma_1^2] = 0. \quad (5.22)$$

The two zero roots of this equation correspond to the symmetries with respect to the phase shift and the shift of the propagation constant conserved when allowance is made for the nonparaxial property. In accordance with (5.22) two nonzero eigenvalues are split off from these (as a result of a shift of the eigenvalue corresponding to the vectors $\delta\mathbf{A}_f$ and $\delta\mathbf{A}_a$ for the nonlinear Schrödinger equation). In order to determine these it is convenient to calculate the matrix elements appearing in (5.22) by going over to the functions $F_{0,1}(r)$ introduced earlier in the integrand expressions (see Fig. 1 and Appendix). We then finally obtain

$$\left(\frac{\tilde{\gamma}}{k}\right)^2 = \frac{1}{k^2}\left(\frac{\gamma_1}{2\Gamma}\right)^2 = -43.35\left(\frac{\Gamma - k}{k}\right)^3. \quad (5.23)$$

In accordance with (5.23), a weakly nonparaxial soliton is stable ($\Gamma > k$). However, the imaginary nature of the eigenvalue γ implies that an "internal mode" occurs whose field distribution as given by (5.4) is close to the soliton field (and is phase shifted by $\pi/2$). The longitudinal period of the oscillations of the perturbed field $2\pi/\tilde{\gamma}$ increases without bound in the paraxial soliton limit $\mu \longrightarrow 0$ ($\Gamma \longrightarrow k$). As a result of the weak radiation damping of these internal modes [23–27] it is difficult for a steady-state soliton to be established under these conditions (an anomalously long nonlinear medium is required).

6. CONCLUSIONS

We have demonstrated for the first time that nonparaxial solitons of electromagnetic radiation are stable in a medium with a Kerr nonlinearity. These solitons may occur at the final stage of self-focusing of supercritical-power radiation. Since the maximum intensity of these weakly nonparaxial solitons is low, the Kerr nonlinearity will be the dominant mechanism (no competing mechanisms of nonlinearity exist at low intensities). We reemphasize that the fields of these solitons do not possess axial symmetry and the polarization structure of the radiation strictly corresponds to elliptic polarization which varies over the cross section. It is important to allow for the vector nature and the nontrivial polarization structure of electromagnetic radiation solitons since their scalar description cannot be quantitative. This approach can not only demonstrate the stability but can also be used to determine the corresponding quantitative characteristic (the eigenvalue γ). This value characterizes the internal modes of these solitons, i.e., the natural modes of small perturbations in an effective light guide induced by the "strong field" of a soliton in a nonlinear medium.

The results obtained here for weakly nonparaxial optical solitons serve as an additional argument to support the "needles of light" identified in [15], i.e., strongly nonparaxial solitons of width less than the wavelength of the radiation in a linear medium. We also note that since the nonlinear Schrödinger equation describes an extremely wide range of phenomena of various physical nature, this approach to analyze the perturbed Schrödinger equation may not be confined merely to optical problems.

ACKNOWLEDGMENTS

The authors thank D.V. Skryabin for important references and comments. This work was supported by the International Scientific-Technical Center (grant no. 666), the Russian Foundation for Basic Research (project no. 98-02-18202), and INTAS (grant no. 1997-581).

APPENDIX

We shall calculate the matrix elements appearing in (5.22) which can be reduced to single integrals (over the radial coordinate) of the functions $F_{0,1}(r)$ and their derivatives. Graphs of the functions are plotted in Fig. 1 and their integrals were calculated numerically:

$$m_{\varphi,\Gamma} = \frac{1}{2\Gamma}\int_0^\infty A_{s0}\frac{\partial \delta A_{s0}}{\partial \Gamma}\rho\, d\rho = \frac{1}{2k^4}\frac{\varepsilon_0}{\varepsilon_2}p_2,$$

$$p_2 = \int_0^\infty \left[3F_0(r)F_1(r) + F_0(r)\frac{dF_1(r)}{dr}\right] r\, dr = 5.821,$$

$$m_{\varphi,a} = -\frac{1}{32\Gamma(\Gamma^2 - k^2)}\int_0^\infty \frac{\partial A_{s0}^2}{\partial \Gamma}\rho^3 d\rho = -\frac{\varepsilon_0}{\varepsilon_2}\frac{1}{k^2}$$

$$\times \frac{1}{32(\Gamma^2 - k^2)}\frac{\partial}{\partial \Gamma}\left[(\Gamma^2 - k^2)\int_0^\infty F_0^2(\sqrt{\Gamma^2 - k^2}\rho)\rho^3 d\rho\right]$$

$$= \frac{\varepsilon_0}{\varepsilon_2}\frac{1}{k^2}\frac{1}{16(\Gamma^2 - k^2)^2}p_3,$$

$$p_3 = \int_0^\infty F_0^2(r) r^3 dr = 2.211,$$

$$m_2 = m_{2s} + m_{2k} + m_{2z},$$

$$m_{2s} = \langle \delta\mathbf{A}^{+}_{\varphi 0}, \mathbf{M}_{2s}\delta\mathbf{A}_f\rangle = \frac{1}{16}\frac{\varepsilon_2}{\varepsilon_0}\frac{1}{\Gamma^2 - k^2}$$

$$\times\int_0^\infty A_{s0}\left[\left(\frac{dA_{s0}}{d\rho}\right)^2 A_{s0}\rho^2 + \left(\frac{d}{d\rho}+\frac{1}{\rho}\right)(A_{s0}\rho^2)\right]\rho d\rho$$

$$= \frac{1}{16k^4}\frac{\varepsilon_0}{\varepsilon_2}p_s,$$

$$p_s = \int_0^\infty\left[\left(\frac{dF_0(r)}{dr}F_0(r)r\right)^2\right.$$

$$\left. + F_0(r)\left(\frac{d}{dr}+\frac{1}{r}\right)\left(\frac{dF_0^2(r)}{dr}F_0(r)r^2\right)\right]rdr = -1.608,$$

$$m_{2k} = \langle \delta\mathbf{A}^{+}_{\varphi 0}, \mathbf{M}_{2k}\delta\mathbf{A}_f\rangle = \frac{k^2}{4(\Gamma^2 - k^2)}$$

$$\times\frac{\varepsilon_2}{\varepsilon_0}\int_0^\infty A_{s0}^3\delta A_{sx0}\rho^3 d\rho = \frac{1}{4k^4}\frac{\varepsilon_0}{\varepsilon_2}p_k,$$

$$p_k = \int_0^\infty F_0^{(3)}(r)F_1(r)r^3 dr = 1.926,$$

$$m_{2z} = \langle \delta\mathbf{A}^{+}_{\varphi 0}, \mathbf{M}_{2z}\delta\mathbf{A}_f\rangle = -\frac{1}{32(\Gamma^2 - k^2)}$$

$$\times\int_0^\infty d\rho\rho A_{s0}\left\{\frac{1}{k^2}[\Delta_\perp - (\Gamma^2 - k^2)]\right.$$

$$\times\left[\Delta_\perp - (\Gamma^2 - k^2) + k^2\frac{\varepsilon_2}{\varepsilon_0}A_{s0}^2\right](A_{s0}\rho^2)$$

$$+\frac{\varepsilon_2}{\varepsilon_0}\left[3A_{s0}^2\Delta_\perp(A_{s0}\rho^2) - 2A_{s0}^2\rho^2\Delta_\perp A_{s0} - (\Gamma^2 - k^2)A_{s0}^3\rho^2\right.$$

$$\left.\left. + k^2\frac{\varepsilon_2}{\varepsilon_0}A_{s0}^5\rho^2\right]\right\} = -\frac{1}{32k^4}\frac{\varepsilon_0}{\varepsilon_2}p_z,$$

$$p_z = \int_0^\infty drrF_0(r)[\Delta_r - 1][\Delta_r - 1 + F_0^2(r)](F_0(r)r^2)$$

$$+3\int_0^\infty drrF_0^3(r)\Delta_r(F_0(r)r^2) - 2\int_0^\infty drr^3F_0^3(r)\Delta_r(F_0(r))$$

$$-\int_0^\infty drr^3F_0^4(r) + \int_0^\infty drr^3F_0^6(r) = 14.899,$$

$$\Delta_r = \frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr}.$$

After substituting these values into (5.22) we obtain (5.23).

REFERENCES

1. S. N. Vlasov and V. I. Talanov, *Self-Focusing of Waves* (Inst. Prikl. Fiz. Ross. Akad. Nauk, Nizhni Novgorod, 1997).
2. G. A. Askar'yan, Zh. Éksp. Teor. Fiz. **42**, 1567 (1962) [Sov. Phys. JETP **15**, 1088 (1962)].
3. N. A. Tikhonov, Dokl. Akad. Nauk SSSR **231**, 592 (1976) [Sov. Phys. Dokl. **21**, 663 (1976)].
4. N. Akhmediev, A. Ankiewicz, and J. M. Soto-Crespo, Opt. Lett. **18**, 411 (1993).
5. J. M. Soto-Crespo and N. Akhmediev, Opt. Commun. **101**, 223 (1993).
6. A. P. Sheppard and M. Haelterman, Opt. Lett. **23**, 1820 (1998).
7. D. I. Abakarov, A. A. Akopov, and S. I. Pekar, Zh. Éksp. Teor. Fiz. **52**, 463 (1967) [Sov. Phys. JETP **25**, 303 (1967)].
8. V. M. Eleonskiĭ and V. P. Silin, Pis'ma Zh. Éksp. Teor. Fiz. **13**, 167 (1971) [JETP Lett. **13**, 117 (1971)].
9. V. M. Eleonskiĭ, L. G. Oganes'yants, and V. P. Silin, Zh. Éksp. Teor. Fiz. **62**, 81 (1972) [Sov. Phys. JETP **35**, 44 (1972)].
10. V. M. Eleonskiĭ, L. G. Oganes'yants, and V. P. Silin, Zh. Éksp. Teor. Fiz. **63**, 532 (1972) [Sov. Phys. JETP **36**, 282 (1973)].
11. D. A. Kirsanov and N. N. Rozanov, Opt. Spektrosk. **87**, 423 (1999) [Opt. Spectrosc. **87**, 390 (1999)].
12. V. M. Eleonskiĭ, N. E. Kulagin, D. V. Turaev, and L. P. Shil'nikov, Dokl. Akad. Nauk SSSR **309**, 848 (1989) [Sov. Phys. Dokl. **34**, 1052 (1989)].
13. V. M. Eleonsky, V. G. Korolev, N. E. Kulagin, and L. P. Shil'nikov, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 385 (1993).
14. S. Chi and Qi Guo, Opt. Lett. **20**, 1598 (1995).
15. V. E. Semenov, N. N. Rozanov, and N. V. Vysotina, Zh. Éksp. Teor. Fiz. **116**, 458 (1999) [JETP **89**, 243 (1999)].
16. N. N. Rozanov, Opt. Spektrosk. **89** (3), 460 (2000) [Opt. Spectrosc. **89**, 422 (2000)].
17. V. E. Zakharov and A. M. Rubenchik, Zh. Éksp. Teor. Fiz. **65**, 997 (1973) [Sov. Phys. JETP **38**, 494 (1973)].
18. A. G. Vladimirov and N. N. Rozanov, Opt. Spektrosk. **89** (5), 795 (2000) [Opt. Spectrosc. **89**, 731 (2000)].
19. N. G. Vakhitov and A. A. Kolokolov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **16**, 1020 (1973).

20. V. I. Talanov, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 303 (1970) [JETP Lett. **11**, 199 (1970)].
21. E. A. Kuznetsov and S. K. Turitsyn, Phys. Lett. A **112A**, 273 (1985).
22. V. M. Malkin and E. G. Shapiro, Physica D (Amsterdam) **53**, 25 (1991).
23. V. E. Zakharov, V. V. Sobolev, and V. S. Synakh, Zh. Éksp. Teor. Fiz. **60**, 136 (1971) [Sov. Phys. JETP **33**, 77 (1971)].
24. D. E. Pelinovsky, V. V. Afanasjev, and Yu. S. Kivshar, Phys. Rev. E **53**, 1940 (1996).
25. C. Etrich, U. Peschel, F. Lederer, *et al.*, Phys. Rev. E **54**, 4321 (1996).
26. D. E. Pelinovsky, Yu. S. Kivshar, and V. V. Afanasjev, Physica D (Amsterdam) **116**, 121 (1998).
27. N. N. Rozanov, D. A. Kirsanov, P. I. Krepostnov, and V. O. Popov, Opt. Zh. **67**, 28 (2000).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1141–1146.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1319–1324.

PLASMA, GASES

On the Energy Flux of Stationary Electromagnetic Waves in Anisotropic Dissipative Media with Spatial Dispersion

M. D. Tokman[a, *], E. Westerhof[b, **], and M. A. Gavrilova[a, *]**

[a] *Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia*
**e-mail: tokman @ appl.sci-nnov.ru*
****e-mail: mina @ appl.sci-nnov.ru*
[b] *FOM-Instituute voor Plasmafysica "Rijnhuizen," Trilateral Euregio Cluster, Associate EURATOM-FOM, 3430 BE Nieuwegein, the Netherlands*
***e-mail: E. Westerhof @rijnh.nl*

Received December 21, 1999

Abstract—The special features of the propagation of electromagnetic waves in gyrotropic medium with dispersion and resonant dissipation (specifically, in a magnetoactive plasma) are studied. Even though the anti-Hermitian components of the permittivity tensor are substantial in magnitude, weakly damped waves can exist in such media. However, the well-known phenomenological expression for the energy flux of waves in a medium with spatial dispersion is inapplicable for them. A theory extending this expression to the case studied is developed. The modified expression for the energy flux is used to construct the Hamiltonian for the ray optics of such media. © *2000 MAIK "Nauka/Interperiodica".*

A well-known complexity of the theory of electromagnetic waves is the absence of phenomenological expressions for the group velocity, energy density, and energy flux of waves in dissipative, dispersive media [1–3]. Interest in this problem has increased recently in connection with the problem of describing the propagation of resonance electromagnetic radiation in a high-temperature magnetoactive plasma. A characteristic feature of the propagation of electromagnetic radiation in the cyclotron resonance bands is the existence of weakly damped waves even in the region where the anti-Hermitian components of the permittivity tensor are substantial [4–6]. Thus we are dealing with the case where for

$$\left|\varepsilon_{pm}^{aH}\right| \sim \left|\varepsilon_{pm}^{H}\right| \tag{1}$$

the dispersion equation

$$D(\omega, \mathbf{k}) = \det\|D_{pm}\| = 0 \tag{2}$$

for real ω possesses a solution with weak damping at wavelengths

$$|\mathrm{Re}\,\mathbf{k}| \gg |\mathrm{Im}\,\mathbf{k}|. \tag{3}$$

(Here

$$D_{pm} = \delta_{pm}k^2 - k_p k_m - \frac{\omega^2}{c^2}\varepsilon_{pm}(\omega, \mathbf{k}),$$

$\varepsilon_{pm}^{aH,H}$ are the anti-Hermitian and Hermitian components of the permittivity tensor ε_{pm} of the medium, and the symbols ω and $\mathbf{k}$ have their standard meanings.)

The weakness of the absorption could be due to the special features of the polarization of normal waves [4–6], and even in spite of the conditions (1) the projection of the active component of the current in the medium on the direction of the hf field vector is a small parameter:[1]

$$\frac{\left|E_p^*\varepsilon_{pm}^{aH}E_m\right|}{\left|E_p^*\varepsilon_{pm}^{H}E_m\right|} = \mu \ll 1 \tag{4}$$

(here $\mathbf{E}$ is a vector of the complex amplitude of hf field).

However, even though the absorption of waves is relatively weak a number of difficulties arise in describing waves when the permittivity tensor is strongly anti-Hermitian.

(i) The standard expression (i.e., obtained for vanishingly small "non-Hermitian component" of the permittivity tensor) for the energy flux $\mathbf{S}_H$ of waves[2]

$$\mathbf{S}_H = \frac{c^2}{16\pi\omega}E_p^*E_m\frac{\partial D_{pm}^H}{\partial \mathbf{k}}, \tag{5}$$

is no longer correct, since the derivative of the small complex quantity $E_p^*E_m\varepsilon_{pm}^{aH}$ with respect to the wave vector is not always sufficiently small (see [5]).

[1] In a magnetoactive plasma the condition (4) holds for nonlongitudinal propagation of electromagnetic waves relative to the magnetic field. In this case the circularly polarized component of the field, rotating in the direction of Larmor rotation of the electrons or ions—near the electronic or ionic cyclotron resonances, respectively—is suppressed (depressed).

[2] The vector $\mathbf{S}_H$ is the sum of the Poynting vector and the energy flux excited in the medium by hf field [1, 2, 7].

1063-7761/00/9106-1141$20.00

(ii) The standard use of the dispersion function $D(\omega, \mathbf{k}, \mathbf{r})$ as the Hamiltonian for ray optics[3] engenders rays in a complex space—by virtue of the strong complexity of the derivatives of the function $D(\omega, \mathbf{k})$ with respect to its arguments—even though Eq. (2) possesses weakly damped solutions (in his connection see [5, 6, 8]).

Physically, however, it seems almost obvious that for weakly damped waves the difficulties in obtaining an expression for the energy flux and the (directly related) difficulties of constructing a geometric optics, even under the conditions (1), are purely formal. In plasma physics the most commonly used method for eliminating these difficulties is to use a purely Hermitian permittivity tensor of a cold collisionless plasma [5, 6]. Then, the energy flux reduces to the standard Poynting vector. However, this approach neglects the characteristic behavior of waves near the cyclotron resonance band, where, as indicated in [9], the spatial dispersion induced components of the energy flux can be determining.[4] The method of investigating the propagation of waves beams in a plane-layered medium was used in [11, 12]. In this method the wave field is expanded in WKB modes satisfying the conditions (3). It follows from the analysis performed in these works that the trajectory of beams in a plane-layered medium can be given by a real ray Hamiltonian $H(\mathbf{r}, \mathbf{k}, \omega)$, if for real ω and $\mathbf{k}$ the condition $H(\mathbf{r}, \mathbf{k}, \omega) = 0$ corresponds to vectors $\mathbf{k}$ coinciding with the real part of the corresponding solution of the dispersion equation (2) with real frequencies.[5] The numerical analysis performed in [11] demonstrated that the direction of propagation of waves beams near the cyclotron resonance bands can differ substantially from the direction of the standard energy flux in a medium with spatial dispersion, given by Eq. (5). In his connection an attempt to obtain an expression for the energy flux of weakly damped waves that is valid at the finite anti-Hermitian component of the permittivity tensor is interesting. It turns out that this problem can be greatly simplified by using the method proposed in [12] for analyzing the energy characteristics of a wave field in dissipative anisotropic media, based on expanding a smoothly nonuniform stationary field in a spectrum of normal waves. We believe that this approach makes it possible to introduce phenomenologically the definition of the energy flux of a stationary electromagnetic wave field; such a definition is the main content of this paper.

This paper is organized as follows. In Section 1 the Poynting relation, which relates an hf field and the currents which it excites in a dissipative and isotropic medium, reduces to an explicitly divergence form, which makes it possible to determine uniquely the direction of transfer of the radiation intensity. In Section 2 the situations where the modified expression for the energy flux differs substantially from the standard relation (5) or, conversely, transforms into it are discussed; the physical meaning of the results obtained is clarified; and, the possibility of using the results obtained to construct a geometric optics of the corresponding media is discussed.

Although the formalism developed in the present paper is valid for an arbitrary anisotropic medium, the exposition is primarily associated with the theory of electromagnetic waves in a magnetoactive plasma.

1. Let a stationary high-frequency field with the wave structure

$$\tilde{\mathbf{E}} = \mathbf{E}(\mathbf{r})\exp(i\mathbf{k}_0 \cdot \mathbf{r} - i\omega t) \tag{6}$$

(here $\mathbf{E}(\mathbf{r})$ is a smoothly nonuniform complex amplitude, and the function $\mathbf{E}(\mathbf{r})$ also takes account of possible wave damping) be excited in a uniform anisotropic medium with dispersion and dissipation. The condition that the amplitude varies "slowly" has the form

$$\frac{1}{k_0}\frac{1}{|E|}\left|\frac{\partial E}{\partial \mathbf{r}}\right| = \varepsilon \ll 1. \tag{7}$$

The existence of a small parameter ε requires quite weak absorption at the wavelength and a relatively narrow angular Fourier spectrum of the wave field. The Poynting theorem for a field of the form (6)

$$\operatorname{div}\left(\frac{c^2}{8\pi\omega}\operatorname{Re}[\mathbf{E}^* \times [\mathbf{k}_0 \cdot \mathbf{E}]]\right) + \frac{1}{2}\operatorname{Re}(\mathbf{j} \cdot \mathbf{E}^*) = 0 \tag{8}$$

(here $\mathbf{j}$ is the complex amplitude of the current excited in the medium) still does not determine explicitly the direction of transfer of the intensity of the wave field in the medium, since the possible dependence of the current $\mathbf{j}$ on the spatial derivatives of hf field can lead to the appearance in Eq. (8) of additional divergence terms. Specifically, in a medium with no dissipation under the conditions (7) it follows from Eq. (8) [1, 7] that

$$\operatorname{div}\mathbf{S}_H = 0, \tag{9}$$

where the vector $\mathbf{S}_H$ is determined by the relation (5).

The situation becomes more complicated if the medium is dissipative, i.e., the permittivity tensor contains Hermitian and anti-Hermitian components.

In first place, if the anti-Hermitian tensor ε_{pm}^{aH} depends on the wave vector $\mathbf{k}$, then according to [5] a relation with additional spatial derivatives follows from equation (8):

$$\operatorname{div}\mathbf{S}_H + q + Q = 0, \tag{10}$$

where

$$Q = \frac{ic^2}{8\pi\omega}(D_{pm}^{aH})\big|_{\mathbf{k}=\mathbf{k}_0} E_p^* E_m$$

[3] The dependence of D on $\mathbf{r}$ occurs in a nonuniform medium.

[4] The modification of the theory, based on the "cold" plasma approximation, is especially needed for waves propagating "along tangents" to the resonance surfaces [10].

[5] The effect of the finite value of $\operatorname{Im}\mathbf{k}$ on the expression for $\operatorname{Re}\mathbf{k}$ can be or need not be taken into account—depending on the accuracy required.

is the standard expression for the intensity of the losses,

$$q = \frac{c^2}{16\pi\omega}\left(\frac{\partial D_{pm}^{aH}}{\partial \mathbf{k}}\right)\Bigg|_{\mathbf{k}=\mathbf{k}_0}\left(E_p^*\frac{\partial E_m}{\partial \mathbf{r}} - \frac{\partial E_p^*}{\partial \mathbf{r}}E_m\right)$$

is an additional term due to the fact that dissipation and spatial dispersion are taken into account simultaneously.

In second place, in an anisotropic dissipative medium, even in the absence of spatial dispersion (i.e., when $q = 0$), the direction of transfer of the intensity of the hf can, in principle, be different from the direction of the vector $\mathbf{S}_H$. The problem is that in an anisotropic medium the dissipative term Q depends strongly on the polarization of hf field, and the polarization in turn depends on the intensity distribution nonlocally, specifically—under the condition (7)—the polarization can depend on the derivatives of the intensity with respect to spatial coordinates[6] and engender additional divergence terms in the Q term.

Let us substitute the hf field (6) in the form of a composition of normal waves of the same type:

$$\mathbf{E}(\mathbf{r}) = \int \mathbf{e}(\lambda)A_k(\mathbf{k}_0 + \Delta\mathbf{k})\delta(\lambda)e^{i\Delta\mathbf{k}\cdot\mathbf{r}}d\Delta\mathbf{k}. \qquad (11)$$

Here $\lambda(\mathbf{k}_0 + \Delta\mathbf{k})$ is the eigenvalue, corresponding to the selected normal mode, of tensor $D_{pm}(\mathbf{k}_0 + \Delta\mathbf{k})$; $\mathbf{k} = \mathbf{k}_0 + \Delta\mathbf{k}$ is the wave vector; $\delta(\lambda)$ is the Dirac function that "selects" the values of $\mathbf{k}$ that satisfy the dispersion equation; $\mathbf{e}(\lambda)$ is the corresponding eigenvector (in the normalization $|\mathbf{e}|^2 = 1$), which is the polarization vector of the given normal mode[7] as $\lambda_i \longrightarrow 0$; A_k is a weighting function of the spectral distribution.

Taking into account the "narrowness," dictated by the condition (7), of the spatial Fourier spectrum ($\Delta k \ll k_0$), the following expression can be obtained from Eq. (11) (see also [12]):

$$\mathbf{E}(\mathbf{r}) = \mathbf{e}(\mathbf{k}_0)A(\mathbf{r}) - i\frac{\partial A}{\partial r_n}\left(\frac{\partial \mathbf{e}}{\partial k_n}\right)\Bigg|_{\mathbf{k}=\mathbf{k}_0}, \qquad (12)$$

where $\mathbf{e}(\mathbf{k}_0)$ is a unit polarization vector (eigenvector) for the "central" value of the wave vector $\mathbf{k} = \mathbf{k}_0$, the quantity $A(\mathbf{r})$ is an effective scalar amplitude of the field,[8] determined by the relation

$$A(\mathbf{r}) = \int A_k(\mathbf{k}_0 + \Delta\mathbf{k})\delta(\lambda)e^{i\Delta\mathbf{k}\cdot\mathbf{r}}d\Delta\mathbf{k}. \qquad (13)$$

We note that the second term on the right-hand side of Eq. (12) describes, specifically, the possible influence of dissipation on the polarization of the normal modes of an anisotropic medium.

We now substitute Eq. (12) into Eq. (10), representing the scalar complex amplitude $A(\mathbf{r})$ in the form

$$A = I^{1/2}\exp(i\phi),$$

where I is a positive real quantity. The result is

$$\operatorname{div}\mathbf{S} + W + \tilde{Q} = 0, \qquad (14)$$

where

$$\tilde{Q} = \frac{ic^2}{8\pi\omega}I\left[\left(1 + \frac{\partial\phi}{\partial r_p}\frac{\partial}{\partial k_p}\right)e_m^* e_n E_{mn}^{aH}\right]\Bigg|_{\mathbf{k}=\mathbf{k}_0} \qquad (15a)$$

$$\approx \frac{ic^2}{8\pi\omega}I(e_m^* e_p D_{mp})\big|_{\mathbf{k}=\mathbf{k}_0+\nabla\phi} \qquad (15b)$$

is the modified expression for the power losses (taking account of the difference of the "local" value of the wave vector from the "central" value $\mathbf{k}_0$),

$$W = \frac{c^2}{16\pi\omega}\frac{\partial I}{\partial \mathbf{r}}\left[D_{pm}^{aH}\left(\frac{\partial e_p^*}{\partial \mathbf{k}}e_m - e_p^*\frac{\partial e_m}{\partial \mathbf{k}}\right)\right]\Bigg|_{\mathbf{k}=\mathbf{k}_0.} \qquad (16)$$

Substituting the relations

$$D_{pm}^{aH}e_m = -D_{pm}^{H}e_m, \quad e_p^* D_{pm}^{aH} = e_p^* D_{pm}^{H}, \qquad (17)$$

which follow from Maxwell's equations, Eq. (14) can be reduced to an explicitly divergence form:

$$\operatorname{div}\mathbf{S}_M + \tilde{Q} = 0, \qquad (18)$$

where

$$\mathbf{S}_M = \frac{c^2}{16\pi\omega}I\frac{\partial}{\partial \mathbf{k}}(D_{pm}^{H}e_p^* e_m)\big|_{\mathbf{k}=\mathbf{k}_0} \qquad (19)$$

is the modified expression for the energy flux.

Concluding this section, we note that the final relation (18) corresponds, generally speaking, to an expansion in the first powers of the small parameters ε and μ. Specifically, Eqs. (17) for the real quantities $\mathbf{k} = \mathbf{k}_0$ can be used with this degree of accuracy. Likewise, the field intensity I, $I = |\mathbf{E}|^2$, appearing in the expansions (15) and (16) can be used in the same approximation. The second term on the right-hand side of Eq. (15a) is, at first glance, of the order of $\varepsilon\mu$. However, in media with

[6] The simplest example where the intensity distribution influences the polarization of the wave field is a "vacuum" plane-parallel beam with TM polarization: the narrower the beam, the stronger the longitudinal components of electric field is.

[7] Formally, the tensor D_{pm} has three eigenvalues $\lambda_i(\mathbf{k})$ and correspondingly three eigenvectors $\mathbf{e}_i(\lambda)$ (here $i = 1, 2, 3$). The number i corresponding to the desired normal waves in the limit $\lambda_i \longrightarrow 0$ is chosen.

[8] The condition (7) makes it possible to interpret the quantity $|A|^2$ as the intensity and the quantity $\arg A$ as the phase, determining the "local" value of the wave vector $\mathbf{k}(\mathbf{r}) = \mathbf{k}_0 + \nabla(\arg A)$. The relation $\int_\infty |A|^2\, d\mathbf{r} = \int_\infty |\mathbf{E}|^2\, d\mathbf{r}$ is exact.

resonant dissipation the derivative of the small "dissipative term" $e_m^* D_{mp}^{aH} e_p$ with respect to the wave vector $\mathbf{k}$ can be quite large: the smallness of the parameter μ still does not guarantee that the corresponding derivatives will be small. The following circumstance is also important. The use of differentiation of the "complex-conjugate" vectors $\mathbf{e}^*\mathbf{k}$ with respect to the argument $\mathbf{k}$ in the expansions (15a), (16), and (19) is determined, strictly speaking, only if the imaginary part of $\mathbf{k}$ is neglected; this assumption is also valid on the basis of the indicated expansion in small parameters.

2. The expression (19) differs from the standard energy flux (5) by the fact that differentiation of the polarization vectors with respect to the wave vector is assumed in it. For a purely Hermitian permittivity, when

$$D_{mp}^{H} e_m = 0, \tag{20}$$

the modified energy flux vector $\mathbf{S}_M$ automatically transforms into the standard flux $\mathbf{S}_H$.[9] Although only the Hermitian part of the tensor D_{pm} appears formally in Eq. (19), the antihermitian components ε_{mp}^{aH} influence the polarization vectors appearing in the expression (19).

It is evident from Eq. (19) that if the effect of dissipation on the polarization of the hf field is sufficiently weak [i.e., the condition (20) is approximately satisfied even when the finite values of ε_{mp}^{aH} are taken into account], then the standard expression (5) for the energy flux can be used to describe the propagation of the waves—this result was also obtained in [12] for the kinetic model of a magnetoactive plasma—with transverse propagation of ordinary and extraordinary waves at the first and second harmonics of the gyrofrequency. In general, however, the correction to the standard quantity $\mathbf{S}_H$ is of the order of

$$|\mathbf{S}_H - \mathbf{S}_M| \sim \frac{c^2 k}{\omega} I \frac{\operatorname{Im}\mathbf{k}}{\delta k}, \tag{21}$$

where $\operatorname{Im}\mathbf{k}$ is the absorption coefficients, δk is the characteristic scale of variation of the function $\mathbf{e}_\mathbf{r}(\lambda(\mathbf{k}))$ in $\mathbf{k}$ space, and $\mathbf{e}_\mathbf{r}$ is the resonance[10] (i.e., determining the absorption) component of the polarization vector of the wave.

Thus, as expected, the corresponding correction is important near the absorption line when the dependence of the polarization on the wave vector is sharp.

As an example we shall consider the quasitransverse propagation of an ordinary wave in a collisionless magnetoactive plasma. Near the electron-cyclotron resonance with

$$1 \gg \frac{k_\parallel}{k_\perp} \gtrsim \beta_T \tag{22}$$

(here $k_{\parallel,\perp}$ are the longitudinal and transverse (relative to the magnetic field $\mathbf{H}_0$) components of the wave vector and β_T is the ratio of the thermal velocity to the velocity of light) we can obtain from the expressions presented in, e.g., [2, 4, 13], for the tensor ε_{pm} that the ratio of the "dissipative" correction forms with the Poynting vector the ratio

$$\beta_T/(k_\parallel/k_\perp)^2,$$

i.e., it can be fundamental. The "correction" to the Poynting vector is oriented primarily along $\mathbf{H}_0$ and is of the same order of magnitude as the "nondissipative" correction due to the spatial dispersion. The fact that near the absorption line the spatial dispersion and dissipation influence wave propagation in the same order is very natural.

In the cold-plasma approximation with collisions (see [2, 13]) we can obtain the estimate

$$|S_H - S_M| \sim (c^2 k/\omega^2) I\nu,$$

where ν is the effective collision frequency. In this case the correction is proportional to be "central" small parameter ν/ω.

Although the relation (18) uniquely determines the direction of transfer of the intensity of the hf field, the existence of dissipation, of course, makes it difficult to give a physical interpretation of the formally introduced vector $\mathbf{S}_M$. Even with the additional assumptions that the system (medium + field) is closed and the processes in the medium are stationary, the difference between the vector $\mathbf{S}_M$ and the Poynting vector determines only part of the total energy flux excited in the medium by the hf field. The complete information about such fluxes under the condition $\varepsilon_{pm}^{aH} \neq 0$ can be obtained, naturally, only by analyzing a correct model of the medium. Such an analysis has been performed in [12] for a kinetic model of a magnetoactive plasma. In this work the transverse propagation of ordinary and extraordinary waves at the first and second harmonics of the gyrofrequency, when Eqs. (20) determine the polarization vectors to high degree of accuracy and the standard relation (5) is valid for the energy flux of the waves, was studied. It was found that an additional (excited by the field) energy flux of particles, which is directed transverse to the magnetic field and transverse to the "field" flux vector $\mathbf{S}_M$, is formed in dissipation zone:

$$S_\perp = \frac{i\omega}{16\pi} e_p^* e_m \frac{\partial \varepsilon_{pm}^{aH}}{\partial k_\perp} I. \tag{23}$$

[9] For vanishingly small non-Hermitian component of the permittivity tensor, Eqs. (5) and (19) are also identical. This case is examined in [15].

[10] In a magnetoactive plasma this is one of the circularly polarized components.

The flux $S_\perp$ leads to a definite shift of the energy-release zone relative to the region of localization of the hf field, but it does not affect the direction of transfer of its intensity (see [12]). It is obvious that such details always remain outside the framework of the formal phenomenological derivation of Eqs. (15), (18), and (19).

The expression for the energy flux (19) can be represented in another form. Since the quantity $e_p^* D_{pm}^H e_m$ is the real part of the eigenvalue of the tensor D_{pm}, it follows from Eq. (19) that

$$\mathbf{S}_M = \frac{c^2}{8\pi\omega} I \frac{\partial}{\partial \mathbf{k}} (\mathrm{Re}\,\lambda)|_{\mathbf{k}=\mathbf{k}_0}. \quad (24)$$

In this form the expression for $\mathbf{S}_M$ can also be obtained from other, more graphic, considerations. Let us consider the plane wave

$$\mathbf{E} = \mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{r}}.$$

Let the relation

$$\lambda(\mathbf{k}) = 0 \quad (25)$$

be the dispersion equation for a given normal mode. For $\mathrm{Re}\,\mathbf{k} = \mathbf{k}_0$ and $\mathrm{Im}\,\mathbf{k} = \kappa$ with $|\kappa| \ll |k_0|$ we have from Eq. (25)

$$\kappa \left(\frac{\partial \mathrm{Re}\,\lambda}{\partial \mathbf{k}} \right)\Bigg|_{\mathbf{k}=\mathbf{k}_0} + \mathrm{Im}\,\lambda(\mathbf{k}_0) = 0. \quad (26)$$

Since

$$\nabla |E_0|^2 = -2\kappa |E_0|^2, \quad i\,\mathrm{Im}\,\lambda = e_p^* D_{pm}^{aH} e_m,$$

we obtain from Eq. (26) a balance relation similar to Eq. (18) (true, for the particular case of a plane wave).

The quantity $\mathrm{Re}\,\lambda$ can be used as a real Hamiltonian of ray optics, since we have proved that the energy flux vector and the derivative $\partial \mathrm{Re}\,\lambda/\partial \mathbf{k}$ are identically directed even in the resonance absorption region, where the derivatives of $\lambda = \mathrm{Re}\,\lambda + i\,\mathrm{Im}\,\lambda$ with respect to $\mathbf{k}_0$ can be strongly complex functions.

The numerical calculations presented in [14] demonstrate that the direction of the quantity $\partial \mathrm{Re}\,\lambda/\partial \mathbf{k}$ is the same as the direction of propagation of wave beams in a nonuniform magnetoactive plasma even when the direction of the standard energy flux (5) differs substantially from them.

CONCLUSIONS

In this paper we examined the energy relations for a stationary wave field in an anisotropic medium with dissipation and spatial dispersion. This question was found to be closely related to the construction of a stationary geometric optics of such media. We showed that if the properties of the polarization of electromagnetic waves guarantee that wave damping is weak even when the Hermitian and anti-Hermitian components of the permittivity tensor are of the same order of magnitude, then the condition $|\varepsilon_{pm}^{aH}| \sim |\varepsilon_{pm}^{H}|$ in itself is not an obstacle for a phenomenological (i.e., not based on a model of the medium) calculation of the energy flux vector.

It has not been ruled out that the approach developed here will make it possible to introduce correctly for nonstationary waves an expression for the energy density of the waves and their group velocity in such media. However, this is a more complicated problem,[11] which is largely of only methodological interest.[12]

It is more important to investigate the transfer of the intensity of an hf field taking account of the modification of the permittivity operator by the nonuniformity of the medium (see [13]). However, the fruitfulness of a general phenomenological approach in this problem is by no means certain, since the model of the medium strongly determines the influence of the nonuniformity on the form of the nonlocal response of the medium in an hf field.

ACKNOWLEDGMENTS

We thank V.E. Semenov, E.V. Suvorov, and A.M. Feigin for a discussion of certain aspects of this paper. This work was supported by the Russian Foundation for Basic Research (project no. 98-02-17204) and the program of the Russian Ministry of Science "Controlled Thermonuclear Fusion and Plasma Processes" (project no. 370).

REFERENCES

1. V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion, and Excitons* (Nauka, Moscow, 1965; Springer-Verlag, New York, 1984).
2. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas* (Nauka, Moscow, 1967; Pergamon, Oxford, 1970).
3. V. L. Ginzburg, *Theoretical Physics and Astrophysics* (Nauka, Moscow, 1967; Pergamon, Oxford, 1979).
4. *Plasma Electrodynamics*, Ed. by A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin, *et al.* (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).
5. A. D. Piliya and V. A. Fedorov, *Reviews of Plasma Physics*, Ed. by B. B. Kadomtsev (Énergoatomizdat, Moscow, 1984; Consultants Bureau, New York, 1987), Vol. 13.
6. V. V. Alikaev, A. G. Litvak, E. V. Suvorov, and A. A. Fraiman, in *High-Frequency Plasma Heating*, Ed. by A. G. Litvak (American Institute of Physics, New York, 1992), p. 1.
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).

[11] See [3] in this connection.

[12] As we have seen above, the question of the energy flux is associated with the problem of constructing a geometric-optics description of wave propagation in certain media.

8. H. Bindslev, in *Contributing Papers of 9th Joint Workshop on Electron Cyclotron Emission (ECE) and Electron Cyclotron Heating (ECH)*, Ed. by J. Lohr (World Scientific, Singapore, 1995), p. 585.
9. E. Westerhof, in *Proceedings of the Joint Varenna-Lausanne International Workshop "Theory of Fusion Plasmas", Italy, 1994*, p. 115.
10. M. D. Tokman, Fiz. Plazmy **23**, 1104 (1997) [Plasma Phys. Rep. **23**, 1023 (1997)].
11. E. Westerhof, Plasma Phys. Controlled Fusion **39**, 1015 (1997).
12. A. I. Smirnov and M. D. Tokman, Zh. Éksp. Teor. Fiz. **110**, 549 (1996) [JETP **83**, 294 (1996)].
13. A. V. Timofeev, in *Reviews of Plasma Physics*, Ed. by B. B. Kadomtsev (Énergoizdat, Moscow, 1985; Consultants Bureau, New York, 1987), Vol. 14.
14. E. Westerhof, M. D. Tokman, and M. A. Gavrilova, in *Contr. Papers of 26th EPS Conference on Controlled Fusion and Plasma Physics, Maastricht, 1999*, p. 729; M. D. Tokman, E. Westerhof, and M. A. Gavrilova, Plasma Phys. Controlled Fusion **42**, 91 (2000).
15. M. Brambilla, *Kinetic Theory of Plasma Waves* (Oxford Univ. Press, Oxford, 1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1147–1162.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1319–1324.

PLASMA, GASES

Instability of Plasma–Dust Systems with a Macroparticle Charge Gradient

O. S. Vaulina, A. P. Nefedov, O. F. Petrov*, and V. E. Fortov

Institute of Heat Physics of Extreme States, Russian Academy of Sciences, Moscow, 127412 Russia

** e-mail: ipdustpl@redline.ru*

Received May 25, 2000

Abstract—An analysis is made of various self-oscillations which appear in plasma-dust systems with spatially varying macroparticle charges. The conditions of establishment and the nature of the evolution of the two main types of instability of these systems are analyzed. Particular attention is paid to cases of vortex particle motion. Dust systems having parameters similar to those in laboratory experiments are modeled numerically. Calculations were made for charged particles in an external electric field and in the Earth's gravitational field using a molecular dynamics method. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

At present, in the field of nonideal dusty plasmas there is a shift in interest away from problems of crystallization and phase conversions and toward problems associated with the establishment and evolution of various instabilities of plasma–dust systems [1–10]. Publications devoted to the analysis of forced and natural macroparticle oscillations are appearing all the time [5–10]. Some studies have been devoted to the experimental observation of the vortex motion of dust particles in various types of plasma: in dc gas-discharge plasmas and nuclear-excited plasmas with various methods of induction [7, 9]. Mention should also be made of a recent attempt to crystallize dust particles under microgravity conditions which resulted in the formation of macroparticle vortices in an experimental rf discharge generator [10].

Stable vortex motion of macroparticles in a dissipative medium such as a dusty plasma can only occur in the presence of potential sources compensating for the energy scattering. Then dynamic equilibrium can be established between the incoming and dissipated energy. These systems belong to a class of nonlinear, thermodynamically nonequilibrium, open systems. In equilibrium systems dissipative effects cancel out any inhomogeneity, i.e., thermodynamic equilibrium is established. In nonlinear open systems, dissipation exhibits a completely different quality. Its combined action with other processes may lead to the establishment of stable steady-state structures and complex oscillatory or chaotic regimes [11–14]. Frequently, qualitatively different processes occur in the same medium, whose direction may be changed by a very small resonant action matched with the internal properties of the system. This behavior is common to nonlinear media.

Most interesting from the point of view of the influence of resonant effects on the self-organization of nonlinear systems are active media (self-oscillatory systems) in which the type of steady-state motion is entirely determined by the properties of the medium and does not depend on the initial conditions. In complex hydrodynamic systems such as a dusty plasma a "soft" oscillation excitation regime is most frequently observed where oscillations appear, as it were, spontaneously without any initial impetus. If for some random reason (for example, as a result of thermal fluctuations) an oscillation having a negligibly small amplitude appears in the active medium, it will develop if its phase favors the pumping of energy into the system [13, 14].

Periodic, quasi-periodic, and stochastic self-oscillating systems exist. One example of stochastic vortex motion is turbulence and examples of periodic motion include autowaves and dissipative structures [11–13]. A dissipative structure is a stable inhomogeneous distribution of concentrations and is a dynamic ordered structure, unlike steady-state ordered structures such as crystals or quasi-crystalline systems of charged macroparticles in a highly nonideal dusty plasma. The classification into self-oscillations and dissipative structures for a bounded dusty system is highly arbitrary. In this sense, the motion of isolated particles can be classified as self-oscillations while the combined result of this motion, which gives a stable inhomogeneous distribution of concentrations in the bulk of the dust cloud, can be classified as dissipative structures.

Vortex motion may appear in "passive" nonlinear media as a manifestation of the soliton behavior of these systems. These solutions are generally determined by the initial conditions of the problem and disappear when allowance is made for dissipative effects. Sources of vortex motion may include gyroscopic or potential fields with frozen-in vortex lines. At present,

however, no data on any appreciable magnetic fields or on phenomena resembling barotropic are available for laboratory plasma–dust systems. Since the vortex motion of macroparticles is observed in various types of plasmas, it is logical to assume that electrical forces provide the source of this motion. Forces associated with ion drag or focusing and also thermal convection or thermophoresis can frequently be neglected. One possible mechanism capable of converting the potential energy of an external electric field into the energy of dust particle motion is the spatial dependence of the macroparticle charge in a dusty system [5, 8]. The presence of a macroparticle charge gradient in a plasma–dust system is determined by their nonuniform conditions of charging such as the temperature and concentration gradient of the surrounding plasma component, spatial changes in the illumination or surface temperature of the dust component when the macroparticles comprise an emitting material, dispersion of the shape or size of the dust particles, and so on [15].

In the present paper we consider cases where various self-oscillations occur in a dusty system as a result of a spatial change in the charge of monodisperse macroparticles. We analyze conditions for the establishment and evolution of two main types of instability of these systems. Particular attention is focused on analyzing the vortex motion of particles. We perform a numerical simulation of macroparticle systems having parameters close to the conditions of existing experiments to observe dusty structures in gas discharges. Numerical calculations were made using a molecular dynamics method for particles situated in an electric field and in the Earth's gravitational field.

2. SELF-OSCILLATIONS OF MACROPARTICLES IN A DUSTY SYSTEM WITH SPATIALLY VARYING CHARGE

2.1. Dispersion Relations in Nonconservative Systems

The dispersion relations $L(\omega, \mathbf{k}) = 0$ are a linear analog of the differential wave equations of motion and determine the functional dependence of the oscillation frequency ω on the wave vector $\mathbf{k}$. A study of the roots $\omega(k)$ of the relation $L(\omega, \mathbf{k}) = 0$ allows us to determine the region of existence of nontrivial and unstable solutions of the wave equations.

The dispersion relations $L(\omega, \mathbf{k}) = 0$ corresponding to conservative systems have no imaginary parts but the wave equations describing the evolution of oscillations in the weakly nonlinear limit describe competition between nonlinearity and dispersion. The nonlinear terms generate harmonics (only provided that the initial wave has sufficient amplitude) which compete with dispersion effects, generating final equilibrium. This situation differs substantially from the case where oscillations exist in nonconservative systems since energy scattering occurs in these systems and energy may be transferred by some mechanism such as a background flux, temperature gradient, or charge gradient. In these systems the dispersion relations $L(\omega, \mathbf{k}) = 0$ or their roots $\omega(k)$ are complex functions $\omega = \omega_R + i\omega_I$ and the wave solutions for $\omega_I > 0$ increase exponentially and are unstable, whereas for $\omega_I < 0$ they have a dissipative nature and decay exponentially.

When energy is transferred or pumped into a system, this potential energy may be used to form oscillations and waves. Nonequilibrium nonlinear systems frequently have some "control parameter" μ such that this nonlinear system may become unstable ($\omega_I > 0$) when this parameter passes a certain critical value $\mu = \mu_c$. In the supercritical region $\mu > \mu_c$ the oscillations will draw energy from the stored potential energy. This means that in this region infinitely small perturbations generated in the system as a result of thermal or other perturbations will grow.

Existing mathematical models developed to study the conditions for the establishment of oscillations in nonequilibrium nonlinear systems are constructed by analyzing differential wave equations. These models exhibit two main types of instabilities: first, dissipative instability for systems in which damping occurs (type I instability) and second, dispersion instability when there is weak or no dissipation (type II instability) [14]. The main difference between these instabilities is that the amplitude equation $f(A, x, t)$ for type I instability has the first order of the time derivative whereas for type II instabilities it contains the second derivative. The simplest equations for type I instability are nonlinear diffusion equations describing various vortex convective motion, dissipative structures, and autowaves [11–14]. The amplitude of the oscillations in these systems is limited by the existence of specific boundary conditions or singularities of the spatial distribution of the medium parameters (for example, background sources of potential energy) which ensure a progressive increase in dissipative losses. Examples of equations for type II instabilities in weakly dissipative systems may include the Van der Pol equations which simulate the oscillations of various self-excited oscillators and the system of Lorentz equations which is one of the main models in stochastic oscillation theory [12–14]. The amplitude and frequency of the oscillations in these systems may be limited as a result of dissipative processes and as a result of dispersion effects (for example, phase mismatch) which play a dissipative role phenomenologically [14].

These instability categories can be classified if we consider the dispersion relation $L(\omega, \mathbf{k})$ for the small perturbation of a stable system G by a harmonic wave of amplitude $\mathbf{b}$:

$$\varphi = \mathbf{b}\exp\{ikx - i\omega t\}. \quad (1)$$

The differential wave equations written in operator form then have the form $\hat{G}(ik; -i\omega; \mu)\mathbf{b}$ and the dispersion relation

$$L(\omega, \mathbf{k}) \equiv \det\hat{G} = 0$$

will show whether this model contains damped terms. If a type I instability exhibits damping, the dispersion relation $L(\omega, \mathbf{k})$ will be a complex function for stable ($\mu < \mu_c$) and unstable ($\mu > \mu_c$) states of the system since it contains even and odd powers i. The roots of the dispersion relation will also be complex, $\omega = \omega_R + i\omega_I$ and Eq. (1) has the form

$$\varphi = \mathbf{b}\exp\{ikx - i\omega_R t\}\exp\{\omega_I t\}.$$

For $\omega_I > 0$ the solution will increase in time and is unstable. The point where $\mu = \mu_c$ and the value of ω_I changes sign is called the bifurcation point of the system.

The second type of instability occurs when there is negligible or no dissipation. In this case, the dispersion relation is a real function and its roots may form the complex-conjugate pair $\omega = \omega_R + i\omega_I$. Equation (1) will have the form

$$\varphi = \mathbf{b}\exp\{ikx - i\omega_R t\}\exp\{\pm\omega_I t\}.$$

Consequently, the solution increases exponentially for any $\omega_I \neq 0$. For stable solutions $\omega_I = 0$ and the harmonic perturbation will propagate dispersively and not decrease as for a dissipative system. In this case, the critical parameter μ_c determines the boundary between the region of neutral stability of the system $\omega_I = 0$ and its unstable state $\omega_I \neq 0$.

One of the important characteristics of active media is the tendency of the system to act predominantly via the mode corresponding to μ_c. In cases of strong dispersion, as a result of the evolution of type II instability, in practice only one mode "survives" and the steady-state motion is a harmonic wave of frequency corresponding to a specific resonant frequency ω_c of the system. Under conditions of weak dispersion the form of the steady-state regular motion may comprise pulses of various shapes far from sinusoidal. It should be noted that in many physical problems associated with modeling the behavior of hydrodynamic systems, the inclusions of even weak viscosity destroys the nature, or completely eliminates the appearance, of dispersion solutions.

2.2. Linearization of the Discrete Problem of Motion

We shall analyze a dust system consisting of N_p charged particles in the electric field of a two-dimensional cylindrical trap in the presence of the spatially varying macroparticle charge $Z = Z(\rho, y)$ where $\rho = \sqrt{x^2 + y^2}$. We write the equation of motion for each particle taking into account the pair interaction $\mathbf{F}_{\text{int}}$, the total external force $\mathbf{F}_{\text{ext}}$ of the gravitational $m_p g$ and electric $\mathbf{E}(\rho, y) = \mathbf{i}E(y) + \mathbf{j}E(\rho)$ fields, the force of friction and the random Brownian $\mathbf{F}_{\text{br}}$ produced by the impacts of the surrounding gas molecules:

$$m_p\frac{d^2\mathbf{r}_k}{dt^2} = \sum_j F_{\text{int}}(r)|_{r=|\mathbf{r}_k-\mathbf{r}_j|}\frac{\mathbf{r}_k-\mathbf{r}_j}{|\mathbf{r}_k-\mathbf{r}_j|} - m_p\nu_{\text{fr}}\frac{d\mathbf{r}_k}{dt} + \mathbf{F}_{\text{br}} + \mathbf{F}_{\text{ext}}. \quad (2)$$

Here r is the interparticle spacing, m_p is the particle mass, and ν_{fr} is the friction frequency. The external force is defined as

$$\mathbf{F}_{\text{ext}} = \mathbf{i}\{E(y)eZ(\rho, y) + m_p g\} + \mathbf{j}E(\rho)eZ(\rho, y)$$

and the pair interaction between the particles is described by the force

$$F_{\text{int}}(r) = -eZ(\rho, y)\frac{\partial\phi_D}{\partial r},$$

where

$$\phi_D = \frac{eZ(\rho, y)}{r}\exp\left(-\frac{r}{D_p}\right)$$

is the screened Coulomb potential with the screening length D_p and e is the electron charge. Hence, the interparticle interaction force and the external forces acting on the particle depend on its coordinates. When the curl of these forces is nonzero, this system may perform positive work, which compensates for the dissipative energy losses. The system (2) is not conservative since it contains energy scattering as a result of friction, and energy may be transferred by means of the combined work of the electric field forces and the Earth's gravitational field.

It is almost impossible to make an analytic analysis of the conditions for the establishment of oscillations for a system described by the Eqs. (2). For a one-dimensional chain of interacting particles the problem can be reduced to a system of nonlinear partial differential equations [12, 14]. However, this transformation is extremely laborious and various additional assumptions on the nature of the particle interaction must be introduced to solve the spatial problem. Thus, we shall confine ourselves to analyzing a two-dimensional equation of motion for an isolated particle situated in the Earth's gravitational field $m_p g$, in the external electric field

$$\mathbf{E}^{\text{ext}}(\rho, y) = \mathbf{i}E^e(y) + \mathbf{j}E^e(\rho)$$

and the total electric field of the charged particles in the dust cloud

$$\mathbf{E}^{\text{int}}(\rho, y) = \mathbf{i}E_y^i(\rho, y) + \mathbf{j}E_\rho^i(\rho, y),$$

assuming that the Brownian force and the collective effects associated with the spatial fluctuations of the

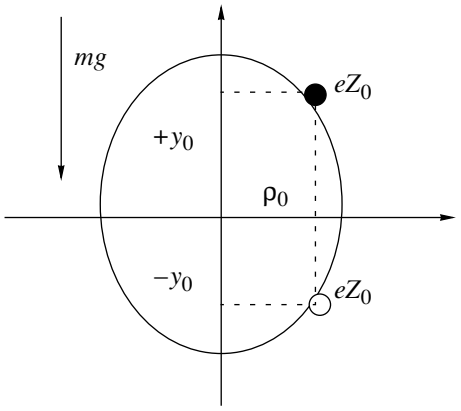


Fig. 1. Schematic showing the position of a macroparticle of charge eZ_0 at the edge of a dust cloud.

particle charges in the dust cloud play a minor role at the time of occurrence of instability.

We shall assume that a particle having a certain charge Z_0 is situated in a stable state at an extreme point in the dust cloud in the position (ρ_0, y_0) relative to its center (Fig. 1) and we shall consider its response to certain deviations (ρ, y) from the equilibrium point (ρ_0, y_0) of the system:

$$m_p\frac{d^2\rho}{dt^2} = -m_p\nu_{fr}\frac{d\rho}{dt} - e\{E^e(\rho_0) + \alpha\rho - E^i_\rho(\rho_0, y_0) - \omega_0 y - \omega^p\rho\}(Z_0 + \beta_\rho\rho + \beta_y y), \quad (3a)$$

$$m_p\frac{d^2y}{dt^2} = -m_p\nu_{fr}\frac{dy}{dt} + e\{E^e(y_0) - \gamma y + E^i_y(\rho_0, y_0) + \omega_0\rho + \omega^y y\}(Z_0 + \beta_\rho\rho + \beta_y y) - m_p g, \quad (3b)$$

where

$$\alpha = \frac{dE^e(\rho)}{d\rho}, \quad \gamma = -\frac{dE^e(y)}{dy}, \quad \omega^\rho = \frac{dE^i_\rho(\rho, y)}{d\rho},$$

$$\omega^y = \frac{dE^i_y(\rho, y)}{dy}, \quad \omega_0 = \frac{dE^i_\rho(\rho, y)}{dy} \equiv \frac{dE^i_y(\rho, y)}{d\rho},$$

$$\beta_\rho = \frac{dZ(\rho, y)}{d\rho}, \quad \beta_y = \frac{dZ(\rho, y)}{dy}$$

are the first derivatives of the system parameters at the point (ρ_0, y_0). The equality

$$\frac{dE^i_\rho(\rho, y)}{dy} = \frac{dE^i_y(\rho, y)}{d\rho}$$

should be satisfied because of the potential nature of the total field both for the case of fixed charged particles and for $Z(\rho, y) = \text{const}$.

The linearized system (3a), (3b) will have the form

$$\frac{d^2\rho}{dt^2} = -\nu_{fr}\frac{d\rho}{dt} + a_{11}\rho + a_{12}y, \quad (4a)$$

$$\frac{d^2y}{dt^2} = -\nu_{fr}\frac{dy}{dt} + a_{22}y + a_{21}\rho, \quad (4b)$$

where

$$a_{11} = -\frac{eZ_0\{\alpha - \omega^\rho\}}{m_p}, \quad a_{12} = \pm\frac{eZ_0\omega_0}{m_p},$$

$$a_{22} = -\frac{eZ_0\{\gamma - \omega^y\} + m_p g\beta_y/Z_0}{m_p},$$

$$a_{21} = \pm\frac{eZ_0\omega_0 + m_p g\beta_\rho/Z_0}{m_p}$$

for the stationary stable state of a particle

$$\{\rho_0 = \rho(t \longrightarrow \infty); \quad \pm y_0 = y(t \longrightarrow \infty);$$

$$E^e(\rho_0) = E^i_\rho(\rho_0, \pm y_0);$$

$$E^e(y_0) \pm E^i_y(\rho_0, \pm y_0) = m_p g/eZ_0\}$$

in the position above $(\rho_0, +y_0)$ or below $(\rho_0, -y_0)$ the center of the dust cloud (Fig. 1). Assuming that the classification of the wave regimes in active media is made by analogy with the oscillating regimes in systems with discrete parameters, we can use the system (4a), (4b) to analyze the conditions for the evolution of instability as a system of linearized amplitude equations neglecting the spatial derivatives [11, 14].

2.3. Conditions for the Establishment of Self-Oscillations in a System of Dust Particles

We shall consider the response of the system to a small perturbation $\varphi = b\exp\{-i\omega t\}$ which appears in the ρ or y direction. In this case, we obtain various "dispersion" relations $L(\omega) \equiv \det\hat{G} = 0$ which determine the region of existence of nontrivial and unstable solutions in the system $\hat{G}(-i\omega; \mu)b$ defined by Eqs. (4a) and (4b):

$$\omega^4 + (a_{11} + a_{22} - \nu_{fr}^2)\omega^2 + (a_{11}a_{22} - a_{12}a_{21}) + i\nu_{fr}\omega\{2\omega^2 + a_{11} + a_{22}\} = 0. \quad (5)$$

An analysis of Eq. (5) shows that in the absence of coupling between the system parameters ($a_{12}a_{21} = 0$) allowing for the fact that a_{22} and a_{11} are real and negative, this system will be stable. For $\nu_{fr} \neq 0$ the solution of the system (4a) and (4b) is also stable and will be damped

asymptotically ($\omega_I < 0$) for any small perturbation if the following conditions are satisfied simultaneously:

$$a_{11}a_{22} - a_{12}a_{21} \geq 0, \tag{6a}$$

$$\nu_{fr}^2 \geq |a_{11} + a_{22}|/2. \tag{6b}$$

The sign of the equality in Eq. (6a) determines the neutral curve of purely dissipative instability ($\omega_R = 0$, $\omega \equiv \omega_I = 0$). The appearance of type I dissipative instability will be determined by the condition

$$a_{11}a_{22} - a_{12}a_{21} < 0. \tag{7}$$

The motion may have a dispersive pattern ($\omega_R \neq 0$, $\omega_I = 0$) near a certain resonant frequency

$$\omega_c^2 = \frac{|a_{11} + a_{22}|}{2},$$

when the friction in the system is balanced by the incoming potential energy so that

$$\frac{a_{11} + a_{22}}{2} = \frac{4a_{12}a_{21} + (a_{11} - a_{22})^2}{4\nu_{fr}^2}. \tag{8}$$

Condition (8) determines the nontrivial solution of the system (4a), (4b). In the more general case, oscillation at a certain frequency ω_c will develop when the damping does not destroy the structures of the dispersion solution, i.e., is fairly weak [of the order of $O(\varepsilon)$ and does not allow any appreciable shifts of the neutral curve where $\omega_I = 0$. In order for the oscillating solutions to be amplified, allowing for the appearance of self-conjugate complex roots and the condition (6a), the following inequality must be satisfied

$$\nu_{fr}^2 \ll \omega_c^2 < \frac{(a_{12} - a_{21})^2}{4\nu_{fr}^2}. \tag{9}$$

Inequality (9) determines the region of dispersion instability of the problem where wave-like solutions may appear provided that the motion of isolated particles is synchronized in the dust cloud. It should be noted that unlike condition (7), inequality (9) is not a criterion for the onset of instability since the upper limit $(a_{12} - a_{21})^2/4\nu_{fr}^2$ of the frequency ω_c is a maximum estimate, which only determines the position of the neutral curve in various particular cases close to the nontrivial solutions of the problem.

A more detailed analysis of the system (4a), (4b) using Eq. (5) is not of particular interest since the effect of eliminating the spatial derivatives may actuate instability mechanisms which do not occur in real systems. For example, for the dispersion wave equations a solution which depends only on time in accordance with the Benjamin–Feir criterion, may be unstable to secondary modes [14].

From the physical point of view small perturbations in the system (4a), (4b) will grow in the two cases considered, first when no restoring force (7) exists for them and second, near a certain resonant frequency of the system when the forces of friction do not suppress the oscillating motion (9). If we use existing analogies with known amplitude equations, in the first case we obtain a diffusion type of motion equation. In the second case, the equation of particle motion will be quadratic with respect to time and the instability observed in the system will be of the dispersion type.

We shall give a brief analysis of the conditions for the evolution of the two different types of instabilities in a dust system using specific values of the parameters a_{ij}.

2.4. Dissipative Instability

Since purely dissipative instability only develops in a system in the absence of a restoring force, the conditions for the onset of bifurcation could be determined using the criterion for stability of the trivial solution in the Lyapunov sense [16]. According to this criterion, the solution of the system (4a), (4b) neglecting the second time derivative will become unstable when the roots of the equation $\det(a_{ij} - r\delta_{ij}) = 0$ have a positive real part. Since the coefficients a_{11} and a_{22} for a particle situated in a steady state at the edge of the cloud are always negative, the condition for the onset of instability will be determined by relation (7), which agrees with the analysis of Eq. (5) presented above. Taking into account the value of the coefficients a_{ij}, we obtain

$$\begin{aligned}(\alpha - \omega^p)(eZ_0\{\gamma - \omega^y\} - \beta_y m_p g/Z_0)\\ - eZ_0\omega_0^2 < \pm\omega_0\beta_\rho m_p g/Z_0.\end{aligned} \tag{10}$$

Bearing in mind that

$$(\alpha - \omega^p)\{\gamma - \omega^y\} > \omega_0^2,$$

we can show that the left-hand side of (10) will be positive for any negative or small β_y and the system will be stable if $\beta_\rho = 0$. Then a particle situated below the center of the cloud ($\omega_0 < 0$) will be unstable for $\beta_\rho > 0$ whereas for $\beta_\rho < 0$ instability (10) will only occur for particles situated in the upper part of the dust system. The case of macroparticle instability accompanying a large positive change β_y in charge in the direction of the Earth's gravitational field is equivalent to the unstable position of a charged particle in an electric field having a positive gradient ($\gamma = -dE^e(y)/dy < 0$) and will not be considered here.

The control parameter μ_c of the system can be either the parameter β_ρ defining the background sources of potential energy or the value of ω_0 which depends on the shape and size of the dust cloud. For a bounded dust cloud whose shape can be approximated by an ellipsoid of revolution having the radius of curvature $R_c(\rho_0, \pm y_0)$ and total charge Q_Σ uniformly distributed over its volume, we have $\omega_0 \sim Q_\Sigma/R_c$. Thus, we can assume that if the charge has a weak spatial dependence, the stability

of the dust system will be destroyed as Q_Σ increases and when clouds having shapes closer to spherical are formed. The total charge Q_Σ will increase as the number N_p of charged particles increases on transition from small cluster systems to extended objects or as the macroparticle concentration n_p in the dust cloud increases.

If the second time derivative can be neglected and the curl of the particle velocity vector $\mathbf{V}$

$$\mathbf{\Omega} = \operatorname{curl}\mathbf{V} \equiv \operatorname{curl}\left(\frac{d\rho}{dt}, \frac{dy}{dt}\right)$$

is nonzero, condition (10) describes the onset of vortex motion of the particle along a certain closed curve which is impossible in a conservative system. The vector $\mathbf{\Omega}$ determines the angular velocity of the macroparticles. For monotonic spatial dependences of the electric field and charge functions the direction of rotation of the particles in the plane parallel to the field of gravity can be determined from the sign of curl $\Omega = g\beta_\rho/\{Z_0\nu_{fr}\}$ of the linearized system of equations (4a), (4b). Hence, the value of the parameter β_y does not influence the conditions of establishment or the direction of the vortex motion.

It should be stressed that these conditions applied to the evolution of instability from the steady-state position of a particle having the velocity $V(\rho_0, y_0) = 0$. When a cloud of weakly interacting particles forms in a highly viscous gas, drift motion having a certain constant velocity $V(\rho, y) \approx$ const compensated by forces of friction near the point $(\rho_0, \pm y_0)$ may become stable. A similar pattern of motion may be achieved in nuclear induction experiments in plasma–dust systems conducted under atmospheric-pressure conditions using weakly charged macroparticles [9]. In this case, the coefficients a_{ij} of the system (4a), (4b) and the condition for the establishment of first-category instability and the value of curl $\mathbf{\Omega}$ also vary.

2.5. Dispersion Instability

A typical difference between dispersion instability and a dissipative bifurcation is that its formation is accompanied by the evolution of perturbations at frequencies ω close to some natural resonant frequency ω_c of the system.

The reliability of any analysis of the conditions for the onset of this instability is determined by the possible consequences arising from the absence of spatial derivatives. Thus, we shall consider a simple example for which the solution of the system (4a), (4b) has the clearest physical meaning.

We shall assume that $a_{11} \approx a_{22}$, $|a_{11}| \gg \nu_{fr}^2$, condition (10) is not satisfied and a restoring force exists, Then, if a particle undergoes a certain displacement from the equilibrium position, a damped harmonic perturbation appears having a certain natural frequency $\omega = \omega_c$ which at the initial stages of evolution may be expressed in the form

$$\frac{dy}{dt} = -\omega_1\rho, \quad \frac{d\rho}{dt} = \omega_2 y \quad (\omega_c^2 = \omega_1\omega_2 \equiv |a_{11}|).$$

In this case, the system (4a), (4b) may be written in the standard form to analyze self-oscillations:

$$\frac{d^2\rho}{dt^2} = \left(-\nu_{\text{fr}} + \frac{a_{12}}{\omega_2}\right)\frac{d\rho}{dt} + a_{11}\rho, \tag{11a}$$

$$\frac{d^2y}{dt^2} = \left(-\nu_{\text{fr}} - \frac{a_{21}}{\omega_1}\right)\frac{dy}{dt} + a_{11}y. \tag{11b}$$

Taking into account the tendency of the system to act via the predominant mode close to the position of the neutral curve [14], we find that ω_c will obey Eq. (8) and $\omega_1 = \omega_2$. The criterion for the evolution of type II instability will then be

$$2\nu_{\text{fr}}\omega_c < |a_{12} - a_{21}|,$$

since the difference between the positive or negative values of ω_c is physically meaningless ($\omega_c^2 > 0$). Allowing for the coefficients a_{12} and a_{21} we obtain

$$\nu_{\text{fr}} < \omega_c < \frac{|\Omega|}{2} = \frac{g|\beta_\rho|}{2Z_0\nu_{\text{fr}}}. \tag{12}$$

Condition (12) is equivalent to criterion (9) from which it follows that when the second type of instability occurs, the frequency ω_c should be close to or below $|\Omega|/2$. Consequently, a decrease in the friction frequency ν_{fr} below a certain threshold value leads to an increase in the parameter $\omega_\Omega = |\Omega|/2$ relative to the resonant frequency ω_c of the system which first promotes the evolution of dispersion instability [when Eq. (10) is satisfied] and second may result in the appearance of new instability against the background of the existing vortex motion.

2.6. Particle Kinetic Energy

One of the important problems in studies of any motion in nonconservative systems is the energy exchanged by the background and the steady-state oscillations. Since all stable motion of charged dust particles in an electric trap is finite, their kinetic energy for a certain direction x will be determined by the amplitude $A \sim \sqrt{\langle x^2\rangle}$ where $\langle x^2\rangle$ is the time-averaged particle displacement, and the characteristic frequency ω:

$$\langle K_x\rangle \approx m_p A^2\omega^2/2.$$

In general, the inflow of potential energy (A or ω) into the dust system should be controlled by dissipative processes.

A linear analysis cannot be used to estimate the amplitude A of the steady-state oscillations and is

merely suitable at the initial stages of the process when the perturbations are small since in the linear approximation for $\omega_I > 0$ the oscillations will increase without bound. The amplitude during the evolution of the first type of instability, i.e., regular vortex oscillations, can be estimated taking into account the scale of the steady-state motion:

$$2l_p < A < L/2,$$

where $l_p = n_p^{-1/3}$ is the average interparticle distance, and L is the characteristic size of the dust cloud. When the second type of instability is established, i.e., wavelike motion of particles in a dust cloud having a structure close to crystalline, the displacement of an isolated particle from the equilibrium position should not exceed the radius of a Wigner–Seitz cell: $A < l_p/2$.

The influence of dissipative effects on the frequency ω of the established motion differs substantially for different cases of instability evolution. The stable regular motion of the particles after the evolution of a dissipative bifurcation should obey the diffusion equation so that the angular velocity of rotation will be compensated by forces of friction so that $\omega^2 \sim 1/\nu_{fr}^2$. When dispersion instability develops, oscillation should be established at a certain resonant frequency $\omega = \omega_c$ determined by the system parameters and condition (12) limits the possible increase in the oscillation frequency ω as a result of progressive dissipative losses.

We shall estimate the kinetic energy $K_{(I)}$ which may be acquired by a dust particle following the evolution of type I dissipative instability for the case of linearly varying charge

$$Z(\rho) \sim Z_0 + \beta_\rho \rho$$

within the particle trajectory. The value of ω^2 can then be taken as

$$\frac{\Omega^2}{4} = \frac{(g\beta_\rho)^2}{(2Z_0\nu_{fr})^2},$$

and the kinetic energy $K_{(I)}$ can be expressed in the form

$$K_{(I)} = \frac{m_p g^2 \xi^2}{8\nu_{fr}^2}, \qquad (13)$$

where the parameter $\xi = A\beta_\rho/Z_0$ determines the relative changes in charge $Z(\rho)$ within the trajectory of a macroparticle. We shall consider the values of the parameter ξ for which $K_{(I)}$ becomes higher than the thermal energy corresponding to room temperature $T_p \approx 0.02$ eV for a macroparticle of radius $a_p = 5$ μm having the material density $\rho = 2$ g/cm^3 ($m_p \approx 10^{-12}$ kg). We take the friction frequency to be $\nu_{fr} = 20$ s^{-1} which approximately corresponds to the pressure $P = 0.5$ Torr in neon at room temperature using the free-molecule approximation [17]. As a result we find that for $\xi \approx 10^{-3}$ the kinetic energy $K_{(I)}$ is almost an order of magnitude higher than T_p. As the buffer gas pressure increases to $P = 5$ Torr or the particle radius decreases to $a_p = 2$ μm, we obtain $K_{(I)}/T_p > 10$ for $\xi > 10^{-2}$. Consequently, even negligible changes ξ in the charge can lead to an effective consumption of potential energy from background sources.

The kinetic energy $K_{(II)}$ of a dust particle for the evolution of dispersion instability can be estimated assuming that for small changes in the macroparticle charge the value of ω_c characterizing the particle motion in a selected direction will be close to the dust frequency of the system

$$\omega_p \approx (2e^2 Z_0^2 n_p \exp(-k)\{1 + k + k^2/2\}/m_p)^{1/2},$$

where $k = l_p/D_p$. In accordance with the conditions achieved in gas-discharge plasma experiments, we shall take X_0 to be equal to the equilibrium charge $\langle Z \rangle$ acquired by a dust particle as a result of charging by ion and electron fluxes from the surrounding plasma:

$$\langle Z \rangle \approx 2 \times 10^3 a_p\,[\mu\text{m}] T_e\,[\text{eV}],$$

where T_e is the electron temperature [18, 19]. The kinetic energy $K_{(II)}$ can then be expressed in the form

$$K_{(II)}\,[\text{eV}] \approx 5.76 \times 10^3 (a_p\,[\mu\text{m}] T_e\,[\text{eV}])^2 \chi^2 c_n / l_p\,[\mu\text{m}], \quad (14)$$

where

$$c_n = \exp(-k)\{1 + k + k^2/2\}, \quad \chi = A/l_p < 0.5$$

is a certain parameter determining the ratio of the amplitude A of the particle trajectory to the average interparticle distance l_p. Bearing in mind the conditions of observation of dust structures in laboratory gas-discharge plasmas, we assume $k \approx 1$–2, $l_p = 500$ μm, and $T_e = 1$ eV and we consider the energy which can be acquired by a particle of radius $a_p = 5$ μm if the amplitude of its steady-state motion corresponds to $\chi = 0.1$. In accordance with Eq. (14), this energy appreciably exceeds T_p and is $K_{(I)} \approx 3$ eV. The maximum kinetic energy of the particles for the evolution of dispersion instability which does not destroy the crystal structure of the cloud is achieved when $\chi = 0.5$ and is

$$K_{(II)}^{\lim} = c_n e^2 Z_0^2 / 4 l_p,$$

from which we obtain

$$\frac{K_{(II)}^{\lim}}{T_p} = \frac{\Gamma}{4} \quad \text{for} \quad k \longrightarrow 0,$$

where Γ is the interparticle interaction parameter.

These estimates of the particle kinetic energy (13) and (14) may be exaggerated because the frequency ω of the steady-state motion is not determined exactly. One source of possible constraint on the value of ω are the various collective effects caused by the presence of strong interparticle interaction forces and phenomeno-

logically playing a dissipative role. Spatial fluctuations of the surrounding particles may change the frequency of the steady-state oscillations of an isolated particle or lead to randomization of its motion. It should be stressed that the establishment of regular motion as a result of the evolution of these instabilities is a particular case of the solution of the nonlinear differential equations. If the action of these random forces F_{rad} has no correlation with the Brownian force F_{br}, the kinetic temperature T_s of the system will increase and its increment is given by

$$\Delta T \sim \langle F_{rad}^2 \rangle / \nu_{fr}(\nu_{fr} + \eta),$$

where η is the characteristic frequency of the action F_{rad} [5, 20].

To conclude, we note that the analysis of the conditions of establishment and the pattern of evolution of the instabilities presented in this section is very approximate since it is based on an appreciable simplification of the problem. However, it shows qualitatively that a cloud of charged dust particles may have two different types of bifurcations. The viability of our conclusions and their applicability to a real system consisting of many charged particles can be checked by means of a numerical simulation of the complete problem of motion (2).

3. NUMERICAL SIMULATION OF DUST SYSTEMS WITH SPATIALLY VARYING MACROPARTICLE CHARGE

3.1. Parameters of the Problem

We used the molecular dynamics method to solve the three-dimensional motion problem (2). We considered the motion of each particle in the cloud allowing interparticle pair interaction, the total external force of the gravitational and electric fields, the friction force, and the random Brownian force. The external electric fields are assumed to be linear:

$$\mathbf{E}^{ext}(\rho, y) = \mathbf{i}(E_G + \gamma y) + \mathbf{j}\alpha\rho.$$

The charge function had the form

$$Z(\rho, y) = Z_{00}(1 + \beta_\rho^* \rho^2 + \beta_y^* y^2)$$

with various coefficients β_ρ^* and β_y^* such that the changes in the particle charge within the dust cloud did not exceed 30%.

For all cases not discussed below the particle temperature T_p was assumed to be room temperature (300 K), the screening radius $D_p \approx 1000$ μm, and the radial field gradient $\alpha \approx 3N_p^{-1/3}$ V/cm^2 (N_p is the number of particles in the system). In calculations with different numbers of particles N_p the parameters of the charge function β_ρ^* and β_y^* varied proportionately as $N_p^{-2/3}$. This dependence of the external field and charge function on N_p could maintain an almost constant particle concentration and charge at the edge of the cloud. The variable parameters of the system (2) were the friction frequency ν_{fr}, the coefficients β_ρ^* and β_y^*, and the ratio of the electric field gradient γ in the direction of the gravitational field to the radial gradient α. It should be noted that reducing the system (2) to dimensionless parameters shows that if the characteristic values $Z_{00}^2/D_p^3 m_p$ and $\alpha Z_{00}/D_p m_p$ are conserved, the problem has similar solutions (pattern of motion, particle kinetic energy, $k = l_p/D_p$, and so on). The dynamic behavior of the macroparticles also depends on the characteristic frequencies ν_{fr}, ω_p, and the angular frequency $\omega_\Omega = |\Omega|/2$ Eq. (12) which allowing for the parameters of our problem may be expressed in the form

$$\omega_\Omega \approx g|\beta_\rho^*| L_\rho / (1 + \beta_\rho^* L_\rho^2)\nu_{fr},$$

where

$$L_\rho \approx 0.5 N_p^{1/3} l_p \sqrt{\frac{\gamma}{\alpha}}$$

is the radial dimension of the dust cloud.

Under the selected conditions we observed various cases of dynamic equilibrium in the system: vortex, oscillatory, and stochastic motion. The average interparticle distance l_p determined from the maximum of the pair correlation function was set in the range ~500–2300 μm depending on the other parameters of the problem. We considered fairly extended and also small cluster systems of macroparticles. Illustrations of numerical calculations for systems containing between 15 and 3000 particles are given below for the case of monotonically varying charge and electric field in the direction of the Earth's gravitational field.

3.2. Vortex Motion of Macroparticles

Unlike the turbulent vortex motion which is a particular case of stochastic self-oscillation, nonlinear systems may exhibit periodic circular motion and also other types of vortices such as spiral, concentric, or pulsating. In this section we consider several different examples of possible vortex motion of particles in systems with a charge gradient.

Vortex motion was established in a system of dust particles when conditions similar to Eq. (10) were satisfied. The stability of the system was destroyed when the asymmetry of the dust cloud decreased ($\gamma/\alpha \longrightarrow \sqrt{2}$) and also as β_ρ^* and the number of particle in small cluster systems increased to $N_p \approx 100$. With further increasing N_p the dynamic characteristics of the dust cloud remained almost constant. The conditions for excitation of oscillations did not depend on the friction forces and β_y^*. The direction of particle rotation was deter-

mined by the coefficient β_ρ^*: for $\beta_\rho^* < 0$ the particles "fell" in the center of the system whereas in the opposite case $\beta_\rho^* > 0$ they "floated."

The nature of the steady-state oscillations did not depend on the initial conditions or the existence of a certain spatial order in the particle configuration. Replacing the quadratic charge function $Z(\rho, y)$ with a linear one also did not alter the qualitative nature of the particle motion and its dependence on the viscosity of the buffer gas. The average kinetic energy of the regular macroparticle motion was

$$\langle K\rangle \sim 1/\nu_{fr}^2.$$

As ν_{fr} decreased below a certain threshold value $\nu_{lim} \approx (0.1–0.15)\omega_p$, the motion became randomized and a near-Maxwellian particle velocity distribution was established in the system having the temperature

$$T_s \approx \frac{2}{3}\langle K\rangle \sim \frac{1}{\nu_{fr}}.$$

A further decrease in ν_{fr} ($\omega_\Omega \longrightarrow \omega_p$) led to resonant pumping of the dust cloud. It can be postulated that this effect is the result of the evolution of the dispersion bifurcation Eq. (12) against the background of vortex particle motion (see Section 2.5).

3.2.1. Spiral rotation of particles. The calculations were made for the radial charge gradient defined by the function

$$Z(\rho) = Z_{00}(1 + \beta_\rho^*\rho^2),$$

where

$$Z_{00} = 4\times 10^4, \quad \beta_\rho^* \approx -12.5 N_p^{-2/3}\ \text{cm}^{-2}.$$

The ratio of the charge number Z_{00} to the particle mass m_p was taken as

$$Z_{00}/m_p \approx 10^{15}\ \text{g}^{-1}.$$

In this case, spiral rotation of isolated particles was observed in the system for $k = l_p/D_p \approx 2$, $\omega_p \approx 70\ \text{s}^{-1}$, and $\omega_\Omega\ [\text{s}^{-1}] \approx 1225 N_p^{-1/3}/\nu_{fr}\ [\text{s}^{-1}]$.

Figures 2a and 2b illustrate the quasi-harmonic motion of 15 particles and the parameters of the problem are indicated in the captions, The oscillations are quasi-periodic with two incommensurable frequencies. The particle motion of lower frequency and larger amplitude was synchronized over a complex spiral about a certain base radius along which the trajectory of the spiral motion of four "high-frequency" particles passed. The amplitude of the regular oscillations was almost independent of the friction frequency ν_{fr} and was determined by the charge gradient and the trap field. Time dependences of the amplitude of the oscillations along the *y*-axis for an isolated particle are plotted in Fig. 3 for various friction frequencies.

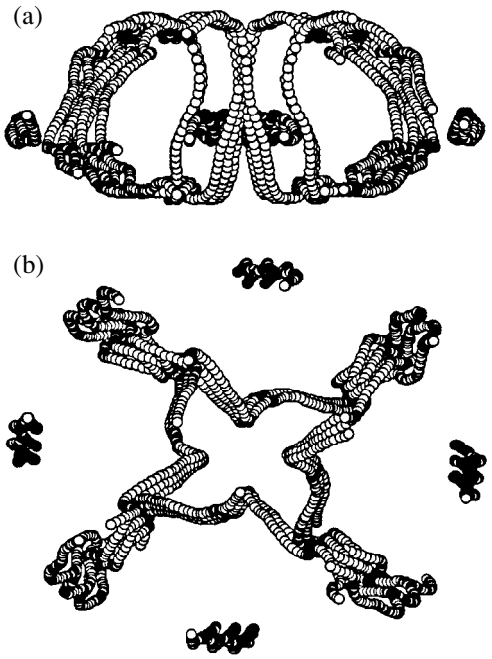


Fig. 2. Particle trajectories over time $t = 200/\nu_{fr}$ in a system having the parameters: $\nu_{fr} \approx 77\ \text{s}^{-1}$, $\gamma/\alpha = 2$, $Z_0 = 4\times 10^4$, $\beta_\rho^* \approx -2\ \text{cm}^{-2}$, $\beta_y^* = 0$, $N_p = 15$; $\langle K_x\rangle = \langle K_z\rangle \approx 0.34$ eV, and $\langle K_y\rangle \approx 1.1$ eV: (a) side view and (b) top view.

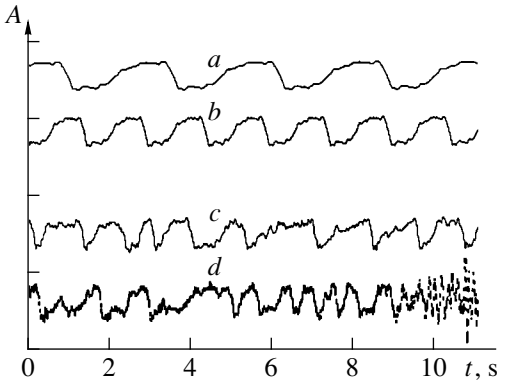


Fig. 3. Time dependences of the amplitude *A* of the oscillations (along the *y*-axis) of an isolated particle with decreasing friction frequency $\nu_{fr} \approx 77\ \text{s}^{-1}$: (*a*) ν_{fr}; (*b*) $\nu_{fr}/2$; (*c*) $\nu_{fr}/6$; and (*d*) $\nu_{fr}/10$ for $t < 9$ s and $\nu_{fr}/20$ for $t > 9$ s.

The velocity of rotation was inversely proportional to ν_{fr} and the average kinetic energy of the particles was $\langle K\rangle \sim 1/\nu_{fr}^2$ (see Fig. 4). As the friction frequency decreased below the threshold $\nu_{lim} \approx 0.15\omega_p$ the oscillations became irregular, the trajectories became more complex, the number of harmonics with similar amplitudes increased, and the oscillations became randomized. The particle velocity distribution became more uniform in terms of directions ($\langle K_y\rangle \approx \langle K_x\rangle = \langle K_z\rangle$) and

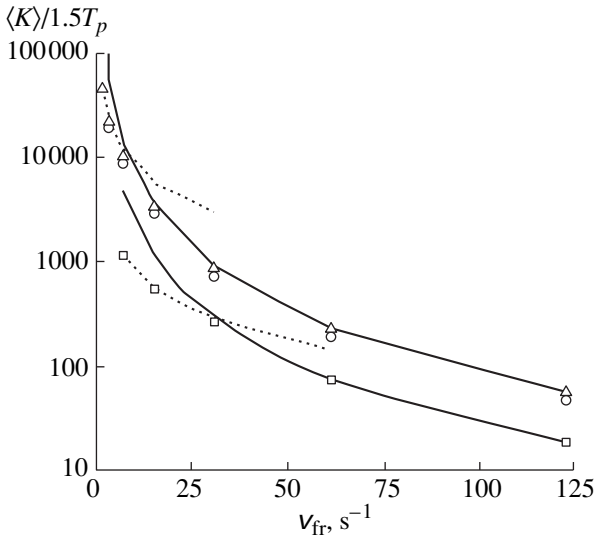


Fig. 4. The ratio $\langle K\rangle/1.5T_p$ vs. the friction frequency ν_{fr} for a dust system having the parameters: $\gamma/\alpha = 2$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -12.5 N_p^{-2/3}$ cm^{-2}, $\beta_y^* = 0$ for various numbers of particles N_p = (□) 15, (o) 60, and (△) 500. The solid curve gives the approximation by the function $f_{ap}(\nu_{fr}) \propto \nu_{fr}^{-2}$ and the dashed curve gives that by the function $f_{ap}(\nu_{fr}) \propto \nu_{fr}^{-1}$.

close to Maxwellian with the temperature $T_s = 2\langle K\rangle/3$, and we observed the establishment of the dependence $\langle K\rangle \sim 1/\nu_{fr}$ (Fig. 4). As the friction frequency decreased further to $\nu_{fr} \approx 7$ s^{-1} ($\omega_\Omega \approx \omega_p \approx 70$ s^{-1}) a resonant buildup of oscillations occurred, leading to "explosion of the system."

When the number of particles increased to 60 or more, the parameters of the vortex motion became similar to those of the large-scale low-frequency oscillation (oscillations of larger amplitude). The motion of isolated particles became more complex but remained spiral with respect to a certain base radius. The amplitude of the oscillations remained independent of ν_{fr} and the oscillation frequency was inversely proportional to the coefficient of friction until its decrease ($\nu_{fr} < \nu_{lim} \approx 0.1\omega_p$) resulted in the periodicity of the motion being destroyed and caused particle "heating" followed by explosion of the system for $\nu_{fr}[\text{s}^{-1}] < 1225 N_p^{-1/3}/\omega_p$ [s^{-1}] and $\omega_p \approx \omega_\Omega \approx 75$ s^{-1}.

It should be noted that in our numerical experiment the possibility of the buildup of resonant oscillations is attributed to the modeling of a spatially unbounded electrical trap. In this case, the system cannot "lose" dust particles. Under real experimental conditions, for example, under the conditions of a dc glow discharge [5–8], the instability which evolves as the number of particles in the dust system increases, can be stabilized

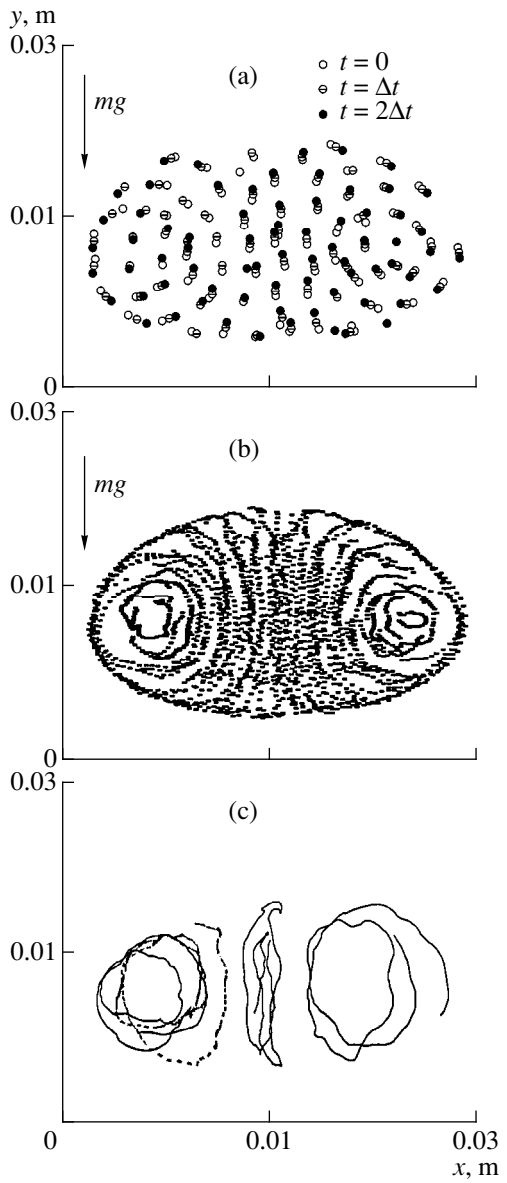


Fig. 5. (a) Successive ($\Delta t = 1/\nu_{fr}$) displacements and (b) distribution of particle concentrations in the central cross section of a trap, and also (c) trajectories of isolated particles over the time $t_{exp} = 150\Delta t$ in a system with $\nu_{fr} \approx 38.5$ s^{-1}, $\gamma/\alpha = 2$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -0.2$ cm^{-2}, $\beta_y^* = 0$, $N_p = 500$; $\langle K_x\rangle = \langle K_z\rangle \approx 4.3$ ev, $\langle K_y\rangle \approx 14$ eV.

by "excess" particles being ejected from the trap bounded by the electric field of a striation.

Calculations made for 500 particles are illustrated in Figs. 5a–5c. Figure 5a shows successive displacements of particles (over the interval $\Delta t = 1/\nu_{fr}$) incident in the central cross section of the cloud in the plane parallel to the force of gravity. Figure 5b shows all the positions of particles observed in the central part of the cloud during the exposure time $t_{exp} = 30\Delta t$. This figure demonstrates that an ordered distribution of particle concentration is

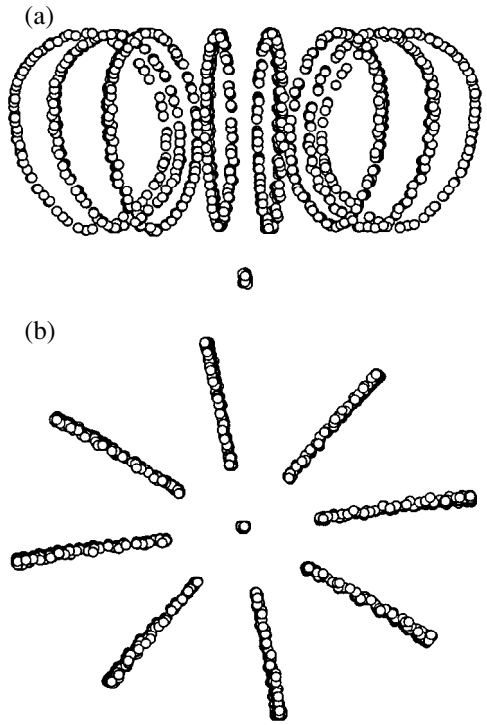


Fig. 6. Particle trajectories over time $t = 100/\nu_{fr}$ in a system having the parameters: $\nu_{fr} \approx 115\ s^{-1}$, $\gamma/\alpha = 1$, $Z_0 = 8 \times 10^3$, $\beta_y^* = \beta_\rho^* \approx 5.5\ cm^{-2}$, $N_p = 25$; $\langle K_x \rangle = \langle K_z \rangle \approx 0.85$ eV, and $\langle K_y \rangle \approx 2$ eV: (a) side view and (b) top view.

established in the dust cloud and identifies the result of the collective particle motion as a dissipative structure. Figure 5c shows trajectories of isolated particles in the bulk of the cloud over time $t_{exp} = 150/\nu_{fr}$.

3.2.2. Annular particle motion. Calculations were made for dust systems having the macroparticle charge function

$$Z(\rho, y) = Z_{00}(1 + \beta_\rho^* \rho^2 + \beta_y^* y^2),$$

where

$$Z_{00} = 8 \times 10^3, \quad Z_{00}/m_p \approx 2 \times 10^{14}\ g^{-1},$$

$$\beta_y^* \equiv \beta_\rho^* \approx 50 N_p^{-2/3}\ cm^{-2} \text{ at } \gamma/\alpha = 1.$$

The choice of this spatial dependence $Z(\rho, y)$ was determined by the assumption of a self-consistent plasma–dust system where the particle charge is proportional to the electric field potential.

The average interparticle spacing corresponded to $k = l_p/D_p \approx 1.6$, $\omega_p \approx 25\ s^{-1}$. The oscillation frequency of the isolated particles is inversely proportional to the coefficient of friction $1/\nu_{fr}$ ($\langle K \rangle \propto 1/\nu_{fr}^2$) until its decrease causes heating and explosion of the dust system similar to the case of spiral rotation. The characteristic frequencies ν_{fr} at which randomization and resonant pumping of the dust system occurred were determined in accordance with the varied parameters of the problem.

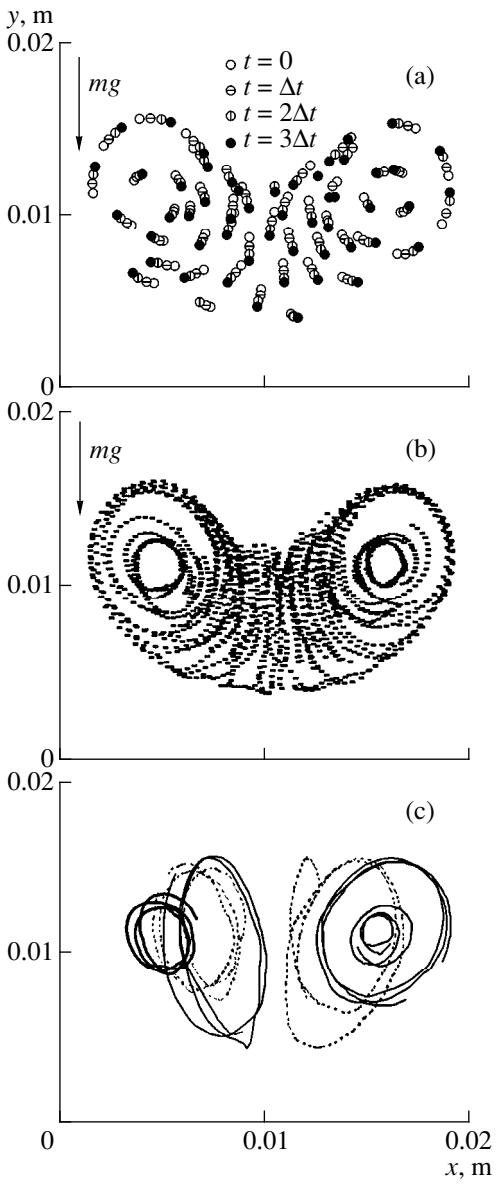


Fig. 7. (a) Successive ($\Delta t = 1/\nu_{fr}$) displacements and (b) distribution of particle concentrations in the central cross section of a trap, and also (c) trajectories of isolated particles over the time $t = 150\Delta t$ in a system with $\nu_{fr} \approx 115\ s^{-1}$, $\gamma/\alpha = 1$, $Z_0 = 8 \times 10^3$, $\beta_y^* = \beta_\rho^* \approx 0.75\ cm^{-2}$, $N_p = 500$; $\langle K_x \rangle = \langle K_z \rangle \approx 2.3$ eV, and $\langle K_y \rangle \approx 5.5$ eV.

Figures 6a and 6b illustrate the quasi-harmonic synchronized motion of 25 particles, showing the large-scale motion of three different particles along eight closed trajectories and the high-frequency motion of an isolated particle at the center of the system. Calculations made for 500 particles for similar system parameters are illustrated in Figs. 7a–7c. Figure 7a gives the successive displacements of particles over the interval

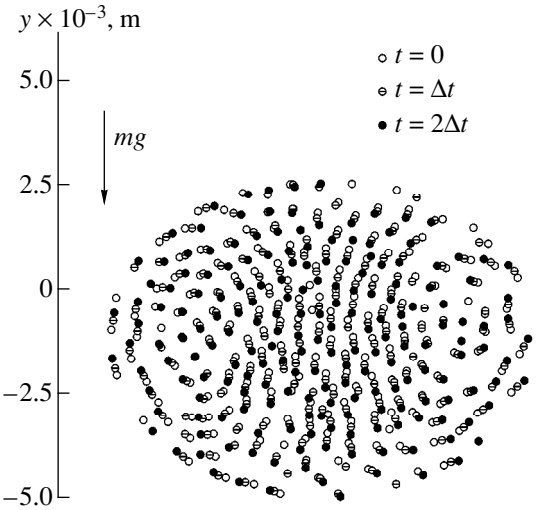


Fig. 8. Successive ($\Delta t = 1/\nu_{fr}$) displacements of particles in the central cross section of a trap in a system with $\nu_{fr} \approx 11.25\ s^{-1}$, $\gamma/\alpha = 1$, $Z_0 = 1.5 \times 10^4$, $\beta_y^* = \beta_\rho^* \approx 0.03\ cm^{-2}$, $N_p = 3000$; $\langle K_x \rangle = \langle K_z \rangle \approx 2.1$ eV, and $\langle K_y \rangle \approx 6.5$ eV.

$\Delta t = 1/\nu_{fr}$ in the central cross section of the trap. Figure 7b shows all the possible positions of particles incident in the central cross section of the cloud over the time $t_{exp} = 30/\nu_{fr}$. Trajectories of isolated particles in the bulk of the dust cloud are shown in Fig. 7c.

The main difference between the pattern of the macroparticle motion and the case examined in Section 3.2.1 is that isolated groups of particles exhibit stable vortex motion whose trajectories are closed in a bounded region of space. From this point of view, this dissipative structure is closer to an annular autowave than a spiral one or convective motion.

To conclude, we shall consider a system of macroparticles (2) having parameters similar to the conditions for the observation of dust structures in a dc glow discharge [5–8]. We shall take the particle radius to be $a_p = 5$ μm, the material density $\rho = 3.1\ g\ cm^{-3}$, and the coefficient of friction $\nu_{fr} \approx 11.25\ s^{-1}$ in accordance with the estimate of the approximation [17] for neon at pressure $P = 0.25$ Torr. The charge is assumed to be $Z_{00} = 1.5 \times 10^4$ and the screening radius $D_p = 580$ μm. Under conditions of dynamic equilibrium the average interparticle distance was $l_p \sim 580$ μm ($k \approx 1$). Figure 8 illustrates the calculations made for 3000 particles. The given positive charge gradient function $\beta_y^* \equiv \beta_\rho^* = 0.03\ cm^{-2}$ ensured that the kinetic energy $\langle K \rangle$ of the particle motion was two orders of magnitude higher than their thermal energy T_p corresponding to 300 K. In this case, the change in the particle charge in the dust cloud was less than 1% of its maximum. These changes in the macroparticle charge may appear, for example, as a result of a radial electron temperature gradient where the slower ones do not reach the walls of the gas-discharge tube at negative floating potential.

3.3. Examples of "Dispersive" Particle Motion

At first glance it may seem that it is fairly difficult to observe dispersion instability in a dust system with a charge gradient since the conditions for the evolution of this bifurcation should simultaneously satisfy Eq. (12) and not come within criterion (10). However, it should be noted that in our analysis we only studied one of the possible cases of spatial dependence of the external electric field when its changes in the direction of the force of gravity and in the radial direction are independent. This situation may not occur in real dust systems where more favorable conditions for the evolution of type II instability can be achieved. As an example, we can give results of a numerical simulation of the anomalous heating of a dust system with a vertical charge gradient which were reported in [5] for an external electric field having interdependent coordinates.

Although our article concentrates on analyzing vortex motion, we shall consider two additional examples of possible particle motion for the evolution of category II instability. These examples may be useful for a qualitative explanation of various effects observed in dusty plasma experiments such as the appearance of acoustic oscillations and anomalous macroparticle heating [5–8, 21–23].

The calculations were made for systems consisting of a small number of particles, between 15 and 60. The constraint of few particles is important in the sense that criterion (12) is easily established for these under conditions fairly close to the evolution of dissipative instability but not resulting in the appearance of vortex motion. Since, by increasing the charge gradient β_ρ we can come within the limits of relation (10), the only available control parameter is the friction frequency ν_{fr}. We found that when the friction frequency fell below a certain critical value $\nu_{fr} = \nu_c$, the modeled systems became unstable and the angular frequency was $\omega_\Omega \approx (1–5)\omega_p$. As a result of the evolution of this instability both regular and stochastic particle motion was established in the dust systems.

3.3.1. Regular particle oscillations. Calculations were made for 15–25 particles. We observed the establishment of various modes of longitudinal and transverse oscillations similar to acoustic or optical ones, and also the formation of complex pulsating oscillations. The steady-state oscillation mode depended on the spatial configuration of the particles similar to the oscillations of atoms and molecules in various types of crystal lattice. As the number of particles increased, any regular oscillations observed in the modeled systems ceased since the spatial symmetry in the macroparticle configuration was broken which resulted in a loss of synchronization of the individual motion. Thus, the existence of a specific spatial orientation in the particle

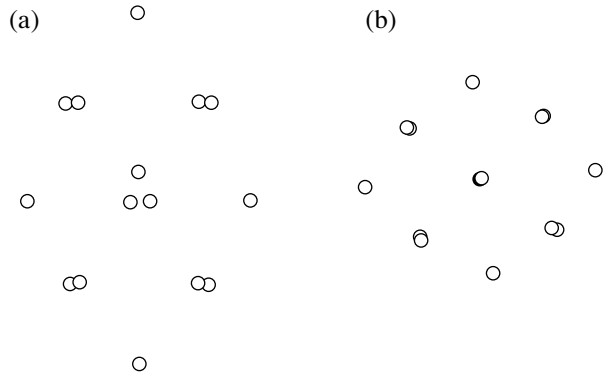


Fig. 9. Particle distribution in a system having the parameters: $\gamma/\alpha = 0.4$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -2$ cm^{-2}, $\beta_y^* = 0$, $N_p = 15$ for $\nu_{fr} > \nu_c \approx 1$ s^{-1}. (a) Side view and (b) top view.

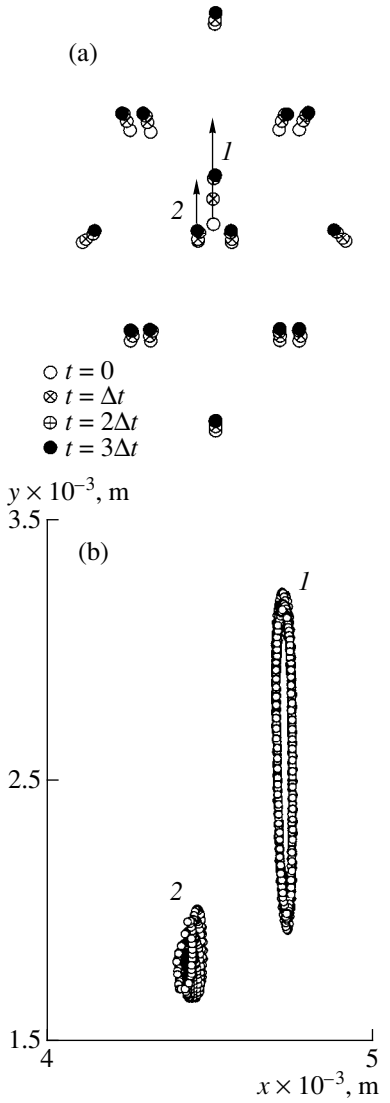


Fig. 10. (a) Successive displacements of particles over the interval $\Delta t \approx T_c/6$ (T_c is the period of the oscillations) and (b) trajectories ($t = 200T_c$) of isolated particles indicated by (*1*) and (*2*), in a system having the parameters: $\gamma/\alpha = 0.4$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -2$ cm^{-2}, $\beta_y^* = 0$, $N_p = 15$, $\nu_{fr} \approx 0.96$ s^{-1}, $\langle K_x \rangle = \langle K_z \rangle \approx 0.15$ eV, and $\langle K_y \rangle \approx 2.5$ eV.

configuration is one of the main factors in the evolution of wave-like motion in a dust cloud. For instance, the excitation of longitudinal oscillations requires a strictly vertical particle configuration (see Figs. 9a, 9b) which is confirmed by experimental data from studies where similar particle motion has been observed [2, 5–8].

We shall confine our illustration to regular oscillations similar to the acoustic oscillations observed in glow discharge experiments [5–8]. Figures 9a and 9b show a steady-state cloud consisting of 15 particles before the onset of instability ($\nu_{fr} > \nu_c \approx 1$ s^{-1}). The parameters of the problem are indicated in the caption. The only difference from the parameters of the problem described in Section 3.2.1 is the choice of ratio $\gamma/\alpha = 0.4$ which characterizes the shape of the cloud. The average interparticle distance corresponded to $k = l_p/D_p \approx 2.3$, $\omega_p \approx 52$ s^{-1}, and the angular frequency ω_Ω [s^{-1}] $\approx 218/\nu_{fr}$ [s^{-1}] for the critical coefficient $\nu_{fr} = \nu_c$ was approximately four times ω_p.

The pattern of particle motion after the evolution of dispersive instability is shown in Figs. 10a and 10b. The particle trajectories differed appreciably from the trajectories of large-scale vortex motion (see Section 2.2) and the frequency ω_c and amplitude of the steady-state oscillations was almost independent of the buffer gas frequency ν_{fr} when this decreased to a certain threshold value $\nu_{lim} \approx 0.38$ s^{-1}. Nevertheless, as ν_{fr} approached ν_{lim}, we observed an appreciable mismatch between the phases of the isolated particle motion. A decrease in the frequency ν_{fr} below the threshold ν_{lim} led to rapid heating of the system followed by pumping and explosion similar to the cases considered in Section 3.2.

Time dependences of the oscillation amplitude for isolated particles and the center of mass of the system along the *y*-axis are plotted in Figs. 11a and 11b. The frequency of the regular particle oscillations $\omega_c \approx 18$ s^{-1} was lower than the characteristic dust frequency ω_p of the system. As we noted in Section 2.6, this factor may be attributed to the existence of various collective or dispersion effects which phenomenologically play a dissipative role.

3.3.2. Heating of the dust cloud. In our analysis of some of the previous examples we have already encountered effects such as macroparticle heating and explosion of the dust system. These effects occurred as the number of equivalent modes in an initial harmonic (or quasi-harmonic) self-oscillating system increased and was followed by the evolution of oscillatory instability leading to parametric resonance. The processes simulating heating of a dust system may take place by a different scenario, for example, as the friction forces

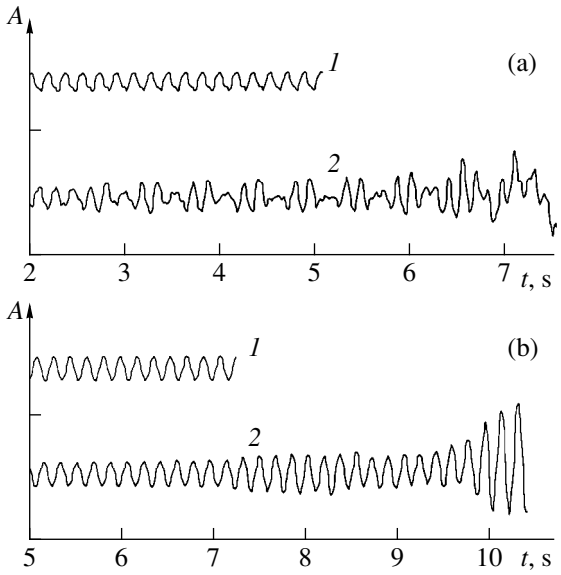


Fig. 11. Time dependences of the amplitude A (along the y-axis) of the oscillations for (a) an isolated particle indicated by (*2*) in Fig. 10a and (b) the center of mass of the system for the cases: (*1*) $0.38\nu_c < \nu_{fr} < \nu_c$; (*2*) $\nu_{fr} < 0.38\nu_c$.

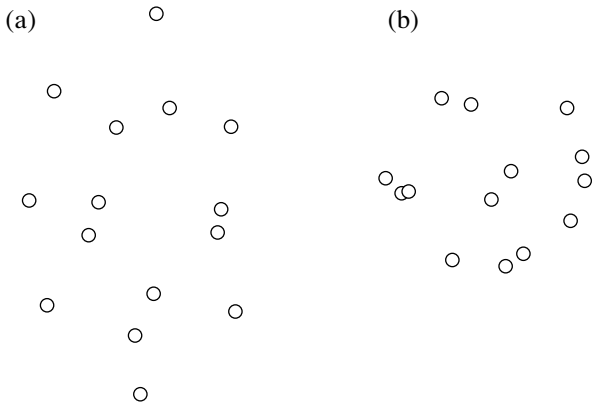


Fig. 12. Particle distribution in a system having the parameters: $\gamma/\alpha = 0.41$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx 2$ cm^{-2}, $\beta_y^* = 0$, and $N_p = 15$ for $\nu_{fr} < \nu_c \approx 4$ s^{-1}. (a) Side view and (b) top view.

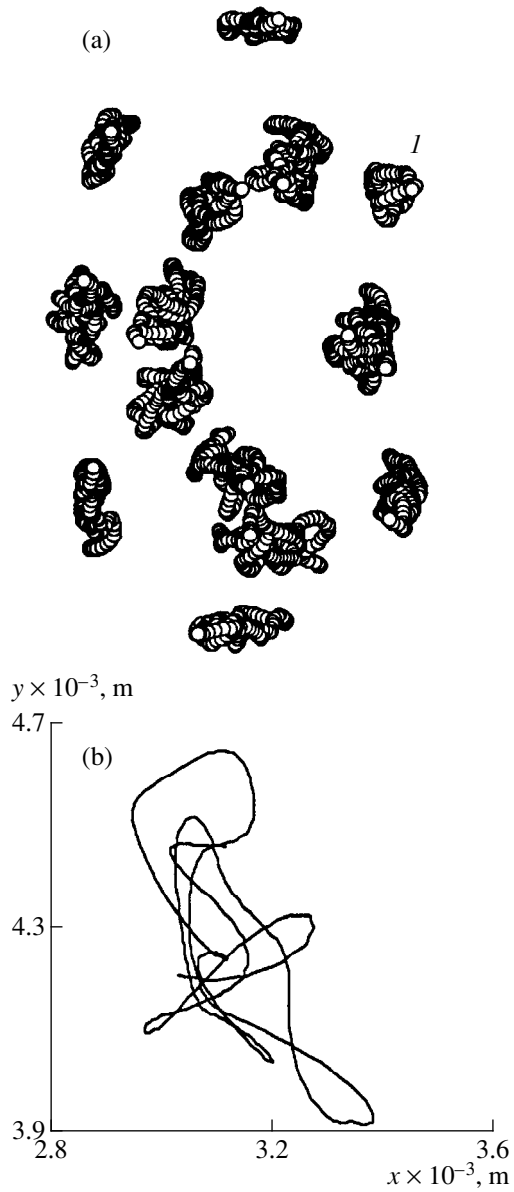


Fig. 13. (a) Illustration of particle motion over time $t = 2/\nu_{fr}$ and (b) trajectory ($t = 20/\nu_{fr}$) of an isolated particle in a system having the parameters: $\nu_{fr} \approx 1.5$ s^{-1}, $\gamma/\alpha = 0.41$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -2$ cm^{-2}, $\beta_y^* = 0$, $N_p = 15$, and $\langle K_x \rangle = \langle K_y \rangle = \langle K_z \rangle \approx 0.5$ eV.

in the particle system decrease where the conditions (10) for the establishment of large-scale vortex motion and strict charge symmetry in selected directions are absent. For illustration Figs. 12a and 12b show a steady-state dust cloud consisting of 15 particles before the onset of dispersive instability ($\nu_{fr} > \nu_c \approx 4$ s^{-1}, $\omega_p \approx \omega_\Omega \approx 55$ s^{-1} for $\nu_{fr} = \nu_c$) in a dust system having parameters similar to those of the problem in Section 3.3.1. The particle trajectories after the evolution of bifurcation ($\nu_{fr} < \nu_c$) are shown in Figs. 13a and 13b for the case $\nu_{fr} \approx 1.5$ s^{-1}.

The particle motion was irregular and the velocity distributions were uniform over directions and close to a Maxwellian distribution with the temperature $T_s \approx 2\langle K \rangle/3$ (see Fig. 14). The particle kinetic energy $\langle K \rangle$ was not inversely proportional to ν_{fr}. For example, when ν_{fr} decreased from 4 to 2 s^{-1}, the value of $\langle K \rangle$ increased from approximately $2.5T_p$ to $20T_p$. On the one hand, this may suggest a correlation between the ensuing random forces and the Brownian motion of the particles. On the other hand, the dependence of $\langle K \rangle$ on ν_{fr} may reflect the sequence of evolving bifurcations as

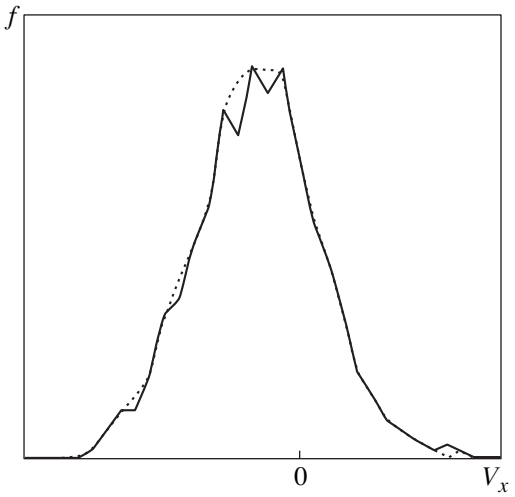


Fig. 14. Distribution $f(V_x)$ of particle velocities V_x in the direction of the x-axis (continuous curve) for a system having the parameters: $\nu_{fr} \approx 1.5\ s^{-1}$, $\gamma/\alpha = 0.41$, $Z_0 = 4 \times 10^4$, $\beta_\rho^* \approx -2\ cm^{-2}$, $\beta_y^* = 0$, and $N_p = 15$, and approximation of this curve (dashed line) by the Maxwell function at temperature $T_p = 2\langle K_x \rangle \approx 1$ eV.

the friction frequency ν_{fr} decreases. As we have noted, one of the main models of stochastic motion is the Lorentz system whose solutions are irregular functions of time over a wide range of parameters. As a result of a small change in the parameters of this system, its solution becomes so complex that it leads to chaos. This "few-mode chaos" occurs as a result of a cascade of bifurcations which give solutions in the form of complex limit cycles. In a computer analysis of the average particle velocities in a dust system this effect will resemble an increase in the kinetic temperature of the macroparticles. This effect was possibly observed in some studies of dust structures in an rf discharge plasma or in investigations of dust crystal "floating" with decreasing buffer gas pressure [21–23].

To conclude we stress that this behavior of dust systems as a result of spatial changes in macroparticle charge successfully models the anomalous heating effect and the formation of various types of ordered dynamic structures.

4. CONCLUSIONS

We have considered one possible mechanism for the instability of plasma–dust systems associated with the presence of a spatially varying macroparticle charge in a dust cloud formed in a trap under the action of an electric field and the Earth's gravitational field. This mechanism is attractive because it can explain a considerable range of phenomena (anomalous heating of the dust system, and the formation of vortices, regular oscillations, and various types of motion) observed in laboratory plasma–dust systems without using any other mechanisms or background energy sources.

We analyzed two main types of instabilities in plasma–dust systems with a charge gradient and described results of theoretical and numerical studies of their conditions of formation and the nature of their evolution. The analytic results can easily be adapted to any other nonconservative system whose linearized equations are similar to those studied here. The numerical simulation showed that as a result of the evolution of these instabilities in a dust system, various regular and stochastic oscillations may appear. Moreover, the excitation of these oscillations does not require any large spatial variations in the macroparticle charge and even a negligible charge gradient (~1–2%) in the dust cloud can cause appreciable rotation of the particles.

The nonequilibrium of the dust system may be caused by various other factors associated with temperature gradients, elastic collisions between macroparticles and surrounding plasma ions, fluctuations of external fields or local charge fluctuations at isolated dust particles as a result of the nonlinearity of the charging processes. However, of these mechanisms only the last two can give reasonable quantitative estimates for the high kinetic energies which can be acquired by dust particles in a laboratory plasma. Nevertheless, even these mechanisms cannot explain the formation of regular collective particle motion without additional sources to compensate for the energy scattering. The existence of a single mechanism and source for the various regular and stochastic oscillations in laboratory plasma-dust systems is unclear at present. The link between the instability of real dust systems and the spatial charge gradient of the macroparticles needs to be checked experimentally.

ACKNOWLEDGMENTS

This work was financed by the Russian Foundation for Basic Research (project no. 98-02-16825) and INTAS (grant no. YSF N 99-215).

REFERENCES

1. M. Salimullah and G. Morfill, Phys. Rev. E **59**, 2558 (1999).
2. V. E. Fortov, A. G. Khrapak, S. A. Khrapak, *et al.*, Phys. Plasmas **7**, 1374 (2000).
3. V. Schweigert, I. Schweigert, A. Melzer, *et al.*, Phys. Rev. Lett. **80**, 5345 (1998).
4. A. Ivlev, D. Samsonov, J. Goree, *et al.*, Phys. Plasmas **6**, 741 (1999).
5. V. Zhakhovskiĭ, V. Molotkov, A. Nefedov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 392 (1997) [JETP Lett. **66**, 419 (1997)].
6. E. Thomas, Jr. and M. Watson, Phys. Plasmas **6**, 4111 (1999).
7. V. Molotkov, A. Nefedov, V. Torchinskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 902 (1999) [JETP **89**, 477 (1999)].

8. V. Zhakhovskiĭ, V. Molotkov, A. Nefedov, *et al.*, in *Proceedings on Physics of Low Temperature Plasma, FNTP-98* (Petrozavodsk, 1998), p. 684.
9. V. Fortov, A. Nefedov, V. Vladimirov, *et al.*, Phys. Lett. A **258**, 305 (1999).
10. G. Morfill, H. Thomas, U. Konopka, *et al.*, Phys. Rev. Lett. **83**, 1598 (1999).
11. G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems* (Wiley, New York, 1977; Mir, Moscow, 1979).
12. T. Akhromeeva, S. Kurdyumov, and G. Malinetskiĭ, *Computers and Nonlinear Phenomena* (Nauka, Moscow, 1988).
13. *Physical Encyclopedia* (Bol'shaya Rossiĭskaya Éntsiklopediya, Moscow, 1998), Vol. 1, pp. 11, 15, 654.
14. R. K. Dodd, J. C. Eilbeck, J. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, New York, 1982; Mir, Moscow, 1988).
15. T. Nitter, Plasma Sources Sci. Technol. **5**, 93 (1996).
16. I. Bronshteĭn and K. Semendyaev, *Handbook on Mathematics* (Nauka, Moscow, 1986).
17. A. P. Nefedov, O. F. Petrov, S. A. Khrapak, *et al.*, Teplofiz. Vys. Temp. **36**, 141 (1998).
18. S. Khrapak, A. Nefedov, O. Petrov, and O. Vaulina, Phys. Rev. E **59**, 6017 (1999).
19. J. Goree, Plasma Sources Sci. Technol. **3**, 400 (1994).
20. O. Vaulina, A. Nefedov, O. Petrov, and S. Khrapak, Zh. Éksp. Teor. Fiz. **115**, 2067 (1999) [JETP **88**, 1130 (1999)].
21. H. Thomas and G. Morfill, Nature **379**, 806 (1996).
22. A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).
23. J. Pieper and J. Goree, Phys. Rev. Lett. **77**, 3137 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1163–1175.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1343–1357.

PLASMA GASES

Excitation and Decay of Low-Lying Nuclear States in a Dense Plasma Produced by a Subpicosecond Laser Pulse

A. V. Andreev[a], R. V. Volkov[a], V. M. Gordienko[a], A. M. Dykhne[b], M. P. Kalashnikov[c], P. M. Mikheev[a], P. V. Nikles[c], A. B. Savel'ev[a, *], E. V. Tkalya[a], R. A. Chalykh[a], and O. V. Chutko[a]

[a]*International Laser Center, M. V. Lomonosov Moscow State University, Moscow, 119899 Russia*
[b]*Troitsk Institute for Innovation and Thermonuclear Research, Troitsk, Moscow oblast, 142092 Russia*
[c]*Max Born Institute of Quantum Optics and Nonlinear Spectroscopy, D-12489 Berlin, Germany*
**e-mail: savelev@femto.phys.msu.su*

Received May 31, 2000

Abstract—The excitation of low-lying nuclear levels in a hot, dense plasma, produced by a subpicosecond pulse with intensity exceeding 10^{16} W/cm^2, is investigated theoretically and experimentally. The basic channels of electronic (inelastic scattering and inverse internal electron convergence) and photon (photoexcitation) excitations of such states as well as the influence of the broadening of a nuclear level on the excitation efficiency and the presence of hot electronic component are examined. The experimental data from measurements of the decay kinetics of the low-lying nuclear level 6.238 keV of the stable isotope ^{181}Ta, which were obtained on two experimental laser systems, are presented. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A number of publications on the experimental observation of nuclear processes accompanying the interaction of subpicosecond laser pulses with solid-state targets have appeared recently [1–3]. These processes become possible primarily because of the efficient generation of superthermal electrons in the plasma formed when a target is irradiated with an ultrashort laser pulse with intensity $I > 10$ PW/cm^2 (1 PW = 10^{15} W). The superthermal electrons can absorb most of the warming radiation. These investigations essentially open up a new field associated with laser stimulation of nuclear reactions. Thus, for intensities greater than the so-called relativistic limit $I\lambda^2 > 10^3$ (PW/cm^2) μm^2, when the electron "temperature" reaches hundreds of kiloelectron volts, direct excitation of a nucleus by electron impact or photoexcitation of the nucleus by hard plasma X-rays become possible. It should be noted that in this case the nuclear reactions occur not directly in the interior volume of the plasma but rather in heavy-metal (Pb, Ag) targets specially arranged around the interaction region. Experiments of this kind have been reported in [2, 3].

A fundamentally different situation arises for "moderate" intensities ranging from 10 to 100 PW/cm^2 [4, 5]. On the one hand, the "temperature" of superthermal electrons in this case ranges from 3 to 10 keV, which is sufficient for direct excitation of low-lying nuclear levels of stable as well as metastable isotopes (the standard methods of nuclear spectroscopy of such levels are based on direct excitation via "normal" states with energies above 100 keV) [6]. On the other hand, to obtain intensities 10–100 PW/cm^2 it is sufficient to use relatively cheap and commercially accessible table-top femtosecond laser systems. The important fact here is that the excitation of low-lying nuclear levels has a number of promising applications ranging from nuclear spectroscopy and separation of isotopes to producing population inversion and gamma lasing [7–9].

The detection of gamma radiation accompanying the decay of low-lying laser level 6.238 keV of the ^{181}Ta isotope, excited by irradiating a tantalum target with 200 fs and 300 PW/cm^2 laser pulses, was first reported in [1]. However, the kinetic decay curve of the excited state could not be measured in [1]. The purpose of the present work was to measure the kinetic decay curve of the excited state of the stable isotope ^{181}Ta and to estimate the number of excited nuclei, as well as to study theoretically the nonstationary excitation of low-lying nuclear states in plasma taking account of various mechanisms of broadening of a nuclear transition in a hot, dense laser plasma.

2. NONSTATIONARY EXCITATION OF LOW-LYING NUCLEAR LEVELS IN A HOT DENSE PLASMA

When a solid target is heated by an ultrashort laser pulse with intensity exceeding 10^{16} W/cm^2 for times less than 100 fs, a strongly ionized plasma is formed with close to solid-state ion density and high electron temperature, right up to 1 keV. In addition, a hot electronic component with "temperature" T proportional to the laser radiation intensity I [10] is formed in such a plasma:

$$T \propto 5 I_{16}^{0.2\ldots1.0} \text{ keV}$$

1063-7761/00/9106-1163$20.00

(here the index 16 signifies that the intensity is normalized to 10 PW/cm^2). These factors together result in a large increase in the excitation efficiency of nuclear transitions with excitation energy ε of the order of the temperature T. In the range of intensities which we considered $I < 10^{17}$ W/cm^2, this limits the energy as $\varepsilon < 15$–20 keV. We shall call such nuclear levels low-lying.

The excitation of low-lying nuclear levels in a hot dense laser plasma is possible in different channels—nuclear-electronic as well as nuclear-photonic [5]. The excitation channels such as inelastic electron scattering by nuclei [11], inverse internal electron conversion (IIEC) [12], and direct photoexcitation of the nucleus by the plasma X-rays [13] make the main contribution. It will be shown below that the latter is the dominant excitation mechanism in the case of a dense hot laser plasma. A rough estimate of the ratio of the number of photoexcited nuclei N^* to the total number of ions N in the plasma volume can be obtained assuming a Planck X-ray radiation spectrum for the plasma [4]:

$$\frac{N^*}{N} \approx \frac{\tau_p \Gamma_r}{\exp(\varepsilon/T) - 1},$$

where τ_p is the duration of the X-ray pulse of the plasma and Γ_r is the radiation width of the nuclear transition. Analysis of the experimental data shows that the hot electronic component makes the determining contribution to the X-ray spectrum of the plasma for photons in the energy range $E > 3$ keV [14]. For typical parameters of the experiment, for the ^{181}Ta isotope (ε = 6.238 keV, $\Gamma_r \sim 2.5 \times 10^3$ Hz, $\tau_p \sim 1$ ps, and $T \sim 4$ keV) this gives 100–1000 excited nuclei in a plasma volume $V \sim 10^{-11}$ cm^{-3}—the quantity detected in an experiment and consistent with the first experimental data [1].

At the same time a systematic description of the excitation dynamics of low-lying nuclear transitions in a hot dense plasma, produced by subpicosecond laser pulses, requires taking account of different excitation channels as well as the specific features of the processes occurring in such a plasma.

In the first place the lifetime of an excited low-lying nuclear level $\tau_{nuc} > 1$ ns is much greater than the heating and cooling times of the plasma (not much different in order of magnitude from the duration of the laser pulse). Therefore, such levels are excited in a strongly nonstationary regime.

It becomes important to take account of the broadening of a nuclear transition as a result of the Doppler, Zeeman, Stark, and other effects. Thus, the plasma ions possess a temperature T_i of several tens of electron volts, so that the Doppler broadening of a nuclear level reaches

$$\Delta\omega_D \approx \frac{\varepsilon}{\hbar}\frac{\sqrt{T_i/M_i}}{c} \approx 10^{14}\ \text{s}^{-1}$$

(M_i is the ion mass and c is the speed of light).

A laser plasma contains quasistatic magnetic fields [15] exceeding 1 MG, which are due to directed motion of electrons and ions. Consequently, Zeeman splitting of the nuclear levels, of the order of

$$\frac{\mu_n H}{\hbar} \approx 5 \times 10^9\ \text{s}^{-1}$$

($\mu_n = 5 \times 10^{-24}$ erg/G is the nuclear magneton and H is the intensity of the magnetic field), is possible.

Another factor leading to broadening of a nuclear transition is homogeneous (in this case collisional) broadening. However, this broadening is difficult to estimate because of the difficulty of determining the frequency of elastic collisions of nuclei in plasma [5]. Consequently, in the present paper we shall consider two cases: homogeneous broadening less than the reciprocal of the plasma emission time, $\gamma \ll \tau_p$, and the opposite case $\gamma \gg \tau_p$.

In addition, broadening exists because a plasma X-ray photon can interact with the nucleus only for a finite transit time through the plasma volume—transit-time broadening. This broadening can be estimated as

$$\Delta\omega_V = cl^{-1},$$

where l is the characteristic size of the plasma. Since the plasma layer can be represented as a flat cylinder with base diameter $d \sim 5$ μm and a long directrix $X \sim 0.5$ μm (the latter quantity is determined by the depth to which the target is heated by a heat-conduction wave in a time of the order of τ_p), the broadening due to the finite transit time of the plasma is of the order of

$$\Delta\omega_V \sim 10^{14}\text{–}10^{15}\ \text{s}^{-1}.$$

Evidently, this broadening is of the order of the Doppler broadening and is much greater than the Zeeman broadening.

Thus, to analyze the dynamics of the interaction of low-lying nuclear states in a hot dense plasma it is important to take account of the nonstationary nature of the excitation process and different kinds of broadening of the nuclear transition. To analyze the excitation dynamics we shall employ the equations for the nuclear density matrix. Assuming that only one low-lying nuclear level is excited, and neglecting transitions from an upper level to the lower level (this is possible, since the time for such transitions is long compared with the lifetime of the laser plasma), we obtain the equation

$$\begin{aligned}\frac{\partial Q(m_2, m_1)}{\partial t} &= -\frac{i}{\hbar}\langle m_2|H_1|m_1\rangle n(m_1) - \gamma Q(m_2, m_1),\\ \frac{\partial n(m_2)}{\partial t} &= \frac{2}{\hbar}\text{Im}\left[\sum_{m_1}\langle m_1|H_1|m_2\rangle Q(m_1, m_2)\right].\end{aligned} \tag{1}$$

Here m_1 and m_2 are indices which enumerate the corresponding lower and upper Zeeman sublevels, Q are the off-diagonal elements of the density matrix between

the sublevels of different levels, n is the population of the sublevels, γ is the relaxation rate of the off-diagonal matrix elements, and H_1 is the interaction Hamiltonian

$$H_1 = \int\left(\rho\varphi - \frac{1}{c}\mathbf{j}\cdot\mathbf{A}\right)dV, \tag{2}$$

where ρ and $\mathbf{j}$ are, respectively, the volume density of charge and the current density of the nuclear transition, $\varphi = \varphi_e + \varphi_f$, $\mathbf{A} = \mathbf{A}_e + \mathbf{A}_f$ are, respectively, the scalar and vector potentials of the electromagnetic field acting on a nucleus and consisting of the electromagnetic fields due to the motion of the electrons and the electromagnetic radiation field of the plasma. We shall represent the Hamiltonian as a multipole expansion in the form

$$H_1 = -\mathbf{d}\cdot\mathbf{E} - \mathbf{m}\cdot\mathbf{H} + \ldots, \tag{3}$$

where $\mathbf{d}$ and $\mathbf{m}$ are, respectively, the electric and magnetic moments of the nuclear transition and $\mathbf{E}$ and $\mathbf{H}$ are the external, with respect to the nucleus, electric and magnetic fields, respectively, and they also have electronic and field components.

In addition, it is necessary to write an equation for the average squared amplitude of the resonance field. For this, we divide the plasma volume into several volumes and assume that in each volume this squared amplitude is independent of the coordinates. To write the equation describing the temporal dynamics of the squared amplitude of the field in each of the volumes we employed the law of conservation of energy of the electromagnetic field:

$$\begin{aligned}&\frac{d}{dt}\int\left(\frac{E^2 + H^2}{8\pi}\right)dV\\ &= -\frac{c}{4\pi}\oint[\mathbf{E}\times\mathbf{H}]d\mathbf{S} - \frac{d\varepsilon_n}{dt} + \frac{d\varepsilon_e}{dt}.\end{aligned} \tag{4}$$

Here the derivative of the energy of the electromagnetic field in the volume stands on the left-hand side. The first term on the right-hand side describes the exchange of energy with the surrounding space through the surface, the second term describes the absorption of X-rays by the nuclei

$$\frac{d\varepsilon_n}{dt} = \sum_{m_2}\frac{\partial n^{ph}(m_2)}{\partial t}\omega_{m_2m_1}n_nV, \tag{5}$$

where $\partial n^{ph}(m_2)/\partial t$ is the derivative of the fraction of nuclei excited by the photon mechanism, n_n is the density of nuclei, V is the volume, and ω is the frequency of the nuclear transition.

The third term on the right-hand side of Eq. (4) describes the energy release in the volume as a result of X-ray emission accompanying the interaction of electrons with ions:

$$\frac{d\varepsilon_e}{dt} = \frac{dW}{d\omega}\Delta\omega_V V, \tag{6}$$

where $dW/d\omega$ is the X-ray spectrum, and the value of $dW/d\omega$ is taken at a point corresponding to the frequency of the nuclear transition.

If the Zeeman splitting is neglected and it is assumed that $\gamma \gg 1/\tau_p$, then an analytic solution can be obtained for Eqs. (1) and (4):

$$\begin{aligned}n_2 &= \frac{2}{\hbar^2\gamma}\left|\left\langle 2\left|H_1^e\right|1\right\rangle\right|^2 t\\ &+ \frac{2d_{21}^2}{\hbar^2\gamma}\frac{4\pi(dW/d\omega)V^{-1/3}}{6V^{-1/3} + 1/l_a}t,\end{aligned} \tag{7}$$

where

$$l_a = \frac{c\hbar^2\gamma}{8\pi d_{21}^2\varepsilon n_n} \tag{8}$$

is the resonance absorption length (d_{21} is the dipole moment of the transition and n_n is the nuclear density), $\left\langle 2\left|H_1^e\right|1\right\rangle$ is the transition matrix element, averaged over the Zeeman sublevels. It can be concluded from the solution obtained that for $l \ll l_a$ the fraction of excited nuclei in plasma is independent of the plasma volume $V = l^3$ (the number of excited nuclei grows as $\propto l^3$). For $l \gg l_a$ the fraction of excited nuclei is directly proportional to l. Of course, the number of nuclei excited in the plasma increases with the volume, but this growth is slower ($\propto l^2$) than for $l \ll l_a$. Consequently, an increase in the linear dimensions of the medium to a magnitude greater than the resonance absorption length appreciably decreases the excitation efficiency of the nuclei.

The matrix elements of the electronic-nuclear processes were calculated in the first Born approximation of scattering theory, and in so doing it was assumed that the IIEC occurs on the 1s level of a hydrogen-like ion. Under these assumptions, the matrix element of the electronic-nuclear processes is

$$\begin{aligned}&\left|\left\langle m_2\left|H_1^e\right|m_1\right\rangle\right|^2\\ &= \int\frac{2\pi\gamma e^2d_{21}^2n_em}{p_1}\ln\left(\frac{p_1 + p_2}{p_1 - p_2}\right)^2 f(p_1)dp_1\\ &+ \frac{16\pi e^2d_{21}^2n_e\hbar^3\gamma m}{p_1^3a^3}f(p_0),\end{aligned} \tag{9}$$

for a $E1$ transition and

$$\begin{aligned}&\left|\left\langle m_2\left|H_1^e\right|m_1\right\rangle\right|^2\\ &= \int\frac{2\pi\gamma e^2d_{21}^2n_e}{c}\ln\left(\frac{p_1 + p_2}{p_1 - p_2}\right)^2 f(p_1)dp_1\\ &+ \frac{16\pi e^2d_{21}^2n_e\hbar^3\gamma}{p_1^2a^3c}f(p_0)\end{aligned} \tag{10}$$

for an $M1$ transition.

Table 1. Characteristics of low-lying nuclear levels of some stable isotopes

Element	ε, eV	α	τ_{nuc}, ns	K	Z_0	Z_1	J_0, eV	J_1, eV
^{57}Fe	14 410	8.5	98	K	26	27	8500	9193
^{169}Tm	8410	285	4.08	M	42	59	2665	5260
^{181}Ta	6238	70.5	6050	M	46	63	3200	6000
^{201}Hg	1561	~10^4	1–10	N	21	43	375	1645

Note: ε, α, and τ_{nuc}—energy, internal conversion ratio, and lifetime of nuclear level; K—spectroscopic signal of the atomic shell making the main contribution to the internal conversion ratio; Z_i and J_i—ionization multiplicity and potentials of ions corresponding to the start ($i = 0$) and ($i = 1$) of the emptying of the main conversion shell.

Here p_1 and p_2 are, respectively, the initial and final momenta of the electron, $p_0 = \sqrt{2m\varepsilon}$, n_e is the electron density, m is the electron mass, and a is the Bohr radius. These matrix elements include the electronic momentum distribution function $f(p)$, which is obtained as the solution of the Boltzmann equation with a relaxational collision integral. This function is of a non-Maxwellian form with an additional maximum, describing the so-called hot electronic component. The average momentum of this electronic component is approximately proportional to the square root of the intensity of the incident laser field $\propto \sqrt{I}$ [10] and decrease with increasing the collision frequency.

Among the stable isotopes with low-lying nuclear level, the most characteristic ones are, from our standpoint, isotopes such as ^{57}Fe, ^{181}Ta, ^{169}Tm, and ^{201}Hg, whose characteristics are presented in Table 1. The low excitation energy of the first level of the isotope ^{201}Hg should result in efficient excitation by the thermal electrons of the plasma. However, the large values of the internal conversion ratio (low probability of gamma decay of the excited state) and probabilities of Auger and Koster–Kronig processes from the M shell of this atom (absence of X-rays when the vacancies in the M shell are filled when internal conversion occurs) leave only one possible detection channel—direct detection of the conversion electrons. Without going into further discussions, we note only the difficulty of experimental detection, using this method, taking account of the large number and wide energy spectrum of the plasma electrons.

Gamma decay of the excited state can be detected for the three other isotopes in Table 1. We note that for the isotope ^{57}Fe 14.41 keV gamma rays can be detected simultaneously with the K_α X-rays, emitted by iron atoms when vacancies in the K shell, which arise in the internal conversion process, are filled. The ratio of the number of photons in these two channels makes possible the direct measurement of the internal conversion ratio of the low-lying excited state.

A qualitative comparison of the characteristics of tantalum and thullium isotopes shows that the efficiency of thullium in this case should be much higher because of the larger width of the excited transition. The short lifetime τ_{nuc} of the excited nuclear state of the thullium isotope makes it possible to count on spectroscopic detection, since the broadening of the plasma spot d on account of the expansion of the plasma into vacuum is sufficiently small to obtain good spectral resolution:

$$d \propto \tau_{nuc}\sqrt{\frac{ZT_e}{M_i}} \leq 0.1 \text{ cm}$$

(Z is the average ionization multiplicity of the plasma and T_e is the temperature of the thermal electrons).

The tantalum isotope possesses two important features for experimental implementation: in the first place, for laser radiation intensity of about 50 PW/cm^2 the temperature of the hot plasma electrons is of the order of the excitation energy and, in the second place, the lifetime of the excited state is quite long, which is convenient for experimental detection of the effect on the basis of the delayed gamma emission of the plasma.

The metastable isotope ^{239}Pu (the excitation energy of the first nuclear level is 7.861 keV) lies next to the row of stable isotopes from Table 1. This isotope apparently possesses the optimal parameters for performing spectral studies (the conversion ratio $\alpha \sim 5950$ and the lifetime of the excited state $\tau_{nuc} = 36$ ps) and, in contrast to other metastable isotopes, its radioactivity is relatively low (the ground-state lifetime $\tau_{nuc}^{(g)} \approx 24000$ yr).

We shall consider the excitation of two stable isotopes: tantalum-181 and thullium-169 (see Table 1). The resonance absorption length estimated from Eq. (8) is 3 cm for the first nucleus and 3.9×10^{-3} cm for the second nucleus:

$$(d_{21}^{(\mathrm{Ta})})^2 \approx 4 \times 10^{-48} \text{ erg cm}^3,$$

$$(d_{21}^{(\mathrm{Tm})})^2 \approx 2.5 \times 10^{-45} \text{ erg cm}^3.$$

We note immediately that these quantities are always greater than the characteristic dimensions of the plasma spot in the plane of the target and inside the target.

The system of Eqs. (1) and (4) was solved numerically for various characteristics of the target and the laser pulse. Unless otherwise stated, in the exposition of the results below the following values are used for

the parameters: laser radiation intensity $I \sim 10$ PW/cm^2, lifetime of the hot plasma $\tau_p \sim 10$ ps, plasma volume $V \sim 10^{-11}$ cm^3 (the characteristic size of the plasma $l \sim V^{1/3}$), homogeneous half-width of the nuclear transition $\gamma \sim 10^{12}$ s^{-1}, and the density of ions was assumed to have the solid-state value with ion temperature $T_i \sim$ 50 eV.

To determine the role of the hot electronic component, a calculation was performed using a purely Maxwellian electronic distribution function. In this case the excitation efficiency of the tantalum nuclei was 5.33×10^{-13}. The presence of a hot electronic component increases the excitation efficiency by almost two orders of magnitude, so that this component must be taken into account in the numerical simulation of the process of excitation of nuclei in a plasma.

The calculation is also performed in the absence of photoexcitation in order to compare the contribution of various excitation channels. Specifically, for tantalum nuclei with the above-indicated parameters of the target and the laser pulse, the efficiency was 7.61×10^{-14} in the absence of photoexcitation and 1.83×10^{-11} in the presence of photoexcitation. For thullium nuclei these values were 7.92×10^{-10} and 1.64×10^{-9}, respectively. Thus, for tantalum photoexcitation is more efficient, while for thullium the contributions of the electron and photon excitation channels are comparable.

It should be noted at the same time that the kinetics of the ionization state of the plasma, neglected in our calculations, can also strongly influence the efficiency of the electronic excitation channels. Thus, the ionization multiplicity of the plasma has a large effect on the cross section for inelastic scattering of an electron by an ion. Moreover, the IIEC is possible only on an unfilled, ionized, atomic shell. Even though the electron temperature and the electron and ion densities are high in the plasma, the ionization rate can be inadequate for emptying the deep shells because of the short lifetime of a hot plasma. Using the average-charge approximation [16], we calculated the ionization kinetics of the isotopes considered in a plasma with solid state density, produced by a 200 fs laser pulse. The calculation assumed that the leading edge of the electronic temperature pulse follows the shape of the laser pulse, and the trailing edge is described by an exponential function with characteristic decay time of the order of 5 ps. The peak electron temperature was varied in the range 500–1000 eV. The computational results are presented in Table 1. The table also contains data on the spectroscopic symbol of the electron shell making the main contribution to the total internal conversion ratio as well as data on the range of ionization multiplicities, which corresponds to the start and end of the emptying of this shell during ionization of the atom. Comparing the obtained maximum ionization multiplicities of the corresponding elements with the values required for efficient switching off of conversion shows that for ^{201}Hg the

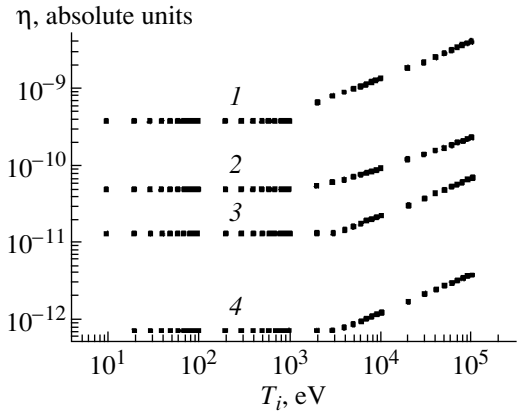


Fig. 1. Excitation efficiency of low-lying nuclear level η versus the ion temperature of the plasma T_i: (*1*) thullium, $\gamma = 10^6$ s^{-1}; (*2*) thullium, $\gamma = 10^{12}$ s^{-1}; (*3*) tantalum, $\gamma = 10^6$ s^{-1}; and (*4*) tantalum, $\gamma = 10^{12}$ s^{-1}.

N shell is essentially completely emptied for a plasma electron temperature of 500 eV, and effective excitation of the low-lying nuclear level in the IIEC channel can be expected. At the same time, for ^{181}Ta and ^{169}Tm, whose ionization multiplicities do not exceed 45 in the electron temperature range considered, the excitation channel via IIEC will be strongly suppressed. The ionization multiplicity for a ^{57}Fe isotope is small and the IIEC process is negligible.

The numerical data are presented in Figs. 1–3. In all figures, the excitation efficiency η for the nuclei—the ratio of the number of excited nuclei to the total number of nuclei located in the plasma volume—is plotted along the ordinate. Figure 1 shows the ion temperature dependences of the excitation efficiency of thullium and tantalum nuclei (i.e., a dependence on the Doppler broadening) for two different values of the homogeneous width γ. Two features should be noted. In the first place the excitation efficiency of nuclei starts to increase when the Doppler width exceeds the width of the exciting X-ray photon. In the second place if the linear size of the medium is much greater than the absorption length (as happens for thullium with $\gamma = 10^6$ s^{-1}), then the excitation efficiency grows more rapidly with increasing Doppler width. Hence follows an important conclusion: the Doppler broadening of a nuclear transition plays the most important role for media of large size (compared with the resonance absorption length). For media of small size Doppler broadening is negligible.

Figure 2 illustrates the temporal dynamics of the excitation of the nuclei. On the nonstationary section ($t < 1/\gamma$) the efficiency increases quadratically, while on the stationary section ($t > 1/\gamma$) it increases linearly. In addition, the difference between the stationary ($\gamma = 10^{12}$ s^{-1}) and nonstationary ($\gamma = 10^6$ s^{-1}) excitation regimes is illus-

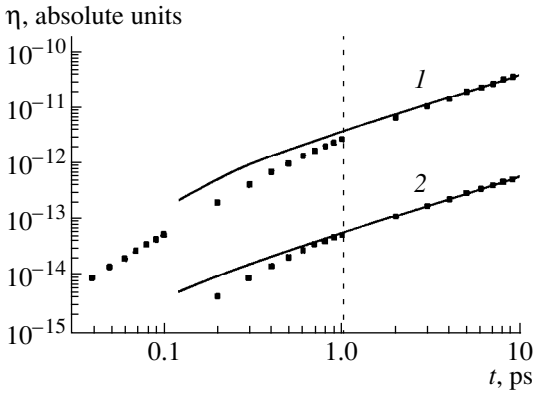


Fig. 2. Time dynamics of excitation of a low-lying nuclear level: (*1*) thullium, (*2*) tantalum. Solid lines—linear approximation in the range 1–10 ps, the dotted line corresponds to a transition from nonstationary to stationary excitation $t = 1/\gamma$.

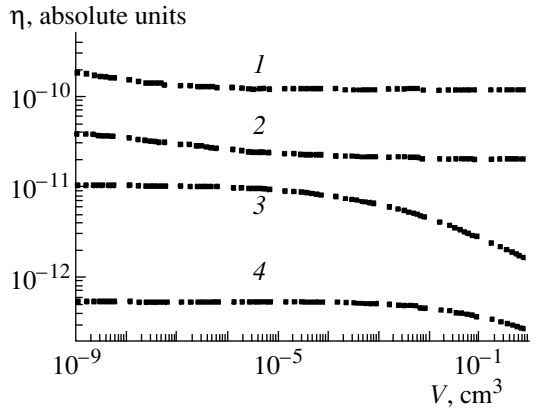


Fig. 3. Excitation efficiency η of a low-lying nuclear level versus the plasma volume *V* (see Fig. 1 for notation).

trated in Figs. 1 and 3. For large γ the excitation efficiency is lower because of the acceleration of the transverse relaxation of the nuclei.

Figure 3 shows the dependence of the excitation efficiency of the nuclei on the plasma volume. On the whole these curves confirm the behavior of the analytical solution (7). For linear size much smaller than the resonance absorption length ($l \ll l_a$), the efficiency is essentially independent of the plasma volume, and the number of excited nuclei grows as the volume. For $l \gg l_a$ the efficiency decreases as $1/l$. The number of nuclei then increases as $l^2 = V^{2/3}$. Thus, increasing the linear dimensions of a plasma above the resonance absorption length will result in a smaller fraction of excited nuclei; this occurs because of the strong absorption of X-rays in plasma. It is here that Doppler broadening of the nuclear transition plays a large role. Our calculations also established that the Zeeman splitting of the nuclear levels does not greatly influence the number of excited nuclei (the excitation efficiency of tantalum nuclei was 1.833×10^{-11} in the presence of this splitting and 1.831×10^{-11} in its absence), which is due to the smallness of this splitting because of the small value of the nuclear magneton.

3. DECAY OF LOW-LYING NUCLEAR STATES EXCITED IN A DENSE HOT LASER PLASMA

Processes occurring when the plasma undergoes cooling and recombination as it expands into the surrounding space can also have a large effect on the decay kinetics of low-lying nuclear states. Indeed, deep ionization of atomic shells can result in complete or partial suppression of internal electronic conversion—the main decay channel of an excited nuclear level [17]. In turn, this can increase the lifetime of an excited nuclear state up to values corresponding to the reciprocal of the radiation width of a level. Thus, analysis of the data in Table 1 shows that such a situation can be relatively simply realized with excitation of the low-lying level of the isotope ^{201}Hg. In addition, for high ionization multiplicity, another decay channel for low-lying nuclear levels can open up—an electronic bridge through the intermediate electronic states of a discrete spectrum [18]. This process is of third-order in the electromagnetic interaction constant for the electrons, but if the probability of conversion decay decreases as a result of ionization, it too can influence the decay kinetics of the low-lying nuclear level.

The fact that the situation in which the conversion channel for decay remains closed for only during several picoseconds, while the laser plasma is quite dense and hot, must be taken into account. However, when the expansion of the plasma into vacuum, accompanied by "quenching" of the ions, is taken into account the lifetime of the closure of the internal electron conversion channel can be expected to increase by analogy to [17]. The crux of the ion "quenching" process consists in the fact that the recombination rate for triple collisions is proportional to the squared electron density, so that as the plasma expands, when the electron density rapidly decreases with increasing distance to the target surface, a situation is created where the ions essentially stop recombining. In this case the ionization state of the plasma "becomes frozen" and the recombination rate of the plasma is determined primarily by the processes occurring in the residual gas in the chamber [16].

The processes occurring in the residual gas in the chamber must be taken into account in order to study the recombination kinetics of an expanding plasma in greater detail. Since the temperature of the residual gas is much lower than the ionization potential of the plasma ions, and the degree of ionization of the residual gas is small, $z \sim 10^{-7}$, the process determining the recombination rate of the expanding plasma will be recombination in ion–atom collisions.

Table 2. Parameters of laser systems used in the experiment

	λ, nm	τ, fs	ω, mJ	*d*, μm	*I*, PW/cm^2
1	616	200	0.3–0.7	3	10–40
2	1053	1000	2000–5000	80	30–70

Note: λ, τ, and ω—wavelength, duration, laser pulse energy; *d* and *I*—diameter of focusing spot and laser radiation intensity on the target.

The plasma expands into vacuum with characteristic velocities of the order of $v \sim 10^7$ cm/s. An ion requires about 1 μs to traverse 10 cm. The recomination rate of an ion with ionization multiplicity 50 in collisions with neutral particles in the gas with density N_a can be estimated as [16]

$$P_{ia} \approx 9 \times 10^{-17} Z^2 v N_a \approx 8 \times 10^6 \text{ s}^{-1} \qquad (11)$$

the residual-gas pressure was assumed to be ~10^{-4} torr). Then in a transit time $t \sim 0.5$ μs through a vacuum chamber of size $l \sim 5$ cm an ion undergoes about 4 recombination acts. For ^{201}Hg this will not result in filling of the conversion *N* shell and restoration of the probability of the conversion channel for decay. Thus, the lifetime of the excited state will increase (since $t \gg \tau_{nuc}$), as a result of which during the entire transit time the decay will occur only in the radiation channel. For ^{181}Ta and ^{169}Tm the partial closure of internal electron conversion can occur only for higher plasma electron temperatures of the order of 1500 eV. The thermal electronic component in the plasma of an ultrashort laser pulse cannot reach such a temperature, since as the intensity increases, the energy of the laser radiation is increasingly more efficiently absorbed by superthermal electrons, and the rate of growth of the temperature of the thermal components slows down substantially. Consequently, to calculate the ionization composition of a plasma the effect of superthermal electrons must be taken into account. The recombination of ions at residual-gas pressure ~10^{-4} torr will result in complete filling of the conversion *M* shell in the transit time up to the chamber walls. As the pressure in the chamber decreases to 10^{-5} torr, the ionization composition of the expanding plasma can be preserved for these isotopes for a long time. Thus, for sufficiently low pressures of the residual gas the conversion channel of decay of the excited state of a nucleus in an expanding plasma can be closed for the entire flight time of the ions.

For high residual-gas pressure, the plasma recombination rate can decrease if the processes in the shock wave formed are taken into account. A region of rarefaction, where plasma ions are confined and undergo recombination, forms behind the shock wave front [16]. Indeed, the gas concentration in the region behind the shock wave front can be hundreds of times less than the initial value, the recombination rate of ions in this region also decreases. In addition, the propagation velocity of the shock wave is of the order of the velocity of sound, which is much less than the ion ablation rate. Thus, the propagation of slowly recombining ions behind the shock wave front can result in prolongation of the closure of the conversion channel of decay right up to relaxation of the shock wave. Estimates show that this situation can be observed for residual-gas pressures of the order of 10 torr, and the relaxation time of the shock wave will be ~100 μs.

4. DECAY KINETICS OF THE 6.238 keV LEVEL OF ^{181}Ta EXCITED IN A HOT DENSE LASER PLASMA

The isotope ^{181}Ta was chosen for the experiments. This was determined by a number of considerations (see Table 1): the relatively low energy of the first excited level, the quite long lifetime of the excited state, making it possible to separate easily in time the gamma rays and the characteristic X-rays of the plasma, and the possibility of performing test experiments with a tungsten target. A tungsten plasma is essentially indistinguishable with respect to its integral characteristics from a tantalum plasma, since its atomic numbers differ by only one unit. At the same time, stable tungsten isotopes do not have a low-lying atomic level.

Our experiments were performed using two laser systems with different laser-pulse parameters (see Table 2). This made it possible to investigate the effect of parameters such as the wavelength of the warming radiation (which determines the temperature of the hot electron component), the laser pulse duration (which determines the lifetime of the hot electronic component, and the laser pulse energy (which, for the same intensity, determines the volume of the plasma) on the excitation efficiency of low-lying nuclear states. The experimental methods chosen were also due to the specific nature of each laser system. In this connection, we shall describe the experimental schemes and results separately for each system.

4.1. Excitation of a Plasma by Femtosecond Laser Pulses

The details of the experimental scheme for detecting gamma decay are presented in [1]. We shall briefly describe only the key aspects of the method chosen.

In the course of the experiment (Fig. 4) the dye laser radiation (no. 1 in Table 2) was focused on the surface of a Ta or W target. To perform the measurements with

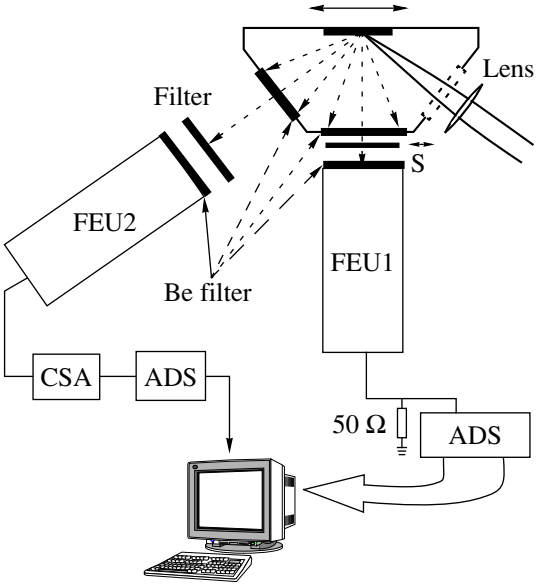


Fig. 4. Arrangement of the experiment using a femtosecond laser system (S is radioactive test source).

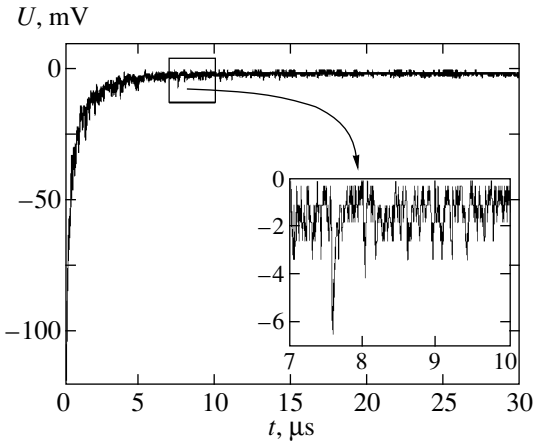


Fig. 5. Typical signal obtained from FEU1.

relatively low laser pulse energy, the radiation was focused into a spot, close in size to the diffraction limit, by a special aberration-free *F*/6 objective. In addition, the small size of the chamber (about 5 cm) made it possible to perform measurements in a large solid angle (see below). The target consisted of a 500 μm thick flat plate and was placed in a chamber with a regulatable residual-gas pressure. Moving the target after each laser burst made it possible to obtain an interaction with a "clean" surface. The experiments were performed with residual-gas pressures in the chamber of 10^{-3} and 10 torr. For 10 torr the shock wave formed with a relaxation time of the order of 100 μs confined ions inside the chamber, while for 10^{-3} torr the ions rapidly escaped from the observation region, which decreased the number of detected gamma rays with long delay times relative to the moment of plasma ignition.

The characteristic plasma X-rays and gamma rays, corresponding to decay of an isomeric level, were detected through the exit window of the chamber, which consisted of 100 μm thick beryllium foil, using two X-ray detectors FEU1 and FEU2 based on FEU-119 and a NaI(Te) scintillator. Both detectors were placed 5 cm from the target; this made it possible to receive radiation in a large solid angle 0.07 rad. The FEU1 detector served to detect the afterglow of the plasma and the gamma rays and was connected through a 50 Ω load to a fast analog-to-digital converter with conversion rate 100 MHz. The signal from the FEU2 detector was fed into a charge-sensitive amplifier (CSA) and detected using an ADC, giving in each laser burst information about the field of plasma X-rays with energy above 3 keV. Thus, the signal from the detector FEU2 made it possible to judge the quality of the focusing of the laser radiation and the plasma parameters in each experimental implementation.

The detectors were calibrated using a radioactive source S (a ^{55}Fe sample, emitting 5.9 keV Mn $K_{\alpha 1,2}$ X-rays). Assuming the signal at the output of the scintillation detector to be a linear function of the energy of the X-rays, for 6.238 keV gamma rays the average amplitude of the output signal was determined to be 1.8 ± 0.5 mV.

The basic features characteristic Ta and W were as follows for the typical implementation of an experiment in the time-resolution regime (Fig. 5) obtained with FEU1.

(1) The initial pulse with amplitude of the order of 5 V, caused by the detection of the characteristic X-ray radiation of the plasma. The trailing edge of this pulse is approximated well by the emission time of the fast component of the scintillator $\tau \sim 250$ ns.

(2) The noise component with amplitude up to 2 mV, much greater than the characteristic noise of the detector (~0.2 mV) and smoothly varying with characteristic time of the order of the emission time of the slow component of the scintillator.

The presence of a large-amplitude pulse impedes analysis at delay times of less than 3 μs relative to the laser pulse; such times are eliminated from further analysis. The gamma rays are detected (see inset in Fig. 5) against the background of a noise component whose amplitude at the initial moments of detection is comparable to the amplitude of the useful signal. In what follows, the presence of a signal of this kind in the experimental realization will be called an "event."

A control experiment with a W target was performed in order to estimate the probability of detecting gamma rays against the background noise signal. Aside from X-rays from the plasma, the X-rays from the source S also struck the FEU1 detector. On the basis of the activity of the source S, we estimated, on the basis of the 100 realizations, the probability of detecting a 5.9 keV

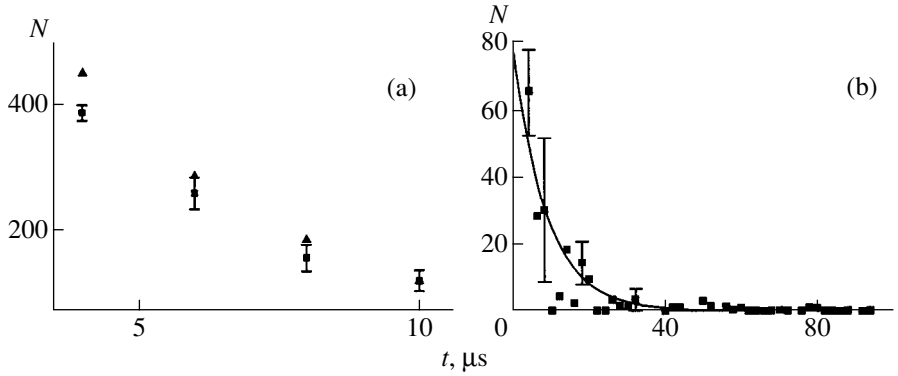


Fig. 6. Result of data analysis for residual-gas pressure in the interaction chamber 10^{-3} torr: (a) the number of events for the W (squares) and Ta (triangles) targets, (b) difference in the number of events on these targets.

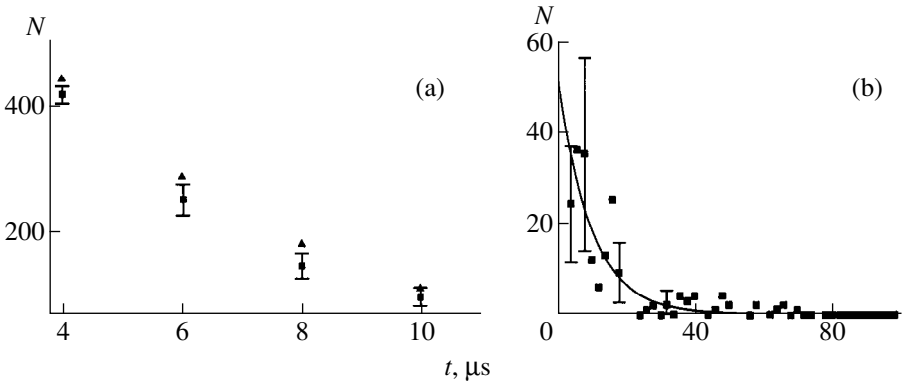


Fig. 7. Result of data analysis for residual-gas pressure in the interaction chamber 10 torr (see Fig. 6 for notation).

X-ray to be 50%. It should be noted that for 6.238 keV gamma rays this probability should be somewhat higher.

Two series of experiments, each with 120 realizations, were performed with residual-gas pressures in the chamber of 10^{-3} and 10 torr, respectively. The difference in the number of events detected on the Ta and W targets enabled us to judge the number of excited nuclei in the Ta plasma. The events detected in the experiments with W were of a purely noise nature, so that we constructed, on the basis of 320 realizations, the statistics of such noise events as a function of the delay time relative to the moment the plasma was produced. This made it possible to estimate the probability with which the number of events detected on a Ta target for a given delay does not fall within the measurement error with respect to the number of events detected under the same conditions on a W target. This error is reflected in Figs. 6 and 7, which display the experimental results for two residual-gas pressures.

Figures 6a and 7a show the delay-time dependences of the number of recorded events for Ta and W targets, and Figs. 6b and 7b show the difference of the number of events on the two targets. Analysis of the data shows that for all delay times the difference signal is greater than zero, and for most points the modulus of the difference is much greater than the computed error. The time-integral excess of the signal on the Ta target above the signal for W was 172 events for 10^{-3} torr pressure and 178 events for 10 torr pressure. These values should be interpreted as the number of gamma decays detected in each series.

Fitting the kinetic curve obtained with a exponentially decaying function

$$N^* = N_0 e^{-t/\tau_{\text{nuc}}} \tag{12}$$

gives for the 10 Torr pressure

$$\tau_{\text{nuc}} = 9.4 \pm 1.7\ \mu\text{s}, \quad N_0 = 80.$$

The estimate obtained for the lifetime of the excited low-lying nuclear state is in good agreement with the published value 8.7 μs [19]. For 10^{-3} torr the experimental curve is strongly distorted by the noise for delay times greater than 10 μs. This makes it impossible to fit parameters by numerical methods. At the same time, a fit of the function

$$N^* = N_0 e^{-t/8.7\ \mu\text{s}}$$

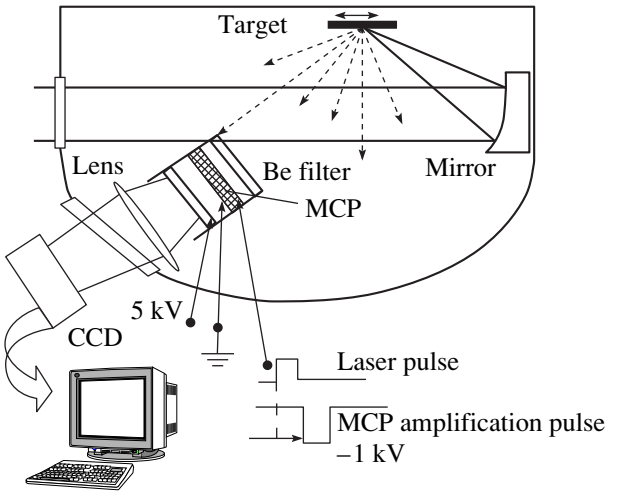


Fig. 8. Experimental arrangement using a subpicosecond laser pulse.

(see solid curve in Fig. 7b) to the experimental dependence gives a completely satisfactory result.

An estimate of the total number of nuclei excited in the laser plasma (taking account of the efficiency of the detection method, the detection solid angle of the detector, the internal electron conversion ratio) gives $(2 \pm 0.5) \times 10^4$ nuclei per laser burst.

4.2. Excitation of a Plasma by Subpicosecond Laser Pulses

In the second series of experiments a subpicosecond laser system based on neodymium glass [20] was used (no. 2 in Table 2). Thus, a much larger number of excited nuclei was expected in this experiment as a result of the higher energy of the laser pulse (and larger volume of the plasma) and as a result of an increase in temperature and the lifetime of the hot electronic component.

The same targets as in the first series of experiments were placed in a cylindrical, 85 cm in diameter, chamber (see Fig. 8). The radiation was focused on the target using an off-axis parabolic mirror with *F*/3 focusing. The target position relative to the waist of the laser beam was chosen so that the laser spot on the target would have a diameter of the order of 80 μm; this gave approximately the same intensity as in the first series of experiments. Since the plasma volume was much larger in these experiments, the number of gamma rays detected in each experimental realization was expected to be large. In this connection, an image converter with a microchannel plate (MCP), which for detection of X-rays with energy of the order of 6 keV had a quantum efficiency of about 10%, was used as a detector for X-rays and gamma rays. The spatial resolution of the image converter with the MCP makes it possible to detect a large number of X-rays simultaneously and independently. The image converter was placed at a distance of about 10 cm in an individual bay of the vacuum chamber, light-insulated from the main volume. This gave approximately the same solid angle of detection as in the first experiment. A 100 μm thick beryllium foil served as the entrance window of the bay. The arrangement of the ICP and the image converter is shown in Fig. 8. The exit surface of the ICP was grounded, and a negative 1.1 kV pulse with a rising leading edge from 10 to 1000 V in 10 ns was applied to the entrance surface. The pulse duration was chosen to be 50 μs, and the time delay relative to the moment of plasma ignition could vary from 100 ns to 100 μs. A constant 5 kV voltage was applied between the exit surface of the MCP and the image converter screen. This arrangement of the MCP prevented detection of slow electrons, knocked out of the walls of the bay by light and corpuscular fluxes. The image detected by the image converter was digitized with the aid of a CCD matrix.

Examples of the images obtained on the CCD matrix for a Ta target and three different delays of the amplification pulse of the MCP are presented in Fig. 9. Each bright dot in the image corresponds to the detection of one X-ray photon—one event. The decrease in the number of such events with increasing delay time is clearly noticeable. Similar images were also obtained for the W target. The presence of events even for this reference, target can be attributed to detection of afterglow of strongly ionized slowly recombining ions, since a natural factor determining the spectrum of the detected radiation in the detection scheme chosen is the transmission spectrum of the beryllium filter (transmittance 0.1 for 2.5 keV photons and 0.5 for 4 keV photons). Special image-recognition algorithms were used to count the number of events. As a result, dependences of the number of events on the delay time were obtained for both targets presented in Fig. 10a. In this case, because of the large number of events at each point of the time scale and their independence from one another, the experimental error was estimated assuming Poisson statistics for the events. Analysis of the difference curve (Fig. 10b) once again, just as in the first series of experiments, shows a statistically reliable excess of the number of events for the Ta target over the analogous value for the W target in the entire range of delay times. Fitting an exponential function of the form (12) to the latter dependence gives for the lifetime $\tau_{nuc} \approx 7 \pm 3$ μs. The total number of excited nuclei, taking account of the solid angle of detection of the MCP, the quantum efficiency of the MCP, and the internal electronic conversion ratio, is found to be of the order of $(5 \pm 2) \times 10^7$ in each laser burst (the fact that the experimental dependence is a time integral of the real kinetic decay curve was taken into account when making the estimate).

In the first experiment the total number of nuclei in a plasma with volume $V \approx 3 \times 3 \times 0.5$ μm^3 is 2×10^{11}. Thus, the excitation efficiency reaches $\eta_{1e} \approx (1 \pm 0.3) \times 10^{-7}$. In the second experiment the total number of

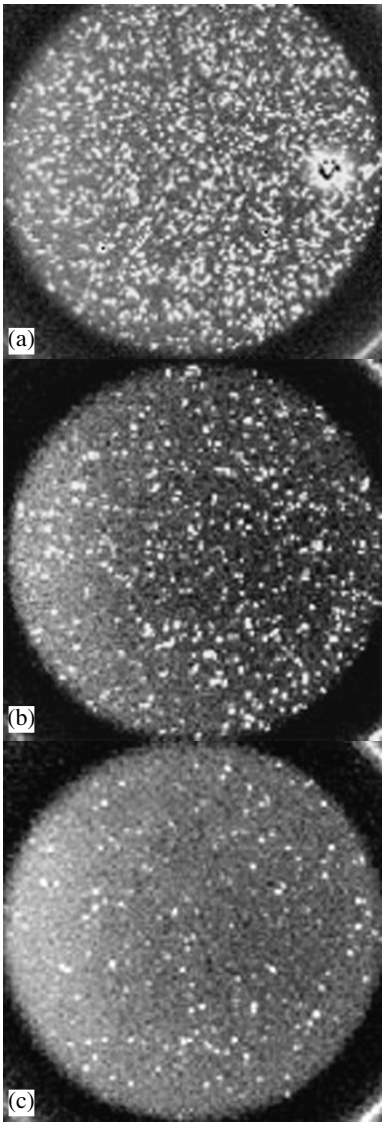


Fig. 9. Diagram of the exit window of an image converter with a MCP with delay of the MCP amplification pulse relative to the laser pulse (a) 0.1, (b) 3, and (c) 10 μs (Ta target).

nuclei in a plasma with volume $V \approx 80 \times 80 \times 0.5\ \mu m^3$ was 1.7×10^{14}, which gives for the excitation efficiency $\eta_{2e} \approx (3 \pm 1) \times 10^{-7}$. The negligible discrepancy in the estimates of the excitation efficiency, obtained in the two independent experiments, could be due to the different temperature of the hot plasma electrons as a result of the difference of the wavelengths and intensities,

$$\frac{T_1}{T_2} \propto \left(\frac{\lambda_1 I_1^2}{\lambda_2 I_2^2}\right)^{1/3} \approx 0.6$$

and to the long duration of the laser pulse in the second experiment.

Comparing the experimental estimates of the excitation efficiency with the theoretical estimates, which give $\eta_c \sim 10^{-9}$–10^{-10} for our experimental conditions (depending on the homogeneous broadening $\gamma \sim 10^6$–$10^{12}\ s^{-1}$) gives a much larger discrepancy. One of the factors that greatly influence the number of excited nuclei could be the effective increase in volume in which excitation occurs as a result of penetration of hot electrons and X-rays into the cold region of the target. Thus, the absorption length L_x of an X-ray with energy $E_x \sim 6$ keV in tantalum is

$$L_x = (\sigma(E_x)\rho)^{-1} \approx 2\ \mu m$$

($\sigma(6\ keV) \approx 300\ cm^2/g$ is the mass attenuation coefficient of metallic tantalum, $\rho = 16.6\ g/cm^3$ is its density), which is much greater than the thickness of the hot plasma layer and, in the case of the first experiment, is comparable to the transverse dimensions of the plasma spot. Therefore, the effective volume in which excitation of low-lying nuclear levels occurs can an order of magnitude or more greater than the characteristic volume of the hot plasma layer, which will proportionally decrease the experimental estimate of the efficiency η_e. It should be noted that the uncertainty in the estimate of the effective excitation volume can be removed by using thin-film samples with thickness of

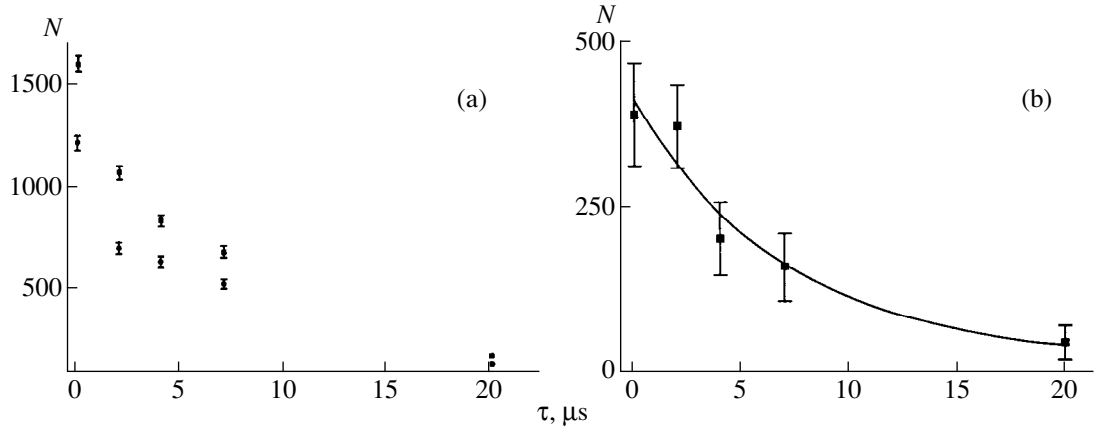


Fig. 10. Number N of events versus the delay of the MCP feed pulse τ: (a) data for Ta (circles) and W (squares) and (b) difference of the data for Ta and W.

the order of or less than the thickness of the plasma layer, i.e., several hundreds of nanometers.

5. CONCLUSIONS

Thus, our calculations have shown that efficient excitation of low-lying nuclear levels with energy of several keV occurs in a hot dense plasma produced by an ultrashort laser pulse. The presence of a nonequilibrium electronic component of the laser plasma (with energy proportional to the intensity of the incident laser radiation) is of great importance. The most important mechanisms leading to broadening of a nuclear level are Doppler broadening due to the thermal motion of ions and homogeneous collisional broadening. The magnitude of the Doppler broadening of a nuclear level has essentially no effect on the efficiency of electronic excitation and strongly influences the efficiency of radiation excitation. This efficiency increases sharply if the product of the Doppler width by the lifetime of the photon exceeds 1 (for tantalum nuclei this occurs at ion temperature 100 eV). A plasma with an anomalously high ion temperature can be obtained by using nanostructural targets [21].

The calculations showed that the optimal longitudinal size of a plasma, from the standpoint of retaining the maximum density of excited nuclei, is equal to the photoabsorption length. Its decrease has virtually no effect on the density of excited nuclei, and an increase results in a decrease of the excitation efficiency. For a low-lying level of a tantalum nucleus, excitation is most efficient in the photon channel, and for a low-lying thullium nucleus the contributions of the photon and electron channels are close.

Our experiments made it possible to detect for the first time the gamma decay of a low-lying nuclear level 6.238 keV of the stable isotope ^{181}Ta, excited in a dense hot plasma produced by an ultrashort laser pulse. Comparing with theoretical results showed that the excitation occurs not only in the volume of the plasma but also in a region with linear dimensions of the order of the order of the X-ray absorption length. Our numerical model needs further elaboration. This concerns primarily taking account of the dynamics of the plasma parameters (temperature and density of ions and electrons, charge composition) in the calculation of the excitation of low-lying nuclear transitions. A more systematic account of the structure of the atomic shells in the calculation of the electronic mechanisms of excitation of low-lying nuclear levels could also be important.

In a hot dense laser plasma the kinetics of ionization and recombination of plasma atoms can have a large effect on the probability of the conversion channel of decay of a isomeric state. Thus, for the isotope ^{201}Hg deep ionization of shells right up to the *N* shell, even at a plasma temperature of 700 eV, can result in closure of the conversion channel of decay with an increase in the lifetime of the excited nuclear state from several nanoseconds to tens of microseconds. The investigation of electronic conversion processes in strongly ionized atoms (with ionization energy of the last "unstriped" shell is of the order of the excitation energy of the nuclear level) could provide unique and direct information about the properties of such states. At the same time, new channels for decay of a nuclear level (electron bridge, and so on), investigation of which is also of great interest, could open in this situation.

Our experiment is essentially the first demonstration of newer experimental methods of nuclear spectroscopy with direct excitation of low-lying nuclear levels in a dense hot laser plasma. From this standpoint the expansion of methods of investigation is undoubtedly of interest: application of high-resolution X-ray spectral devices and detection of the energy spectrum of conversion electrons.

This field of research has undergone strong expansion because low-lying nuclear levels of metastable isotopes are being studied. One of the problems associated with low-lying levels of metastable isotopes could become the search for appropriate candidates for producing population inversion in three- and four-level schemes for a gamma laser [7, 9]. Here pairs of close-lying levels (the splitting between the levels being 1–20 keV), one of which is metastable, is necessary. The ultrashort pulses of a superstrong light field could make it possible to create such metastable isotopes (including short-lived) with the aid of ultrashort laser pulses with ultrarelativistic intensity [2, 3], as well as excitation of the metastable isotope to a close lying level with the aid of an ultrashort laser pulse of "moderate" intensity.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 97-02-17013, 98-02-16070, 99-02-18343, 00-02-17302), and project nos. 00-15-96651 and 00-15-96726 for the support of leading scientific schools. Part of this work was performed on the basis of an agreement on scientific collaboration between the International Teaching-Science Laser Center at the Lomonosov Moscow State University and the Max Born Institute of Quantum Optics and Nonlinear Spectroscopy and was financed by the Deutsche Forschungsgemeinschaft according to the project "Interaction of Superintense Laser Pulses with Matter."

REFERENCES

1. A. V. Andreev, R. V. Volkov, V. M. Gordienko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 343 (1999) [JETP Lett. **69**, 371 (1999)].
2. K. W. D. Ledingham, I. Spencer, T. McCanny, *et al.*, Phys. Rev. Lett. **84**, 899 (2000).
3. T. E. Cowan, A. W. Hunt, T. W. Phillips, *et al.*, Phys. Rev. Lett. **84**, 903 (2000).

4. A. V. Andreev, V. M. Gordienko, A. M. Dykhne, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 312 (1997) [JETP Lett. **66**, 331 (1997)].
5. A. V. Andreev, R. V. Volkov, V. M. Gordienko, *et al.*, Kvantovaya Élektron. (Moscow) **26**, 55 (1999).
6. L. V. Groshev and I. S. Shapiro, *Spectroscopy of Atomic Nuclei* (Moscow, 1952).
7. A. V. Andreev, Vestn. Mosk. Univ., Ser. Fiz., Astron. **35**, 28 (1994).
8. A. V. Andreev, V. M. Gordienko, and A. B. Savel'ev, Laser Phys. **10**, 537 (2000).
9. A. V. Andreev, V. M. Gordienko, and A. B. Savel'ev, in *Proceedings of Induced Gamma Emission'97, IGE Foundation, 1999*, p. 291.
10. P. Gibbon and R. Forster, Plasma Phys. Controlled Fusion **38**, 769 (1996).
11. E. V. Tkalya, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 441 (1991) [JETP Lett. **53**, 463 (1991)].
12. V. I. Gol'danskiĭ and V. A. Namiot, Pis'ma Zh. Éksp. Teor. Fiz. **23**, 495 (1976) [JETP Lett. **23**, 451 (1976)].
13. V. S. Letokhov and E. A. Yukov, Laser Phys. **4**, 382 (1994).
14. V. G. Babaev, M. S. Dzhidzhoev, V. M. Gordienko, *et al.*, Laser Phys. **8**, 637 (1998).
15. M. Borghesi, A. J. Mac-Kinnon, A. R. Bell, *et al.*, Phys. Rev. Lett. **81**, 112 (1998).
16. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1966).
17. A. Attallah, M. Aiche, J. F. Chemin, *et al.*, Phys. Rev. Lett. **75**, 1715 (1995).
18. V. F. Strizhov and E. V. Tkalya, Zh. Éksp. Teor. Fiz. **99**, 697 (1991) [Sov. Phys. JETP **72**, 387 (1991)].
19. *Nuclear Data Sheets*, Ed. by M. J. Martin (Academic, New York, 1966–1986), Vol. 46.
20. M. P. Kalashnikov, P. V. Nickles, and W. Sandner, Opt. Commun. **133**, 216 (1997).
21. V. M. Gordienko and A. B. Savel'ev, Usp. Fiz. Nauk **169**, 78 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1176–1182.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1358–1365.

PLASMA GASES

Model of Collective Ion Acceleration in a Vacuum Discharge Based on the Concept of a Deep Potential Well

S. A. Barengol'ts[a, *], G. A. Mesyats[b, **], and É. A. Perel'shteĭn[c, *]**

[a] *Natural-Science Research Center, Institute of General Physics, Russian Academy of Sciences, Moscow, 117942 Russia*
[b] *Institute of Electrophysics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620016 Russia*
[c] *Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
**e-mail: sb@nsc.gpi.ru*
***e-mail: mesyats@pran.ru; mesyats@nsc.gpi.ru*
****e-mail: perel@nu.jinr.ru*

Received June 20, 2000

Abstract—An ecton mechanism for the operation of the cathode spot and the concept of a deep nonstationary potential well are used as the basis to propose a model of collective ion acceleration at the spark stage of a vacuum discharge. It is shown that in principle a deep potential well can form in the presence of an external electric field and the conditions for its formation in an explosive-emission diode are clarified. The proposed model can explain the main processes leading to collective ion acceleration and shows good agreement with the experimental results. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A vacuum discharge is initiated by processes which lead to concentration of energy in microvolumes of the cathode and the formation of a plasma source which propagates into the interelectrode gap. The formation of this source concludes the vacuum breakdown and the discharge is converted to the spark stage. A spark discharge in vacuum is a high-current, self-sustained discharge where the resistance of the vacuum gap falls rapidly with time. The duration of a spark discharge is determined by the time taken for the plasma to fill the interelectrode gap after which the discharge goes over to the arc stage.

Considerable progress in the study of a vacuum spark discharge was achieved following the discovery of explosive electron emission (EEE) in 1966, when it was established that the spark current is an EEE current generated as a result of microscopic explosions at the cathode surface. Since such a microscopic explosion is short-lived, the emission of electrons in EEE takes place in separate portions known as ectons. The ecton concept has been used to obtain a logical explanation for experimental data obtained from measurements of the parameters of a cathode plasma and the erosion characteristics of cathodes, mechanisms for the self-sustainment of a spark discharge in vacuum have been determined, and so on. The current state of vacuum discharge theory is described in a recently published book [1].

However, although we now have a fairly clear idea of the physics of a spark discharge in vacuum, one of the most interesting effects accompanying the operation of this discharge has remained the subject of discussion and argument for some forty years. This effect was first observed by Plyutto in a plasma diode and involves the generation of anomalously accelerated positive ions in the form of short-lived clusters moving from the cathode toward the anode [2]. The energy of these ions is considerably higher than the potential difference U_0 applied to the gap: for instance, at 300 kV the ion energies in a vacuum diode reached 10–15 MeV [3]. An important characteristic of the light ion spectrum was that the maximum energy was proportional to the ion charge $\sim 3ZeU_0$ where Z is the ion charge. These ion energies can only be attributed to the existence of strong collective interactions between electrons and ions of the cathode plasma. Studies have shown that the anomalously high ion acceleration in the spark stage of a vacuum discharge only takes place in the unstable regime of a vacuum spark [2, 3]. This regime is characterized by abrupt bursts of current density whose amplitude is two to five times as great as the average and is accompanied by an appreciable increase in the electron beam density in the direction of ion acceleration. The electron energy in this case may reach values of the order of $3eU_0$ [4]. The current instability in an unstable regime [5] also leads to the appearance of positive ions moving toward the cathode having energies corresponding to 60–80% of the potential difference applied to the diode. In this case, the ion energy was proportional to their charge both for light and heavy ions.

In these studies a unique relationship was established between the current instabilities of a vacuum spark and the appearance of anomalously accelerated ions. A detailed study of the unstable EEE current extraction regime

was made by the authors of [6–8]. The main results described in these studies are as follows:

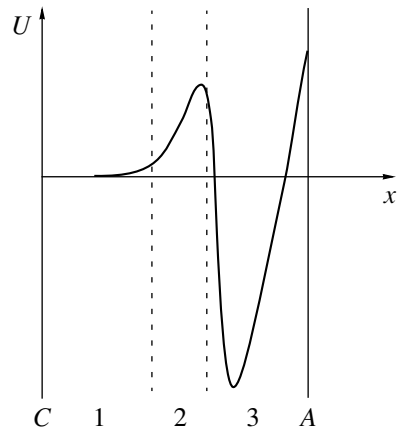
1. In the stable regime current extraction obeys the 3/2 law whereas in the unstable regime the emission current exceeds the Langmuir limit, which is manifest as abrupt short-lived ($\sim 5 \times 10^{-9}$ s) bursts whose amplitude is two to five times higher than the density of the current preceding the burst.

2. An emission current burst is accompanied by an increase in the potential of the plasma layers directly adjacent to the front of the cathode jet, by up to 80% of the potential difference U_0 applied to the diode.

Since the processes in vacuum and plasma diodes have much in common (in a vacuum diode the emission also takes from the outer plasma layers of the cathode jet), common mechanisms were proposed to explain the anomalous ion acceleration in these diodes. These mechanisms included ambipolar acceleration of ions by electrons in an expanding plasma, pinching of the plasma jet at a high rate of current growth, acceleration of a plasma cluster with a frozen-in magnetic flux, and others (see, for example [2, 9–12]). Most of these are the same as the mechanisms proposed to explain ion acceleration accompanying beam injection in a neutral gas [12]. A common disadvantage of many of the proposed models is that the anomalous ion acceleration was considered separately from the entire set of physical processes accompanying this phenomenon so that the conclusions are qualitative and the initial assumptions are not linked to the experimental data. In order to construct an accurate model of anomalous ion acceleration we need to establish a link between the bursts of explosive emission current, the increase in potential at the front of the cathode jet plasma, and finally the appearance of high-energy ions and electrons. In the present study we make a first attempt to combine these processes and obtain an overall picture of the effect.

The proposed model belongs to a class of electrostatic models in which the ions are accelerated by the electric field of the electron beam space charge, and in particular by the self-induced electric field of a virtual cathode when conditions are created for its formation at the front of a cathode jet in a vacuum diode. A virtual cathode which plays an important role in ion acceleration during free expansion of a plasma (ambipolar diffusion [2, 9]) has been observed experimentally at the front of a cathode jet plasma [13] and was used to explain ion acceleration in a diode [14]. However, the assumption that a virtual cathode is accelerated by transmitted electrons made in [14] is highly artificial.

The present model is based on the concept of a nonstationary deep potential well of a virtual cathode [12, 15] which was successfully used in a model of collective ion acceleration accompanying electron beam injection in a gas.[1] The fundamental possibility of a deep potential well being formed in a diode is shown in



Potential distribution in a diode gap when a deep nonsteady-state well is formed: (*1*) Region of quasi-neutral plasma, (*2*) region of positively charged plasma, and (*3*) deep potential well, *C* is cathode, and *A* is anode.

Section 2. Various regimes for extraction of current from a cathode jet plasma are put forward in Section 3. In Section 4 we discuss the dynamics of current bursts, the appearance of high potential at the plasma front, and collective ion acceleration during current bursts.

The scenario for collective ion acceleration in a vacuum diode may be described as follows. In the initial (comparatively short) part of the voltage pulse, the current in an explosive-emission diode is provided by thermionic emission from the edge of the cathode jet plasma. As the cathode plasma expands and the diode voltage increases, conditions of current saturation are established, limited by the emissivity of the plasma boundary layer [1, 6]. Small changes in the potential of the ionic layer near the cathode and in the potential of the plasma column, with some delay at the front of the plasma jet, promote a flow of current which corresponds to the limiting Langmuir current as the diode voltage increases [7]. This quasi-steady-state increase in current is achieved when the potential of the plasma front is close to the cathode potential.

The current behavior is then determined by the nature of the plasma influx from the cathode, which is observed as current fluctuations caused by the ecton mechanism for operation of the cathode spot. These fluctuations may lead to an abrupt increase in current in the anode circuit as a result of a substantial reduction in the space charge within the accelerating gap accompanied by quenching of the cathode current and, consequently, an abrupt increase in the potential at the edge of the plasma layer of the cathode jet as far as values comparable with the applied voltage. As a result of lag effects fast processes at the front of the jet do not appear in the cathode region and the cathode current does not change. The high potential at the front of the jet gives rise to a large electron current from the rare edge plasma and the electrons are accelerated to high energies. In the accelerating gap a deep nonstationary potential well forms near the front of the jet (see figure). Effects associated with

[1] The possible existence of an anomalous ion acceleration mechanism in a diode similar to the collective acceleration of ions in a drifting electron beam was noted by Olson [12].

its formation become significant at high diode voltages. Most of the plasma ions near the front of the jet are trapped in this well and the oscillation energy of the ions trapped in the well is higher than the energy corresponding to the applied voltage. When the well is destroyed, ions having high energies and a large energy spread move toward the anode and the cathode. Electrons having energies higher than could be produced by the potential difference applied to the diode also appear. The process of formation and destruction of the potential well may be repeated, results in multiple bursts of anode current and pulse fluxes of accelerated ions.

2. DEEP NONSTEADY-STATE POTENTIAL WELL IN A DIODE, THRESHOLD CURRENT DENSITY

It has been established experimentally that in a vacuum diode we repeatedly have the situation where a potential comparable to the anode potential appears at the front of the cathode jet. Then an electron beam having a high initial velocity and high density is injected into this gap from plasma layers adjacent to the front. The process of formation of the potential well formed by the space charge of the electrons leaving the plasma has much in common with the formation of a deep nonsteady-state well during electron beam drift beyond the anode plate in vacuum, when the depth of this well for the ions is determined by [15]

$$e\varphi_W = -\frac{8}{3}W, \tag{1}$$

where W is the electron kinetic energy, i.e., as the beam drifts behind the anode a potential well forms whose depth is approximately 2.7 times greater than the electron kinetic energy. As the deep well evolves, it becomes flattened and the depth is reduced to values close to the energy of the incoming electrons. In the limit, a nearly steady-state potential well with small fluctuations in depth is formed [15]. This well is stable with respect to relatively small deviations of the beam parameters.

In order to show that it is possible for a deep nonsteady-state potential well to form in an EEE diode, we shall consider a problem similar to [15] but including the external electric field. The solution of the problem, as in [15], is sought for planar geometry using Lagrange coordinates, where x_0 is the initial electron coordinate and t is the time. A monoenergetic beam having the electron velocity v_0 and density $n_0 = \text{const}$ is injected into the half-space $x > 0$ ($x = 0$ is the cathode plane) at time $t = 0$. The electrons propagate beyond the cathode plane in an external electric field E_0. Since we are interested in the case where space charge effects are large, we must set the condition that the beam stops (and forms a virtual cathode) at a distance from the front of the cathode jet (arbitrarily the cathode in this section) considerably shorter than the cathode–anode gap. In addition, in our problem the front of the cathode jet which moves in space, is at floating potential. However, over the times of formation of the virtual cathode the changes in the position of the front of the cathode jet and its potential are negligible and we assume that the cathode layer is fixed and at a constant potential (the positive charge needed for this is formed by the plasma ions).

The equation for the electron motion allowing for an external electric field has the form

$$\left(\frac{\partial^2 x}{\partial t^2}\right)_{x_0} = \frac{eE_0}{m} + \omega^2 x_0, \tag{2}$$

where m is the electron mass, and $\omega^2 = 4\pi e^2 n_0/m$ is the square of the Langmuir frequency at the front of the cathode jet. The first integral of this equation with the condition $v(t = -x_0/v_0) = v_0$ is given by

$$\left(\frac{\partial x}{\partial t}\right)_{x_0} = \left(\omega^2 x_0 + \frac{eE_0}{m}\right)\left(t + \frac{x_0}{v_0}\right) + v_0. \tag{3}$$

Using the condition $(\partial x/\partial t)_{x_0} = 0$ we find the time within which the first electron reflection occurs:

$$t_r = \frac{2}{\omega}\left(1 + \frac{\alpha}{2}\right), \tag{4}$$

where

$$\alpha = \frac{eE_0}{m\omega v_0}.$$

Integrating Eq. (2) with the boundary condition $x(t = -x_0/v_0) = 0$, we obtain

$$x(x_0, t) = \frac{1}{2}\left(t + \frac{x_0}{v_0}\right)^2\left(\omega^2 x_0 + \frac{eE_0}{m}\right) + v_0 t + x_0. \tag{5}$$

Quite clearly, particles which began to move from $x_0 = 0$ will make the greatest progress toward the anode at the instant of the first reflection. The condition for formation of a virtual cathode for a gap of finite length d has the form

$$x(0, t_r) = \frac{2v_0}{\omega}\left(1 + \frac{3}{2}\alpha + \alpha^2 + \frac{1}{4}\alpha^3\right) < d. \tag{6}$$

From inequality (6) we obtain the natural condition $j > j_L$, where $j = en_0v_0$ is the density of the current injected into the gap and j_L is the density of the Langmuir current. Thus, we can assume that a potential well will form when the emission current exceeds the Langmuir current (the steady-state potential well is well-known in electron tubes).

Using Eqs. (2) and (5), we can easily show that the potential minimum in the presence of an electric field at the time of the first reflection of the particles is situated at the distance $x_{\min} = 2v_0/\omega$ from the cathode (the condition that this distance should be small compared with the gap, required for our analysis to be valid, is satisfied for EEE diodes). The depth of the nonstationary potential

well does not depend on the accelerating field (if the well formation condition is satisfied) and is given by

$$e\varphi_W = -2W$$

$$\times \int_{-2-\alpha}^{-\alpha} \left[\frac{1}{2}(2+\alpha+u)^2(\alpha+u) + 2 + \alpha + u\right] du = -\frac{8}{3}W. \quad (7)$$

Thus, a deep potential well forms when current is injected into the diode gap, not only at zero applied potential (which is well-known and has been observed in experiments to study collective ion acceleration in an electron beam drift space [12]) but also at nonzero applied potential.

The solution (7) is valid for all values of α satisfying condition (6). Low values of α correspond to a vacuum diode with a thermionic cathode in the initial part of the current–voltage characteristic (the anode current is much smaller than the saturation current), electron beam injection into the half-space behind a metal anode, an EEE vacuum diode when the plasma density at the front of the plasma jet is high in a weak electric field, a plasma diode under suitable conditions, and the thermal, free expansion of a plasma, which includes recent experiments to study the exposure of macroparticles to high-power ultrashort laser pulses [16]. High values of α correspond to an electron tube in limiting current saturation regimes and a low-density plasma in strong electric fields.

When $\alpha \ll 1$ condition (6) leads to the well known formation of a virtual cathode when a beam is injected into a diode without an electric field, under the condition $v_0/(\omega d) < 1/2$ [17]. If $v_0/(\omega d) \ll 1/2$, the influence of the anode plate becomes negligible and the virtual cathode is situated at a distance of the order of $v_0/\omega \approx r_D$ from the cathode (r_D is the Debye length). Behind the virtual cathode along the beam path the space charge density is almost zero. However, it should be noted that the condition $j \gg j_L$ is far stronger than the condition for the formation of a nonsteady-state well for the case $\alpha \ll 1$.

This analysis has shown that in principle a deep nonstationary potential well can form in an explosive emission diode when the current exceeds the limiting (Langmuir) value. In this case, the virtual cathode corresponds to polarization of the charges at the plasma front with a characteristic distance of the order of the Debye length. We shall now analyze the real physical processes in an EEE diode which leads to the formation of a deep potential well at the front of the cathode jet plasma.

3. STABLE CURRENT FLOW, CURRENT BURSTS, AND MECHANISM FOR FORMATION OF A DEEP POTENTIAL WELL

As a result of a detailed experimental study of the transition from stable to unstable current extraction in an EEE vacuum diode reported in [6], it was established that this transition was caused by a reduction in the emission current from the edge of the expanding plasma of the cathode jet to a level equal to the space charge limit (Langmuir) in the gap between the front of the jet and the anode. The results of the previous section agree with this conclusion. In addition, on this basis the dynamics of the collective ion acceleration in the diode can be arbitrarily divided into two regimes corresponding to stable and unstable current flow.

In the initial part of the stable current flow regime a dense plasma having a very high emissivity appears at the front of the cathode jet. The emission current is considerably higher than the Langmuir current of the gap and most of the emission current is returned. Near the front of the jet a virtual cathode having a high charge density exists for a long time [7]. Since the main plasma column is at a low potential, a high positive ion charge forms at the front of the jet so that the overall system consisting of ions at the front of the jet and a virtual cathode is almost electrically neutral. The applied electric field is compensated by the negative space charge of the gap and a regime of thermal plasma expansion into free space is achieved [7, 9, 14, 16], i.e., collective acceleration of ions at the front of the jet by the self-induced electric field of the virtual cathode electrons. This acceleration can continue until the ion velocity at the plasma front is equal to the electron thermal velocity. In fact, the velocity of the front of the cathode jet measured in experiments using EEE vacuum diodes, which is the same as the ion velocity at the front, is determined by the electron thermal velocity. The limiting energies of the accelerated ions are higher than the electron energies in relation to their masses although these energies are considerably lower than those recorded experimentally [3].

As the cathode jet expands, the plasma density at its front decreases, and the electron density in the virtual cathode therefore decreases and tends to zero when the emission current approaches the Langmuir limiting current. The second stage then begins, i.e., unstable current flow when, as the voltage increases, the current in the diode increases, limited only by the space charge in the gap but the emission current from the front of the expanding plasma does not allow this condition to be satisfied. At this stage, the bursts of current and potential at the front of the plasma jet and, in some cases, ions accelerated to high energies arise.

In the unstable current flow regime the plasma column of the cathode jet begins to play an important role. A slow increase in current as the applied voltage increases at the previous stage is provided by the cathode current (the quasi-neutrality of the plasma is conserved) and the positive ion charge near the cathode causes an increase in potential to values of the order of the electron thermal energy at very small distances from the cathode ($\sim v_T/\omega_i$, where v_T is the electron thermal velocity and ω_i is the ion Langmuir frequency near the cathode).

An estimate of the conductivity of the plasma under typical experimental conditions in both regimes gives

voltage drops at the column of fractions of a volt for typical experimental values of the currents. However, large short-lived potential bursts comparable with the external applied voltage are observed experimentally. In order to explain this effect, very important information has been provided by experiments [7] to study the initial part of a current pulse with good time resolution, in which lag effects were observed, i.e., a finite signal propagation time, time variations of the macroscopic plasma parameters from the cathode to the front of the jet. The propagation of this type of signal in a plasma is in fact determined by the ion sound velocity in the plasma

$$v_s = v_T\left(\frac{Zm}{AM}\right)^{1/2},$$

where m is the electron mass, M is the nucleon mass, and Z and A are the averaged charge and mass number of the ions in the multicomponent plasma column. The ion sound velocity is of the same order of magnitude as the plasma expansion velocity in the hydrodynamic model. A typical value is $v_s \sim 10^6$ cm/s. The experimental results of [7] also give similar values for the propagation velocity of a perturbation of the plasma parameters at the cathode and the plasma expansion velocity.

For each position of the jet front d_f, we can introduce the characteristic time for propagation of a perturbation of the plasma parameters at the front into the plasma (toward the cathode):

$$t_d = d_f/v_s.$$

When the processes at the front are slow (these include the "quiescent" sections of the current pulse) and the characteristic time of variation of the plasma parameters at the front is $\tau \geq t_d$, the potential can be restored to an almost constant value along the plasma column and its changes at the front are negligible. Of greater interest are the short-lived changes in the parameters at the front when $\tau \ll t_d$. In this case, during the perturbing pulse the plasma parameters at the jet front cannot change as a result of rearrangement of the plasma inside the jet.

This situation is achieved, for example, as a result of plasma instability in the cathode region where, taking into account the ecton mechanism of plasma formation, current quenching can occur within a very short time. Hence the current source disappears at the jet front. The space charge in the gap between the jet front and the anode decreases within a time which can be estimated as the time taken for an electron to fly through this diode gap:

$$t_f[\mathrm{s}] \geq 3.5 \times 10^{-7}\frac{(d-d_f)[\mathrm{cm}]}{\sqrt{U_0[\mathrm{V}]}}. \qquad (8)$$

Note that this time determines the finite electron time of flight effects in nonsteady-state electron tube theory. In EEE diodes the time t_f is much shorter than t_d and determines the time of the steep rise in anode current during bursts. When the space charge is completely removed from the gap a high positive potential appears near the front of the cathode jet and decreases rapidly over the length of the plasma column. This is directly evidenced by the results of experiments described in [8], where it was established using probe diagnostics that the potential far from the plasma front does not respond to abrupt current bursts. The situation is very similar to the plasma disruption and the formation of large potential drops over short lengths proposed in [18]. A characteristic feature of the process considered here is that plasma disruption implies "emission" of electrons from the edge region of the jet and the formation of a charged plasma within a short time (of the order of the reciprocal ion Langmuir frequency), i.e., the formation of a dense ion cluster at the edge of the jet. The potential in the region adjacent to the front provides an electron current considerably higher than the Langmuir limit in the gap between the jet front and the anode. The electron flux density remains the same as in the plasma at the edge of the jet and the electron velocity increases to values close to the speed of light in the experiments [3]. As a result, a current exceeding the limit and having a high electron density is injected into the gap between the jet front and the anode and a deep nonstationary potential well is formed.

With time, the well shifts toward the anode and its depth decreases to values close to the depth of a steady-state well in a diode with an emission current exceeding the Langmuir limit. For long voltage pulses conditions are created for the appearance of repeated bursts and the process is repeated, as was observed experimentally.

4. ANALYSIS OF EXPERIMENTAL RESULTS ON CURRENT BURSTS AND ION ACCELERATION CONDITION WHEN THE DEEP POTENTIAL WELL IS DESTROYED

In order to check the statement that the short-lived increase in potential at the edge of the plasma jet is caused by the disappearance of space charge in the gap between the jet edge and the anode, we shall give additional estimates using experimental results [3, 6]. We shall consider the time interval T corresponding to a current burst. We shall determine the change in the potential ΔU at the jet edge assuming that the capacitance of the gap between the jet front and the anode is approximately equal to the capacitance of the complete diode gap C_d and we assume that current burst pulse is close to rectangular:

$$\Delta U \approx I_A T/C_d, \qquad (9)$$

where I_A is the amplitude of the anode burst current. Substituting into formula (9) the burst current 5 kA, the time $T = 1.5 \times 10^{-8}$ s, and the gap capacitance 250 pF [3], we obtain $\Delta U \approx 300$ kV.

In order to obtain an order-of-magnitude estimate of the potential jumps ΔU in the experimental studies [6], we shall take the corresponding approximate parameters: capacitance per unit area of anode for an equivalent planar diode $c_d \approx 200$ pF/cm^2, $j_A = 10^3$ A/cm^2, $T =$ 5 ns, and then in accordance with Eq. (9) $\Delta U \approx 25$ kV, i.e., as in the experiments [3], ΔU is of the order of the applied voltage.

Thus, using experimental data [3, 6] we have shown that assuming that the current burst involves dumping of space charge in the diode, we can predict a large potential burst at the plasma front. The values of the potential at the plasma front may reach the potential applied to the diode. Note that an estimate of the burst time using Eq. (8) gives values $T \geq t_f = 2$ ns and $T \geq 1$ ns too low for the experiments described in [6] and [3] because the potential at the jet front is, in our case, a floating potential.

In order to estimate the time taken for recovery of the space charge as a result of the formation of a virtual cathode in the gap between the plasma front and the anode and thus the time taken for recovery of almost zero potential at the jet front, we shall again use Eq. (9) with the anode current replaced by the Langmuir current I_L since the current leaving the plasma has the same order of magnitude. Bearing this in mind, we can write the characteristic decay time for the potential U, i.e., the time taken to form a virtual cathode,

$$T_v \approx C_d U / I_\mathrm{L}. \tag{10}$$

It follows from Eqs. (9) and (10) in particular that the intervals between the bursts are greater than the duration of the bursts approximately in the ratio j_A/j_L which is consistent with the oscilloscope traces of the current given in [6] where the ratio of the burst durations and the ratio of the intervals between them are of the same order and $j_A/j_\mathrm{L} \approx 2$–5. Substituting into Eq. (10) corresponding data from [3] $U_0 \approx 300$ kV, $I_\mathrm{L} \approx 1$ kA, and $C_d \approx 250$ pF, we obtain $T_v \approx 10^{-7}$ s; however in these experiments the maximum voltage was achieved within approximately the same time and naturally no repeated abrupt current peaks were observed.

We shall now consider the possibility of ion acceleration as a result of the formation and destruction of a deep potential well. Effective acceleration can be achieved if the lifetime of the well is comparable with the oscillation period of the ions trapped in the well or is at least more than half this. In order to estimate the half-period of the oscillations we shall assume that the depth of the well is $3U_0$ and the half-width is L. The square of the ion oscillation frequency is estimated as

$$\Omega^2 \approx \frac{3Z}{A}\frac{eU_0}{ML^2}. \tag{11}$$

For the acceleration of singly charged aluminum ($A = 27$), which was observed experimentally [3], at $U \approx 300$ kV the lifetime of a deep potential well should be greater than

$$T_W[\mathrm{s}] \approx 10^{-8} L. \tag{12}$$

Since for the experiments [3] this time can be estimated as $\leq 10^{-8}$ s the characteristic half-width of the potential well should be $L \leq 1$ cm.

However, the half-width of the potential well in accordance with Section 2 is estimated as

$$L \approx 2v_0/\omega. \tag{13}$$

This formula holds for the weak relativity which occurred in these experiments.

In the stable current flow regime the current density in the experiments [3] reached values of 100 A/cm^2. In accordance with the assumed concepts of current in the stable regime, this is provided by electrons having energies of the order of the thermal energy at the jet front, i.e., of the order of 1 eV. The corresponding plasma density is of the order of 10^{13} cm^{-3} and the Langmuir electron frequency ω is approximately 2×10^{11} s^{-1}.

Substituting these values of the Langmuir frequency and the electron velocity into Eq. (13), we obtain the half-width of the deep nonstationary potential well $L \approx$ 0.3 cm (the gap length in these experiments is 2 cm). Condition (12) required for the appearance of ions accelerated to maximum energies of around $3ZW$ (W is the energy of the accelerated electrons) during rapid destruction of the potential well is satisfied. Bearing in mind that a high-current spark discharge contains ions having degrees of ionization around 10 [1], and that the kinetic energy of the electrons leaving the plasma is of the order of eU_0, for the experimental conditions [3], i.e., $U_0 = 300$ kV, we obtain maximum ion energies of 10 MeV. Ions having these energies were recorded in these experiments. Thus, the concept of a deep nonstationary potential well can explain the collective acceleration of ions in an EEE vacuum diode.

5. CONCLUSIONS

We have shown that it is possible for a deep nonstationary potential well to form in the presence of an electric field in the diode gap and we have proposed a model of the current flow process and the formation of such a well in an EEE vacuum diode taking into account an ecton mechanism for the operation of the cathode spot, i.e., portioned emission from the cathode. The proposed model of collective ion acceleration in a vacuum discharge based on the concept of a deep potential well shows good agreement with known experimental observations. Problems involved in refining the model and allowing for two-dimensional effects (filament formation process, and so on) will be explained in a subsequent simulation and experiments.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 99-02-16462).

REFERENCES

1. G. A. Mesyats, *Ectons in Vacuum Discharge: Breakdown, Spark, Arc* (Nauka, Moscow, 2000).
2. A. A. Plyutto, Zh. Éksp. Teor. Fiz. **39**, 1589 (1960) [Sov. Phys. JETP **12**, 1106 (1960)].
3. E. D. Korop and A. A. Plyutto, Zh. Tekh. Fiz. **40**, 2534 (1970) [Sov. Phys. Tech. Phys. **15**, 1986 (1970)].
4. K. V. Suladze and A. A. Plyutto, Zh. Tekh. Fiz. **37**, 72 (1967) [Sov. Phys. Tech. Phys. **12**, 48 (1967)].
5. E. D. Korop and A. A. Plyutto, Izv. Vyssh. Uchebn. Zaved., Fiz. **4**, 131 (1973).
6. D. I. Proskurovskiĭ, V. P. Rotshteĭn, A. F. Shubin, *et al.*, Zh. Tekh. Fiz. **45**, 2135 (1975) [Sov. Phys. Tech. Phys. **20**, 1342 (1975)].
7. G. P. Bazhenov, O. B. Ladyzhenskiĭ, E. A. Litvinov, *et al.*, Zh. Tekh. Fiz. **47**, 2086 (1977) [Sov. Phys. Tech. Phys. **22**, 1212 (1977)].
8. G. P. Bazhenov, O. B. Ladyzhenskiĭ, S. M. Chesnokov, *et al.*, Zh. Tekh. Fiz. **49**, 117 (1979) [Sov. Phys. Tech. Phys. **24**, 67 (1979)].
9. A. V. Gurevich, L. V. Pariĭskaya, and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **63**, 516 (1972) [Sov. Phys. JETP **36**, 274 (1973)].
10. E. D. Korop, Zh. Tekh. Fiz. **46**, 2187 (1976) [Sov. Phys. Tech. Phys. **21**, 1284 (1976)].
11. I. A. Krinberg, J. Phys. D **29**, 2049 (1996).
12. C. L. Olson and U. Shumacher, *Collective Ion Acceleration* (Springer-Verlag, Berlin, 1979).
13. S. P. Gorbunov, V. I. Krasov, and V. L. Paperny, J. Phys. D **30**, 1922 (1997).
14. A. A. Plyutto, K. V. Suladze, S. V. Temchin, *et al.*, At. Énerg. **27**, 418 (1969).
15. J. W. Poukey and N. Rostoker, Plasma Phys. **13**, 897 (1971).
16. A. Mackinnon, C. Brown, T. Cowan, *et al.*, in *Proceedings of the 41rs Annual Meeting of the Division of Plasma Physics, Seattle, 1999.*
17. R. B. Miller, *Introduction to the Physics of Intense Charged Particle Beams* (Plenum, New York, 1982; Mir, Moscow, 1984).
18. A. A. Plutto, K. V. Suladze, E. D. Korop, *et al.*, in *Proceedings of the V International Symposium on Discharges and Electrical Insulation in Vacuum, Poland, 1972*, p. 145.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1183–1189.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1366–1373.

FLUIDS

Phase Separation and Vortex States in the Binary Mixture of Bose–Einstein Condensates¶

S. T. Chui[a], V. N. Ryzhov[b, *], and E. E. Tareyeva[b]

[a]*Bartol Research Institute, University of Delaware, Newark, Delaware, 19716 USA*
[b]*Institute for High Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
**e-mail: ryzhov@hppi.troitsk.ru*

Received May 31, 2000

Abstract—The phase separation and vortex states in a two-component Bose–Einstein condensate consisting of $|F = 1, m_f = -1\rangle$ and $|2, 1\rangle$ internal spin states of ^{87}Rb atoms are considered in the framework of the Thomas–Fermi approximation. It is shown that in the nonrotating system, the atoms in the state $|1, -1\rangle$ form a shell around the atoms in the state $|2, 1\rangle$. The critical angular velocity for each state is calculated. These velocities depend drastically on the relative concentrations of the components, the critical angular velocity of the outer component being less than the angular velocity of the inner one. It is shown that the atoms in the $|1, -1\rangle$ state can form a rotating ring around the resting core of the atoms in the $|2, 1\rangle$ state. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The implementation of Bose–Einstein Condensation (BEC) in dilute atomic gases offers new opportunities in studying quantum degenerate fluids [1]. These condensates, which contain thousands of atoms confined to microscale clouds, have similarities to superfluidity and laser and provide a new testing ground for multibody physics.

The modern theoretical description of dilute BEC originates from Bogoliubov's seminal 1947 paper, in which he showed that a weak repulsive interaction qualitatively changes the excitation spectra from the quadratic free particle form to a linear phononlike structure. To describe the trapped condensates at T = 0, one can use the Gross–Pitaevskii (GP) (nonlinear Schrödinger) equation for the condensate wave function [2]. This equation generalizes the Bogoliubov theory to the inhomogeneous phase. It was widely used in discussing the ground-state properties and collective excitations in BEC.

Bulk superfluids are distinguished from normal fluids by their ability to support dissipationless flows. This ability is closely related to the existence of stable quantized vortices. These vortices have been extensively studied in superfluid ^{4}He. Recently, clear experimental evidence of the existence of a vortex in trapped BEC [3–5] was reported. Unlike superfluid helium, the trapping potential makes alkali BEC nonuniform. Theoretical work has been concentrated on the critical angular velocity of the vortex creation, the collective excitations of BEC in the presence of a vortex, and vortex stability considerations [6–13]. It was shown that in contrast with superfluid helium, where the vortex is locally stable, quantized vortices existing in weakly interacting gases can be stable only in a driven system and become unstable without an imposed rotation. Therefore, this system cannot be superfluid [6, 10, 11].

The most interesting behavior found in the study of quantum fluids was in fluid mixtures. Two experimental groups have shown trapped multiple condensates in a magnetic trap in rubidium [14] and in an optical trap in sodium [15]. In these experiments, the spatial separation of condensates has been observed. One can distinguish two types of spatial separation: (a) the potential separation caused by external trapping potentials and (b) the phase separation that can occur in the absence of external potentials due to the interaction between two condensates. It is the latter type of phase separation that we consider in this paper. This type of phase separation has been observed in experiments on simultaneously trapped condensates consisting of the ^{87}Rb atoms in the $|2, 1\rangle$ and $|1, -1\rangle$ spin states (states 2 and 1, respectively) [14]. In this case, the respective intraspecies and interspecies scattering lengths denoted as a_{11}, a_{22}, and a_{12} are in the proportion $a_{11} : a_{12} : a_{22} = 1.03 : 1 : 0.97$ with the average of the three being 55(3) Å [14, 16].

In this paper, we consider the behavior of the binary mixture of Bose–Einstein condensates of alkali atoms. We calculate the critical angular velocity needed to create stable vortices in either component in the rotating frame. This quantity is crucially important in view of the experimental possibility of creating vortices by rotating the confining trap [5].

The physics of interpenetrating Bose fluids is very rich and far from complete understanding. While the properties of rotating single-component Bose fluids

¶This article was submitted by the authors in English.

have been studied very well [6–13], the rotating mixture of alkali atoms presents a new frontier and is essentially virgin territory.

In order to derive analytic results, some approximations must be used. A commonly used one is the Thomas–Fermi approximation (TFA), which ignores the kinetic energy terms. It has been shown that for one-component condensates, the TFA results agree well with the numerical calculations for large particle numbers except for a small region near the boundary of the condensate [7, 8]. In fact, even for a small number of particles, the TFA still usually gives qualitatively correct results. The TFA provides an excellent starting point of study. However, the TFA should not be relied upon when a quantitative comparison of experiment and theory is important. In this case, a numerical approach based on the Monte Carlo simulation becomes necessary.

2. PHASE SEPARATION IN THE NONROTATING CONDENSATE

We first consider the phase separation in the binary mixture without rotation.

For a two-species condensate, we let $\psi_i(\mathbf{r})$ ($i = 1, 2$) denote the wave function of species i with the particle number N_i. We can then write two coupled nonlinear Schrödinger (Gross–Pitaevskii) equations

$$-\frac{\hbar^2}{2m_1}\nabla^2\psi_1(\mathbf{r}) + \frac{1}{2}m_1\omega_1^2(x^2 + y^2 + \lambda^2 z^2)\psi_1(\mathbf{r}) - \mu_1\psi_1(\mathbf{r}) + G_{11}|\psi_1(\mathbf{r})|^2\psi_1(\mathbf{r}) + G_{12}|\psi_2(\mathbf{r})|^2\psi_1(\mathbf{r}) = 0, \quad (1)$$

$$-\frac{\hbar^2}{2m_2}\nabla^2\psi_2(\mathbf{r}) + \frac{1}{2}m_2\omega_2^2(x^2 + y^2 + \lambda^2 z^2)\psi_2(\mathbf{r}) - \mu_2\psi_2(\mathbf{r}) + G_{22}|\psi_2(\mathbf{r})|^2\psi_2(\mathbf{r}) + G_{12}|\psi_1(\mathbf{r})|^2\psi_2(\mathbf{r}) = 0. \quad (2)$$

Equations (1) and (2) were obtained by minimizing the energy functional of the trapped bosons of masses m_1 and m_2 given by

$$E(\psi_1, \psi_2) = \int d^3r\Big[\frac{\hbar^2}{2m_1}|\nabla\psi_1(\mathbf{r})|^2 + \frac{1}{2}m_1\omega_1^2(x^2 + y^2 + \lambda^2 z^2)|\psi_1(\mathbf{r})|^2 + \frac{\hbar^2}{2m_2}|\nabla\psi_2(\mathbf{r})|^2 + \frac{1}{2}m_2\omega_2^2(x^2 + y^2 + \lambda^2 z^2)|\psi_2(\mathbf{r})|^2 + \frac{G_{11}}{2}|\psi_1(\mathbf{r})|^4 + \frac{G_{22}}{2}|\psi_2(\mathbf{r})|^4 + G_{12}|\psi_1(\mathbf{r})|^2|\psi_2(\mathbf{r})|^2\Big]. \quad (3)$$

The chemical potentials μ_1 and μ_2 are determined by the relations $\int d^3r\,|\psi_i|^2 = N_i$. The trap potential is approximated by an effective three-dimensional harmonic-oscillator potential well, which is cylindrically symmetric about the z-axis, with the ratio $\lambda = \omega_{zi}/\omega_i$ of the angular frequency in the axial direction ω_{zi} to that in the transverse direction. The experimental value of λ is $\lambda = \sqrt{8}$. The interaction strengths, G_{11}, G_{22}, and G_{12} are determined by the s-wave scattering lengths for binary collisions of the same and distinct bosons: $G_{ii} = 4\pi\hbar^2 a_{ii}/m_i$ and $G_{12} = 2\pi\hbar^2 a_{12}/m$, where $m^{-1} = m_1^{-1} + m_2^{-1}$.

We now consider the phase separation due to the interaction between the two condensates. In this case, we have

$$\frac{1}{2}m_1\omega_1^2 = \frac{1}{2}m_2\omega_2^2. \quad (4)$$

We can simplify the equations using dimensionless variables. We define the length scale

$$a_\perp = \left(\frac{\hbar}{m_1\omega_1}\right)^{1/2} \quad (5)$$

and the dimensionless variables

$$\mathbf{r} = a_\perp\mathbf{r}', \quad (6)$$

$$E = \hbar\omega_1 E', \quad (7)$$

$$\psi_i(\mathbf{r}) = \sqrt{N_i/a_\perp^3}\,\psi_i'(\mathbf{r}'). \quad (8)$$

The wave function $\psi_i'(\mathbf{r}')$ is normalized to 1. In terms of these variables, the Gross–Pitaevskii energy functional takes the form

$$E' = \frac{1}{2}\int d^3r'\Big[N_1|\nabla'\psi_1'|^2 + N_1(x'^2 + y'^2 + \lambda^2 z'^2)|\psi_1'|^2 + N_2\beta^2|\nabla'\psi_2'|^2 + N_2(x'^2 + y'^2 + \lambda^2 z'^2)|\psi_2'|^2 + \frac{1}{2}N_1u_1|\psi_1'|^4 + \frac{1}{2}N_2u_2\beta^2|\psi_2'|^4 + \frac{2\pi a_{12}}{a_\perp}\frac{m_1}{m}N_1N_2|\psi_1'|^2|\psi_2'|^2\Big], \quad (9)$$

where $\beta^2 = m_1/m_2 = \omega_2^2/\omega_1^2$ and $u_i = 8\pi a_{ii}N_i/a_\perp$. In deriving Eq. (9), we used Eq. (4). Equations (1) and (2) can be rewritten as

$$-\nabla'^2\psi_1' + (x'^2 + y'^2 + \lambda^2 z'^2)\psi_1' - \mu_1'\psi_1' + u_1|\psi_1'|^2\psi_1' + \frac{4\pi a_{12}N_2}{a_\perp}\frac{m_1}{m}|\psi_2'|^2\psi_1' = 0, \quad (10)$$

$$-\beta^2\nabla'^2\psi_2' + (x'^2 + y'^2 + \lambda^2 z'^2)\psi_2' - \mu_2'\psi_2' + u_2\beta^2|\psi_2'|^2\psi_2' + \frac{4\pi a_{12}N_1}{a_\perp}\frac{m_1}{m}|\psi_1'|^2\psi_2' = 0, \quad (11)$$

where $\mu_i' = 2\mu_i/\hbar\omega_1$.

In the TFA, Eqs. (9), (10), and (11) are further simplified by omitting the kinetic energy. The phase segregated condensates do not overlap, and we can therefore neglect the last terms in Eqs. (9), (10), and (11). In separate regions that do not overlap, Eqs. (10) and (11) are then reduced to simple algebraic equations

$$\left|\psi_1'(\mathbf{r}')\right|^2 = \frac{1}{u_1}[\mu_1' - (\rho'^2 + \lambda^2 z'^2)], \tag{12}$$

$$\left|\psi_2'(\mathbf{r}')\right|^2 = \frac{1}{u_2\beta^2}[\mu_2' - (\rho'^2 + \lambda^2 z'^2)], \tag{13}$$

where $\rho'^2 = x'^2 + y'^2$. From Eqs. (12) and (13), one can see that the condensate density has an ellipsoidal form.

In the case of phase separation, the energy of the system can be written as

$$E = E_1 + E_2, \tag{14}$$

where

$$E_1 = \frac{1}{2}\hbar\omega_1 N_1\left[\mu_1' - \frac{1}{2}u_1\int d^3r'\left|\psi_1'\right|^4\right], \tag{15}$$

$$E_2 = \frac{1}{2}\hbar\omega_1 N_2\left[\mu_2' - \frac{1}{2}u_2\beta^2\int d^3r'\left|\psi_2'\right|^4\right]. \tag{16}$$

Equations (12) and (13) have been used in deriving Eqs. (15) and (16).

To investigate the phase separation in the mixture, we first assume that the condensate 1 atoms form an ellipsoidal shell around the condensate 2 atoms (we refer to this configuration as configuration *a*). This form of the condensate is determined by the form of the confining potential. To find the position of the boundary between the condensates, we use the thermodynamic equilibrium condition [17], according to which the pressures exerted by each condensates must be equal:

$$P_1 = P_2. \tag{17}$$

The pressure is given by [18]

$$P_i = \frac{G_{ii}}{2}\left|\psi_i\right|^4. \tag{18}$$

Condensate 2 has the form of an ellipsoid with a long semiaxis q,

$$\rho'^2 + \lambda^2 z'^2 = q^2. \tag{19}$$

Equations (12), (13), and (17)–(19) imply the equation for q:

$$\mu_1' - q^2 = \kappa\mu_2' - \kappa q^2, \tag{20}$$

where $\kappa = \sqrt{a_{11}m_2/(a_{22}m_1)}$.

The chemical potentials and μ_2' can be obtained using the normalization conditions

$$\int d^3r'\left|\psi_1'\right|^2 = \int d^3r'\left|\psi_2'\right|^2 = 1$$

and are given by

$$\mu_1' = \frac{\mu_1^0}{\left(1 - \frac{5}{2}q'^3 + \frac{3}{2}q'^5\right)^{2/5}}, \tag{21}$$

$$\mu_2' = \frac{3}{(\mu_1')^{3/2}q'^3}\left(\frac{2\beta^2(\mu_2^0)^{5/2}}{15} + \frac{(\mu_1')^{5/2}q'^5}{5}\right), \tag{22}$$

where $q = \sqrt{\mu_1'}\,q'$ and

$$\mu_i^0 = \left(\frac{15\lambda u_i}{8\pi}\right)^{2/5}. \tag{23}$$

Equations (21)–(23) allow us to find the chemical potentials μ_1' and μ_2' and the semiaxis of the phase boundary ellipsoid q as functions of N_1 and N_2. The energy $E_a = E_{a1} + E_{a2}$ of configuration *a* is given by

$$E_{a1} = \frac{1}{2}\hbar\omega_1 N_1\left\{\mu_1' - \frac{15}{4}\frac{(\mu_1')^{7/2}}{(\mu_1^0)^{5/2}} \times\left[\frac{8}{105} - \left(\frac{q'^3}{3} - \frac{2}{5}q'^5 + \frac{q'^7}{7}\right)\right]\right\}, \tag{24}$$

$$E_{a2} = \frac{1}{2}\hbar\omega_1 N_2\left\{\mu_2' - \frac{15}{4}\frac{(\mu_1')^{3/2}}{\beta^2(\mu_2^0)^{5/2}} \times\left(\mu_2'^2\frac{q'^3}{3} - 2\mu_2'\mu_1'\frac{q'^5}{5} + \mu_1'^2\frac{q'^7}{7}\right)\right\}. \tag{25}$$

We now consider the opposite case, where the condensate 2 atoms form an ellipsoidal shell around the condensate 1 atoms (configuration *b*). In this case, Eqs. (20)–(25) can be rewritten as

$$\mu_1'' - q_1^2 = \kappa(\mu_2'' - q_1^2), \tag{26}$$

$$(\mu_2'')^{5/2} = \frac{\beta^2(\mu_2^0)^{5/2}}{1 - \frac{5}{2}q_1'^3 + \frac{3}{2}q_1'^5}, \tag{27}$$

$$\frac{15}{2}\frac{(\mu_2'')^{3/2}}{(\mu_1^0)^{5/2}}\left(\frac{\mu_1''q_1'^3}{3} - \frac{\mu_2''q_1'^5}{5}\right) = 1, \tag{28}$$

$$E_b = E_{b1} + E_{b2},$$

$$E_{b1} = \frac{1}{2}\hbar\omega_1 N_1\left\{\mu_1'' - \frac{15}{4}\frac{(\mu_2'')^{3/2}}{(\mu_1^0)^{5/2}} \times\left(\mu_2''^2\frac{q_1'^3}{3} - 2\mu_1''\mu_2''\frac{q_1'^5}{5} + \mu_2''^2\frac{q_1'^7}{7}\right)\right\}, \quad (29)$$

$$E_{b2} = \frac{1}{2}\hbar\omega_1 N_2\left\{\mu_2'' - \frac{15}{4}\frac{(\mu_2'')^{7/2}}{\beta^2(\mu_2^0)^{5/2}} \times\left[\frac{8}{105} - \left(\frac{q_1'^3}{3} - \frac{2}{5}q_1'^5 + \frac{q_1'^7}{7}\right)\right]\right\}, \quad (30)$$

where μ_1'' and μ_2'' are the chemical potentials in configuration *b*, $q_1 = \sqrt{\mu_2''}\,q_1'$ is the long semiaxis of the boundary ellipsoid, and E_b is the energy of configuration *b*.

To find which configuration is stable, we must compare E_a and E_b. We first consider the limiting cases, where $n_2 = N_2/N_1 \ll 1$ and $n_1 = N_1/N_2 \ll 1$.

In the first case, where $n_2 \ll 1$, the approximate solution of Eqs. (20)–(22) is given by

$$q' = q_0\left[1 + \frac{1}{3}\left(1 - \frac{2}{5}\kappa\right)q_0^2 - \frac{5}{6}q_0^3\right], \quad (31)$$

$$\mu_1' = \mu_1^0(1 + q_0^3), \quad (32)$$

$$\mu_2' = \frac{\mu_1^0}{\kappa}[1 + (\kappa - 1)q_0^2 + q_0^3], \quad (33)$$

where

$$q_0 = \left(\frac{2n_2}{5\kappa}\right)^{1/3}. \quad (34)$$

It follows from Eqs. (26)–(28) that

$$q_1' = 1 - p_0 - \frac{5}{2}\left(\kappa - \frac{2}{3}\right)p_0^2 - \frac{25}{8}\left(\kappa^2 - \frac{14}{5}\kappa + \frac{224}{225}\right)p_0^3, \quad (35)$$

$$\mu_1'' = \mu_1^0(1 + 3\kappa^2 p_0^2), \quad (36)$$

$$\mu_2'' = \mu_1^0\left(1 - 2(\kappa - 1)p_0 + \frac{1}{3}(6\kappa^2 + 4\kappa - 1)p_0^2\right), \quad (37)$$

where

$$p_0 = \left(\frac{2n_2}{15\kappa^2}\right)^{1/2}. \quad (38)$$

Using Eqs. (31)–(38), we easily show that

$$\Delta E = E_a - E_b = \frac{1}{2}\hbar\omega_1 N_1\mu_1^0\frac{1 - \kappa}{\kappa}n_2. \quad (39)$$

This shows that $\Delta E < 0$ for $\kappa > 1$, and configuration *a* is therefore stable.

We now consider the case where $n_1 \ll 1$. An approximate solution for configuration *a* is given by

$$q' = 1 - x - \frac{5(3 - 2\kappa)}{6\kappa}x^2 - \frac{224\kappa^2 - 630\kappa + 225}{72\kappa^2}x^3, \quad (40)$$

$$\mu_1' = \mu_2^0\beta^{4/5}\left(1 + 2\frac{\kappa - 1}{\kappa}x - \frac{\kappa^2 - 4\kappa - 6}{3\kappa}x^2\right), \quad (41)$$

$$\mu_2' = \mu_2^0\beta^{4/5}\left(1 + \frac{3}{\kappa^2}x^2\right), \quad (42)$$

where

$$x = \left(\frac{2\kappa^2 n_1}{15}\right)^{1/2}. \quad (43)$$

For configuration *b*, the solution has the form

$$q_1' = y + \frac{\kappa - 2/5}{3\kappa}y^3 - \frac{5}{6}y^4, \quad (44)$$

$$\mu_1'' = \beta^{4/5}\mu_2^0 \times\left[\kappa - (\kappa - 1)y^2 + \kappa y^3 - \frac{2(\kappa - 1)(5\kappa - 2)}{15\kappa}y^4\right], \quad (45)$$

$$\mu_2'' = \beta^{4/5}\mu_2^0(1 + y^3), \quad (46)$$

where

$$y = \left(\frac{2\kappa n_1}{5}\right)^{1/3}. \quad (47)$$

The energy difference is

$$\Delta E = \frac{1}{2}\hbar\omega_1 N_2\mu_2^0 n_1(1 - \kappa). \quad (48)$$

We see from Eqs. (39) and (48) that configuration *a* has a lower energy if $\kappa = \sqrt{a_{11}m_2/(a_{22}m_1)} > 1$. For $m_1 = m_2$, this is consistent with the qualitative assertion and the experimental observation that it is energetically favorable for the atoms with the larger scattering length to form a lower density shell about the atoms with the smaller scattering length [14, 19].

To evaluate ΔE in general case, it is useful to first estimate the energy of the phase boundary that arises due to the gradient terms omitted in the TFA. The sur-

face tension $\sigma = E_s/S$, where E_s is the surface energy and S is the interface area, can be written as [20, 21]

$$\sigma = \frac{\hbar\omega_1}{2\sqrt{2}a_\perp^2}\left(\frac{a_{12}}{\sqrt{a_{11}a_{22}}} - 1\right)^{1/2}(u_1u_2N_1N_2)^{1/4} \times |\psi_1'||\psi_2'|(N_1|\psi_1'|^2 + N_2|\psi_2'|^2)^{1/2}. \quad (49)$$

Taking into account that the surface area of the ellipsoid with the semiaxis $a_\perp q$ is given by

$$S = 2\pi a_\perp^2 q^2\left(1 + \frac{1}{\lambda\sqrt{\lambda^2-1}}\log\frac{1}{\lambda - \sqrt{\lambda^2-1}}\right), \quad (50)$$

we can estimate the contribution of the surface energy $E_s = \sigma S$ to the total energy of each configuration. To be specific, we use the parameters corresponding to the experiments on ^{87}Rb atoms. In this case, $m_1 = m_2$, $a_\perp = 2.4 \times 10^{-4}$ cm, and $N = N_1 + N_2 = 0.5 \times 10^6$ atoms.

In Fig. 1a, we show the energies of configurations *a* and *b* (including the surface energy) $E_a/(\hbar\omega_1 N)$ (solid line) and $E_b/(\hbar\omega_1 N)$ (dashed line) as functions of $\log n_2$. One can see that E_a is always lower than E_b. Figure 1b represents the difference $\Delta E = (E_a - E_b)/(\hbar\omega_1 N)$. The behavior of ΔE for small and large values of n_2 is well described by Eqs. (39) and (48). Figure 1c illustrates the behavior of the surface energy as a function of n_2. We note that the surface energy is much smaller than the interaction energy because the scattering lengths a_{ij} hate very close values (see Eq. (49)).

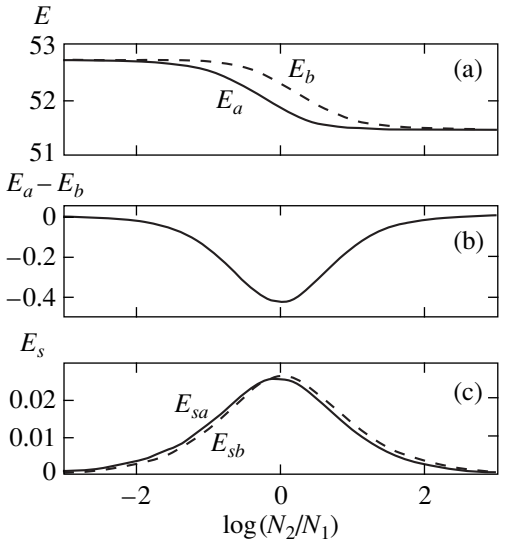


Fig. 1. (a) The total energies of configurations *a* and *b*, $E_a/(\hbar\omega_1 N)$ (solid line) and $E_b/(\hbar\omega_1 N)$ (dashed line), as functions of $\log n_2$. (b) The difference $\Delta E = (E_a - E_b)/(\hbar\omega_1 N)$ as a function of $\log n_2$. (c) The surface energies as functions of $\log n_2$. Solid line corresponds to surface energy E_{sa} for the configuration *a*, dashed line to the surface energy E_{sb} for configuration *b*.

3. VORTEX STATES IN THE ROTATING CONDENSATE

We now consider a trap rotating with the frequency along Ω the *z*-axis.

For a vortex excitation with the angular momentum $\hbar l_j$, the condensate wave function is given by

$$\psi_{l_j}(\mathbf{r}) = |\psi_{l_j}(\mathbf{r})|e^{il_j\phi}. \quad (51)$$

In the rotating frame, the energy functional of the system is

$$E_{\rm rot}(l_1, l_2) = E(\psi_{l_1}, \psi_{l_2}) + \int d^3r(\psi_{l_1}^* + \psi_{l_2}^*)i\hbar\Omega\partial_\phi(\psi_{l_1} + \psi_{l_2}). \quad (52)$$

With the wave function for the vortex excitation in Eq. (51) inserted in Eq. (52), the effective confinement potential for the bosons becomes $l_1^2\hbar^2/2m_1\rho^2 + l_2^2\hbar^2/2m_2\rho^2 + V_1 + V_2$, where $V_i = m_i\omega_i(\rho^2 + \lambda^2z^2)/2$ and $\rho^2 = x^2 + y^2$. In the nonoverlapping regions, in the TFA the density of the vortex state is therefore given by

$$|\psi_1'(\mathbf{r}')|^2 = \frac{1}{u_1}\left[\mu_1'(l_1) - (\rho'^2 + \lambda^2z'^2) - \frac{l_1^2}{\rho'^2}\right], \quad (53)$$

$$|\psi_2'(\mathbf{r}')|^2 = \frac{1}{u_2\beta^2}\left[\mu_2'(l_2) - (\rho'^2 + \lambda^2z'^2) - \frac{\beta^2l_2^2}{\rho'^2}\right]. \quad (54)$$

The important new qualitative feature of a vortex in the TFA is the appearance of a small hole of the radius ξ_i such that $\xi_i^2 \propto l_i^2/\mu_i(l_i)$, with the remainder of the condensate density essentially unchanged. The fractional change in the chemical potentials caused by the vortex $[\mu_i'(l_i) - \mu_i']/\mu_i'$ can be shown to be small [6, 9]: on the order of $1/N^{4/5}$. In calculating physical quantities involving the condensate density, it is sufficient to retain the nonvortex density and simply cut off any divergent radial integrals at the appropriate core sizes $\xi_1^2 = l_1^2/\mu_1'$ or $\xi_2^2 = \beta^2l_2^2/\mu_2'$. We note that calculating the vortex properties using the unperturbed density corresponds to the hydrodynamic limit.

For the phase segregated condensate, we find from Eqs. (51), (52) and (15), (16) that the energy change

Table

Ω/ω_1	n_2	l_1	l_2	$E_{\rm tot}$, 10^7
0.1	1.0	1	0	1.296863
0.1	10.0	2	0	1.287497
0.15	1.0	2	0	1.295151
0.2	1.0	3	1	1.292344

due to the presence of the vortices, $\Delta E = E_{\rm rot}(l_1, l_2) - E_{\rm rot}(0, 0)$, has the form

$$\begin{aligned}\Delta E &= \Delta E_{N_1} + \Delta E_{N_2} \\ &= \frac{1}{2}\hbar\omega_1 N_1 \int d_3 r' \left(\frac{l_1^2}{\rho'^2}|\psi_1'|^2 - \frac{2\Omega l_1}{\omega_1}|\psi_1'|^2 \right) \\ &+ \frac{1}{2}\hbar\omega_1 N_2 \int d_3 r' \left(\frac{l_2^2\beta^2}{\rho'^2}|\psi_2'|^2 - \frac{2\Omega l_2}{\omega_1}|\psi_2'|^2 \right).\end{aligned} \tag{55}$$

In the hydrodynamic limit, ψ_i' is given by Eqs. (12) and (13).

We now consider the stable configuration *a*. In the hydrodynamic limit, the location of the phase boundary is given by Eq. (19). It follows from Eq. (55) that

$$\begin{aligned}\frac{\Delta E_{N_1}}{\frac{1}{2}\hbar\omega_1 N_1} &= \frac{5 l_1^2 (\mu_1')^{3/2}}{(\mu_1^0)^{5/2}} \left\{ \left[\ln\frac{2\mu_1'}{l_1} - \frac{4}{3} \right] \right. \\ &\left. - \frac{3}{2} q' \left[\left(1 - \frac{1}{3} q'^2 \right) \ln\frac{2\mu_1' q'}{l_1} - \left(1 - \frac{q'^2}{9} \right) \right] \right\} - \frac{2\Omega l_1}{\omega_1},\end{aligned} \tag{56}$$

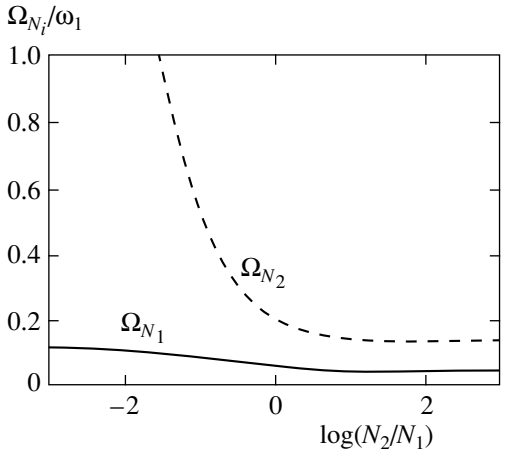


Fig. 2. Critical angular velocities Ω_{N_1}/ω_1 and Ω_{N_2}/ω_1 as functions of $\log n_2$ for configuration *a*. Solid line corresponds to the outer condensate 1, dashed line to the inner condensate 2.

$$\begin{aligned}\frac{\Delta E_{N_2}}{\frac{1}{2}\hbar\omega_1 N_2} &= \frac{15 l_2^2 (\mu_1')^{1/2} q'}{2(\mu_2^0)^{5/2}} \left[\left(\mu_2' - \frac{1}{3}\mu_1' q'^2 \right) \right. \\ &\left. \times \ln\frac{2\sqrt{\mu_1'\mu_2'}q'}{l_2\beta} - \left(\mu_2' - \frac{\mu_1' q'^2}{9} \right) \right] - \frac{2\Omega l_2}{\omega_1}.\end{aligned} \tag{57}$$

The critical angular velocities required to produce the vortex states in each condensate can be determined from the conditions $\Delta E_{N_1} < 0$ and $\Delta E_{N_2} < 0$ and are given by

$$\begin{aligned}\frac{\Omega_{N_1}}{\omega_1} &= \frac{5 l_1 (\mu_1')^{3/2}}{2(\mu_1^0)^{5/2}} \left\{ \left(\ln\frac{2\mu_1'}{l_1} - \frac{4}{3} \right) \right. \\ &\left. - \frac{3}{2} q' \left[\left(1 - \frac{q'^2}{3} \right) \ln\frac{2\mu_1' q'}{l_1} - \left(1 - \frac{q'^2}{9} \right) \right] \right\},\end{aligned} \tag{58}$$

$$\begin{aligned}\frac{\Omega_{N_2}}{\omega_1} &= \frac{15 l_2 (\mu_1')^{1/2} q'}{2(\mu_2^0)^{5/2}} \left[\left(\mu_2' - \frac{1}{3}\mu_1' q'^2 \right) \right. \\ &\left. \times \ln\frac{2\sqrt{\mu_1'\mu_2'}q'}{l_2\beta} - \left(\mu_2' - \frac{\mu_1' q'^2}{9} \right) \right].\end{aligned} \tag{59}$$

To find the asymptotic behavior of the critical angular velocities (58) and (59) for $N_2 \ll N_1$ and $N_1 \ll N_2$, we use approximate solutions (31)–(34) and, thus, obtain

$$\frac{\Omega_{N_1}}{\omega_1} = \frac{5 l_1}{2\mu_1^0}\left(\ln\frac{2\mu_1^0}{l_1} - \frac{4}{3} \right) - \frac{15 l_1}{4\mu_1^0}\left(\ln\frac{2\mu_1^0 q_0}{l_1} - 1 \right) q_0, \tag{60}$$

$$\begin{aligned}\frac{\Omega_{N_2}}{\omega_1} &= \frac{3 l_2 \beta^2}{2\mu_1^0 q_0^2}\left[\ln\left(\frac{2\mu_1^0 q_0}{\kappa^{1/2} l_2 \beta} \right) - 1 \right] \\ &- \frac{l_2\beta^2}{\mu_1^0}\left[\ln\left(\frac{2\mu_1^0 q_0}{\kappa^{1/2} l_2 \beta} \right)\left(1 - \frac{4}{5}\kappa \right) + \frac{7}{12}\kappa - \frac{3}{4} \right].\end{aligned} \tag{61}$$

We see from Eqs. (60) and (61) that, as $n_2 \longrightarrow 0$, the critical angular velocity of the external condensate Ω_{N_1} tends to that of the pure condensate with the scattering length a_{11} [see Eq. (26) in [9]]. The critical angular velocity of the inner condensate Ω_{N_2} tends to the infinity as $n_2 \longrightarrow 0$. However, this consideration cannot be applied to rapidly rotating gases with Ω comparable to ω_1, where the form of the condensate depends on Ω [13].

In the opposite limit $n_1 = N_1/N_2 \ll 1$, the critical angular velocities can be written as

$$\frac{\Omega_{N_1}}{\omega_1} = \frac{l_1}{2\mu_2^0\beta^{4/5}}\ln\frac{2\mu_2^0\beta^{4/5}}{l_1} + \frac{l_1}{3\kappa\mu_2^0\beta^{4/5}}\left[(3-\kappa)\ln\frac{2\mu_2^0\beta^{4/5}}{l_1} + 3 - 2\kappa\right]x, \quad (62)$$

$$\frac{\Omega_{N_2}}{\omega_1} = \frac{5l_2\beta^{6/5}}{2\mu_2^0}\left(\ln\frac{2\mu_2^0}{l_2\beta^{1/5}} - \frac{4}{3}\right) + \frac{15l_2\beta^{6/5}}{2\mu_2^0\kappa^2}\left(\ln\frac{2\mu_2^0}{l_2\beta^{1/5}} - 1\right)x^2. \quad (63)$$

In deriving Eqs. (62) and (63), we used the approximate solutions (40)–(43). We note that, as $n_1 \longrightarrow 0$, the critical angular velocity (63) has the same form as the critical velocity for the pure condensate with the scattering length a_{22} and with the chemical potential $\mu^0\beta^{4/5}$.

Figure 2 shows the critical angular velocities for the external (Ω_{N_1}) and the inner (Ω_{N_2}) condensates as functions of $n_2 = N_2/N_1$ for $l_1 = l_2 = 1$.

Using Eqs. (52) and (55), one can find the vortex configurations corresponding to the energy minimum for a given angular velocity Ω/ω_1. In the table, we represent angular momenta of condensates that correspond to the minimum of the total energy $E_{\text{tot}}/\hbar\omega_1$ of the system for different values of the angular velocity and n_2. In the calculation, we use the parameters for the ^{87}Rb. E_{tot} is calculated as the sum of E (Eq. (14)), the surface energy E_s [Eqs. (49) and (50)], ΔE_{N_1} [Eq. (56)), and ΔE_{N_2} (Eq. (57)].

4. CONCLUSIONS

We have shown that, in the ^{87}Rb condensate, the atoms in state 1 form a shell about the atoms in state 2, the critical angular velocity for each state being drastically dependent on the relative concentrations. The critical angular velocity of the outer component is less than the angular velocity of the inner one. When the ratio of the number of state 2 atoms to the number of state 1 atoms is sufficiently small, the critical angular velocity of the inner state becomes very large; in the framework of the hydrodynamic approximation, it is larger than the oscillator frequency characterizing the confining potential and decreases smoothly with increasing the number of atoms in state 2. From the table, one can see that atoms in state 1 can form a rotating ring around the resting core of the atoms in state 2.

ACKNOWLEDGMENTS

This work was supported in part by NATO grant no.PST.CLG.976038. V.N. Ryzhov and E.E. Tareyeva acknowledge the financial support from the Russian Science Foundation through the grant no. 98-02-16805. S.T. Chui was partly supported by NASA under grant no. CRG8-1427.

REFERENCES

1. M. N. Anderson, J. R. Ensher, M. R. Matthews, *et al.*, Science **269**, 198 (1995).
2. L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **40**, 646 (1961) [Sov. Phys. JETP **13**, 451 (1961)]; E. P. Gross, Nuovo Cimento **20**, 454 (1961); J. Math. Phys. **4**, 195 (1963).
3. M. R. Matthews, B. P. Anderson, P. C. Haljan, *et al.*, Phys. Rev. Lett. **83**, 2498 (1999).
4. J. E. Williams and M. J. Holland, Nature **401**, 568 (1999).
5. K. W. Madison, F. Chevy, W. Wohlleben, *et al.*, Phys. Rev. Lett. **84**, 806 (2000).
6. A. L. Fetter, cond-mat/9808070; cond-mat/9811366.
7. G. Baym and C. J. Pethick, Phys. Rev. Lett. **76**, 6 (1996).
8. F. Dalfovo and S. Stringari, Phys. Rev. A **53**, 2477 (1996).
9. S. Sinha, Phys. Rev. A **55**, 4325 (1997).
10. D. S. Rokhsar, Phys. Rev. Lett. **79**, 2164 (1997).
11. A. A. Svidzinsky and A. L. Fetter, cond-mat/9811348.
12. A. L. Fetter and D. Rokhsar, Phys. Rev. A **57**, 1191 (1998).
13. D. A. Butts and D. S. Rokhsar, Nature **397**, 327 (1999).
14. D. S. Hall, M. R. Matthews, J. R. Ensher, *et al.*, cond-mat/9804138; E. A. Cornel, D. S. Hall, and M. R. Matthews, cond-mat/9808105.
15. J. Stenger, S. Inouye, D. M. Stamper-Kurn, *et al.*, Nature **396**, 345 (1998).
16. M. R. Matthews, D. S. Hall, D. S. Jin, *et al.*, cond-mat/9803310.
17. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
18. L. P. Pitaevskiĭ, Usp. Fiz. Nauk **168**, 641 (1998) [Phys. Usp. **41**, 569 (1998)].
19. H. Pu and N. P. Bigelow, Phys. Rev. Lett. **80**, 1130 (1998).
20. E. Timmermans, Phys. Rev. Lett. **81**, 5718 (1998).
21. P. Ao and S.-T. Chui, Phys. Rev. A **58**, 4836 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1190–1197.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1374–1383.

FLUIDS

Formation of a Grating of Submicron Nematic Layers by Photopolymerization of Nematic-Containing Mixtures

R. Caputo[a], A. V. Sukhov[b, *], C. Umeton[a], and R. F. Ushakov[b]

[a]*Dipartamento di Fisica, Università della Calabria, Arcavacata di Rende, CS, 87036 Italia*
[b]*Institute of Problems of Mechanics, Russian Academy of Sciences, Moscow, 117526 Russia*
**e-mail: sukhov@ipmnet.ru*

Received June 26, 2000

Abstract—A submicron, spatially periodic, structure consisting of a sequence of oriented layers of a nematic liquid crystal (NLC), separated by isotropic polymeric walls, was obtained experimentally. This was accomplished by polymerization induced by the interference pattern of UV laser radiation in a NLC-containing prepolymer mixture. It was established that such a structure occurs when phase separation and nematic ordering are prevented during the polymerization process. These structures are the diffraction elements, whose efficiency is much greater than that of a standard grating of polymer-dispersed liquid crystals [1–4] which is obtained in the same initial mixture. Specifically, a diffraction efficiency of 60–70% was obtained for structures with the period $\Lambda = 0.2$ μm, even in mixtures where a grating with $\Lambda < 6$ μm cannot be obtained at all by the standard technique.

1. INTRODUCTION

Starting with the works of Sutherland *et al*. at the beginning of the 1990s [1, 2], the use of polymer-dispersed liquid crystals (PDLC) for electrically controllable diffraction and holographic components became just about the main topic of interest in electro-optic effects based on PDLC. The reason for such interest is obvious (see, e.g., [3–5])—it is due to the possibility of obtaining inexpensive elements for commercial devices.

To obtain high diffraction efficiency and high optical quality for gratings and holograms it is necessary to obtain lines with quite uniform morphology. Then, optical nonuniformities (due to nematic drops) with the same spatial size as the wavelength of the radiation used for diffraction can be avoided. There are two obvious ways to accomplish this.

The first one, used in [1–4], consists in obtaining drops of a liquid crystal (LC) which are much smaller in size than the wavelength mentioned above. This is the approach employed in the conventional methods of obtaining polymer-dispersed liquid crystals using UV radiation with spatially uniform intensity. In this case the only internal spatial scale in the problem is the minimum size of a thermodynamically stable drop and, possibly, the size of the radiation intensity nonuniformities arising after drop formation starts.

The second method is based on the possibility of obtaining uniform stripes of polymer and nematic rather than, say, PDLC stripes with nematic content varying from the maximum to the minimum of the intensity distribution, or if polymer stripes alternating with PDLC stripes. Of course, this situation is impossible in the traditional method of obtaining PDLC using spatially uniform radiation. Indeed, in this case phase separation is a threshold (with respect to the radiation intensity) stochastic process, starting in a sample with initially uniform concentrations of a nematic and a monomer and accompanied by efficient mass transfer. A process of this kind leads to a quite random morphology of the sample. However, when the initial mixture is irradiated with a spatially periodic intensity distribution (an interference pattern) and therefore a periodically nonuniform degree of polymerization it is logical to infer that mass transfer starts long before phase separation occurs (and it could be completed before the isotropic-liquid–nematic (*I–N*) phase transition starts). Thus, if such a scenario is possible, then it will also be possible to obtain not drops but rather quite uniform regions of nematic and polymeric isotropic phases, provided that during photoinduced polymerization phase separation and a transition are prevented by means of an additional external action. Then no random processes will occur, and one can validly expect to obtain quite regular modulation of the nematic concentration within a period of the grating.

On the other hand if the indicated preventative action is suppressed after polymerization is completed, then the rigid polymeric structure will make it impossible for random deformations of the polymer stripes to occur during a *I–N* transition, which can occur in NLC rich regions of the sample.

The successful results of such an approach are demonstrated in the present paper.

1063-7761/00/9106-1190$20.00

2. EXPERIMENTAL APPARATUS

Three types of initial systems were used. They consisted of a free polymeric mixture (monomer–photoinitiator) and nematic components with high solubility in the prepolymer.

The first type consisted of the prepolymer composition SAM-114 (Merck), diluted with a 5CB nematic. This is the conventional acrylate prepolymer system. Its components are highly mutually soluble, the initial weight concentrations C_N of the NLC from 0 to 95% were thermodynamically stable and formed a spatially uniform isotropic mixture.

The second prepolymer system employed was a SAM-114 composition, diluted by the nematic BL036 (Merck), which possesses a nematic–isotropic (*N–I*) transition temperature above 90°C and virtually the same room-temperature optical properties as the 5CB nematic. The maximum thermodynamically stable concentration of this nematic in the prepolymer is $C_N = 55\%$.

The third mixture was a standard prepolymer mixture NOA-65, diluted with the nematic 5CB. This mixture also exhibits mutual solubility in the range 0–100% of the NLC concentration with the resulting mixture being up to 92% isotropic.

The samples consisted of thin (11 μm thick) plane-parallel layers between two glass plates with standard transparent electrodes (ITO).

The optical arrangement is shown in Fig. 1. A transverse–unimodal beam from an argon laser, operating at the wavelength $\lambda_B = 0.3548$ μm (Coherent Innova 90C) with power monitored with a standard measuring device in a range 3–100 mW, was expanded with a telescope BE to a diameter of approximately 25 mm. The beam edges were cut off with an iris diaphragm, so that the intensity in the remaining aperture (2–5 mm in diameter) was uniform to within 4–5%. Next, a beam splitter BS split the beam into two beams with approximately the same intensity ($I_1/I_2 = 0.95 \pm 0.02$). These beams interesected one another on the entrance plane of the sample, creating an interference pattern whose spatial period varied in range $\Lambda = 6.3$–0.2 μm. The polarization was of *s* type, i.e., the **E** vector of the radiation was directed along the lines of the grating.

The diffraction efficiency of the grating was measured with a slightly focused (beam diameter about 1 mm inside the sample), approximately 1 mW, probe beam, which was also *s* polarized, from an He–Ne laser ($\Lambda_R = 0.63$ μm). The power of the transmitted probe beam (zeroth order of diffraction) and in the first diffraction order were recorded with $PD_{T,D}$ photodiodes, respectively. It should be noted that everywhere below the data for the diffraction efficiency η are scaled to the power of the transmitted beam before polymerization onset.

W note that in short-wavelength part of the abovementioned range of grating periods the wave mismatches

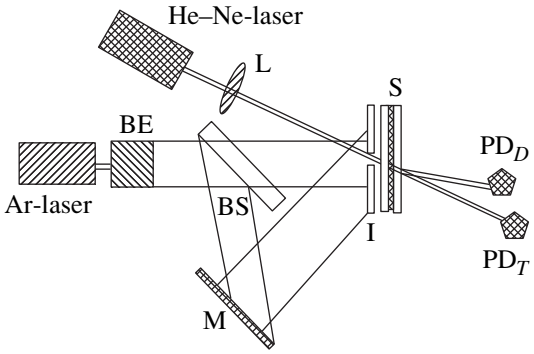


Fig. 1. Optical arrangement for writing and investigating gratings (explanation in text).

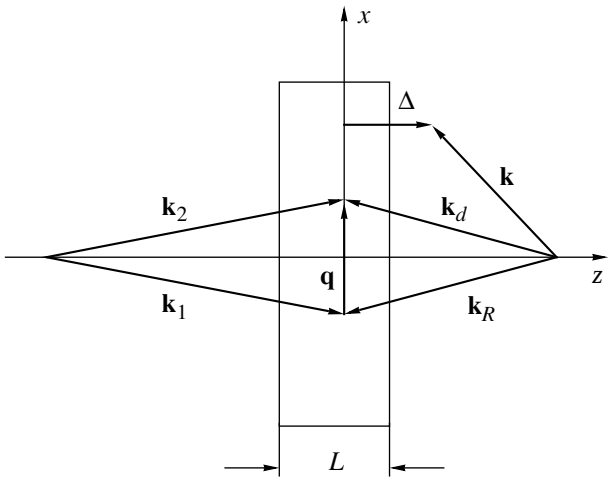


Fig. 2. The diffraction geometry employed: $\mathbf{k}_{1,2}$ are wave vectors of the writing beams, **q** is wave vector of the grating, $\mathbf{k}_{R,d}$ are wave vectors of the probe and the first order of diffraction, **k** is away factor of the second order of diffraction, Δ is wave mismatch of the latter, *L* is sample thickness, and *x* and *z* are coordinate axes.

Δ*L* for the first order of diffraction can reach values of about 3–4 rad, thereby suppressing the required to diffraction maximum. The only possibility of avoiding these mismatches for an individual diffraction order is to make the appropriate choice of the required angle of incidence of the probe beam (see Fig. 2). The angle of incidence of the probe been $\mathbf{k}_R$ is chosen so that the bisector of the angle between the beams $\mathbf{k}_{1,2}$ producing the pattern is also the bisector of the angle between the beams $\mathbf{k}_R$ and $\mathbf{k}_d$, the latter of which is some order of its diffraction (in our case—the first order). It is easy to see that for this, and only for this, maximum the wave mismatch $\Delta L = 0$, while for all other possible maxima $\Delta L \neq 0$ and therefore they are suppressed to a greater or lesser degree ($\propto \sin^2(0.5\Delta L)/(0.5\Delta L)^2$). Therefore, in all experiments described below the probe been was adjusted appropriately so as to correspond to the geometry shown in Fig. 2. Therefore, none of the diffraction efficiency data contain errors associated with wave mismatches.

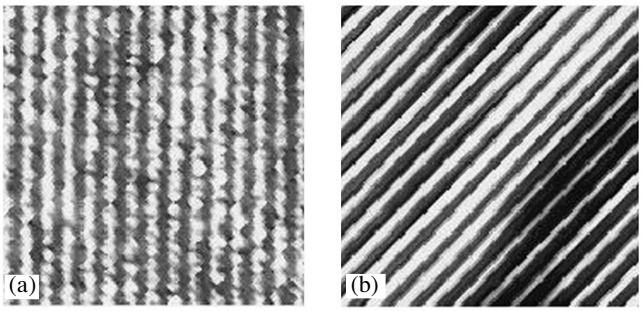


Fig. 3. Photomicrographs of gratings obtained in a sample with C_N = 59.9% with intensity 5 mW/cm^2: (a) T = 22°C and (b) T = 55°C. Here and below the vertical arrangement of the lines in the photomicrographs corresponds to the stochastic orientation of the axes of the drops, and an arrangement inclined by 45° corresponds to a uniform orientation along the lines.

It should also be noted that for $\Lambda < 0.317$ μm it is impossible to perform diffraction efficiency measurements of this kind, since a diffraction maximum is a nonuniform wave. Consequently, for gratings with a shorter spacing the data were obtained when one beam, forming the interference pattern, was used as the probe after the polymerization process was completed.

To avoid ambiguity associated with such a difference of the wavelengths of the probe beam for large and small Λ in the interpretation of the data, the diffraction efficiency of the gratings was compared for both wavelengths (λ_R and λ_B) in range 0.5 μm > Λ > 0.35 μm. This comparison revealed that in entire indicated range both the diffraction efficiencies differ only by a constant factor associated with the difference of the optical thicknesses nL/λ of the sample for these wavelengths. For convenience in making comparisons, the data "renormalized" to the wavelength λ_R = 0.63 μm are presented everywhere below.

The morphology of the samples obtained was monitored with a standard polarized microscope with resolution of the order of 0.5 μm and recorded using a standard CCD camera.

Naturally, the individual lines of a grating were not resolved for short periods. In this case the microscope was used only for monitoring the orientation of the anisotropic medium obtained.

3. EXPERIMENTAL RESULTS

Our approach to the problem of searching for an external action that prevents phase separation and an *I–N* transition during the polymerization process consisted of the following.

In the first place, as already mentioned above, we used a monomer and a nematic which dissolve well in one another. Moreover, the nematic concentrations used in the experiments were only slightly higher than the threshold for phase separation when the mixture was irradiated with a uniform beam with the same intensity range. Thus, the amount of the nematic in the experimental sample corresponded to almost complete solubility not only in the initial monomer but also in the corresponding polymer.

In second place, before irradiation with an interference pattern the sample was heated (with a standard liquid optical thermostat) to temperatures above the *N–I* transition temperature of a pure nematic component. Such heating made it possible to prevent an *I–N* transition during the polymerization process, simultaneously increasing the above-indicated solubility and thereby preventing phase separation during the polymerization process.

Under such conditions mass transfer can be due only to the spatial modulation of the degree of polymerization within a period of the interference pattern. The corresponding mass transfer itself consists of diffusion of the nematic (and also, possibly, the monomer and low-molecular reaction products) during and after the completion of the photoinduced polymerization. Processes of this kind have been observed and investigated in a recent work [6] for the first of the above-enumerated types of initial mixtures. Specifically, it was established that processes of this kind can result in very high degrees of spatial modulation of the nematic concentration (up to 90–95%). This modulation, even in the absence of an *I–N* transition engenders a diffraction grating (called in [6] a "pregrating"), which has a quite low diffraction efficiency (1–2%) but is of high-quality (there is no spontaneous scattering characteristic of PDLC).

3.1. SAM-114–5CB Mixtures

The range of nematic concentrations C_N = 50–70% was investigated for mixtures of this type. Samples of thickness L = 11 μm were irradiated for 10 min, which is sufficient for polymerization to be completed in this system (see [6]). The grating morphology obtained for Λ = 6.3 μm is presented in Fig. 3. Both gratings presented were obtained in the same sample with C_N = 60% using the same radiation intensity I_B = 5 mW/cm^2. The first one (Fig. 3a) was irradiated at room temperature, so that phase separation started during irradiation. It is evident that this is an ordinary PDLC grating (see, for example, [1–3]) with stochastically oriented nematic drops. At the same time the second grating (Fig. 3b), irradiated at temperature 55°C and slowly (in 40–50 min) cooled to room temperature, contains uniformly oriented stripes of an extended nematic phase with very sharp, undistorted changes.

The typical temperature dependences of the diffraction efficiency of such gratings are presented in Fig. 4. It follows from the figure that, as expected, the dependences decay rapidly near the point of the *N–I* transition, which occurred for mixtures of this type near 25°C (the *N–I* transition of a pure nematic 5CB starts at

37°C, but the fact that the substance in the nematic stripes of the grating is strongly diluted by nonmesogenic reaction products must be taken into account). It is evident from the figure that the diffraction efficiency reaches 25%, while for the grating obtained at room temperature the diffraction efficiency for mixtures of this type does not exceed 7%. Such a sharp increase in the diffraction efficiency is explained primarily by the absence of losses due to stochastic scattering in PDLC.

We did not perform systematic measurements of the dependences of the diffraction efficiency on the radiation intensity and the period of the pattern, since the reproducibility of such measurements was poor because the temperature of the *N–I* transition was close to room temperature. On the other hand there is also room for improvement of the longevity of the gratings obtained in these mixtures. Although in some samples the grating morphology remained unchanged after one year of storage, in most samples the nematic ordering at room temperature vanished in hours—several days, i.e., the samples transformed into a "pregrating" state [6]. Such a loss of ordering is due to slow diffusion of the nematic from the stripes into the surrounding exposed region of the solution and back diffusion of the monomer from the solution.

3.2. SAM-114-BL036 Mixtures

Samples with a thickness of 11 μm were irradiated at 65°C for 10 min, which, as experiments showed, was adequate for polymerization to be completed. This means that irradiation for a longer time would not change the properties of the gratings obtained.

The results obtained to reduce, qualitatively, to the following. In the first place, it must be underscored that SAM-BL036 mixtures were found to be more than unsatisfactory from the standpoint of the obtaining gratings by the conventional methods [1–3] (i.e., at room temperature). Specifically, even for optimal irradiation intensities (see below) it was impossible to obtain diffraction efficiencies above $\eta_1 = 2$–3%. Even such values could be obtained only in a very narrow range of nematic concentrations $C_N = 33$–35%; for all other concentrations the diffraction efficiency did not exceed the "pregrating" level (<1%). This was due to the extremely high losses of the probe beam to stochastic scattering. Even the above-indicated values of the diffraction efficiency could be reached only for relatively large grating periods $\Lambda = 6.3$ μm. For smaller values of Λ the diffraction efficiency was below the threshold of experimental detection, which is 0.2%.

This is due to the morphology of the gratings obtained (see Fig. 5a). The grating lines are almost overlapped by individual drops of the nematic, which typically are at least 4 μm in size. Probably, the surface tension at the nematic–polymer interface for mixtures of this type is very large and the indicated drop size is the minimum thermodynamically admissible value.

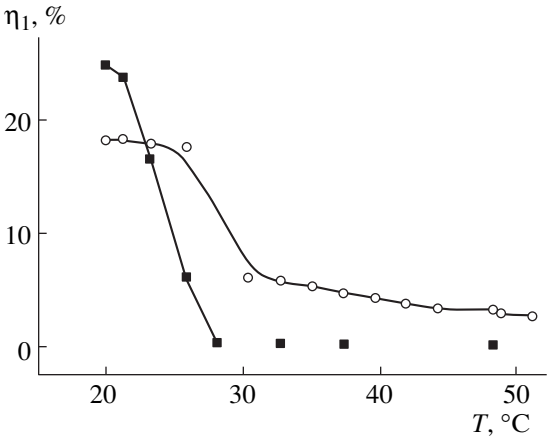


Fig. 4. Temperature dependences of the diffraction efficiency of gratings recorded at $T = 55$°C; $C_N = 59.9\%$; $I =$ (○) 15 and (■) 10 mW/cm^2.

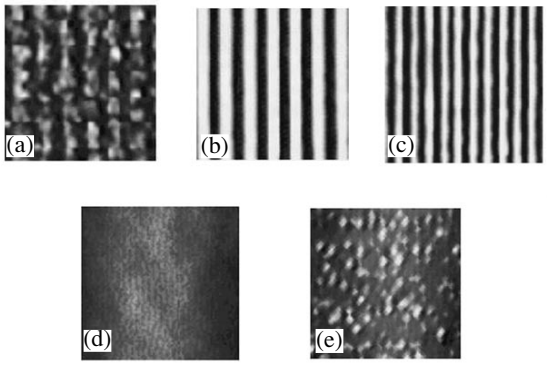


Fig. 5. Photomicrographs of gratings: (a) $\Lambda = 6.3$ μm, $T =$ 22°C; (b–d) $\Lambda = 6.3$, 4.8, and 1.44 μm, respectively, $T =$ 65°C; and (e) $\Lambda = 1.44$ μm, the phase separation threshold is exceeded; (d, e) the microscope does not resolve individual lines.

The results obtained using the above-described heat cycle differ radically from the results presented. On the whole it should be noted that diffraction efficiencies exceeding 20% were observed for any grating periods in the range $\Lambda = 6.3$–0.22 μm. Moreover, the grating morphologies presented in Figs. 5b–5d exhibit, just as for mixtures of the first type, a sequence of layers of an oriented nematic phase with sharp straight boundaries separated by an isotropic polymer. There are no drop defects, and consequently there is no stochastic scattering of the probe wave. Of course, for a more accurate proof of the nematic nature of the layers obtained it would be desirable to perform electron-microscopic investigations. However, an indirect confirmation of the nematic nature is that extended disclinations, characteristic of nematics, are sometimes present in the anisotropic stripes observed (see Fig. 6).

We shall now present the quantitative results. We shall confine ourselves to the case of the "best" (for

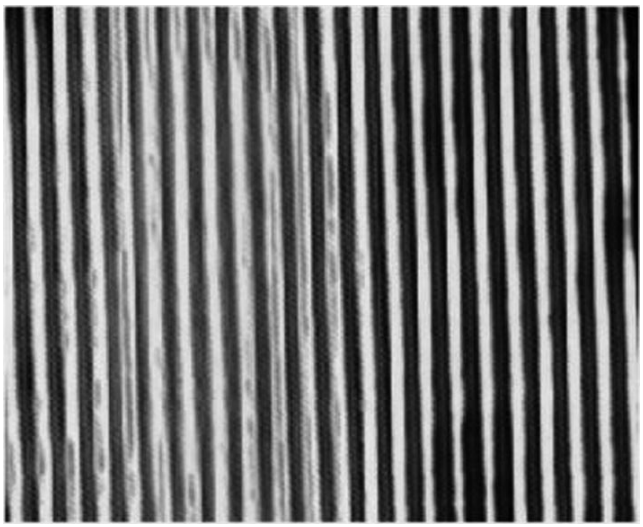


Fig. 6. Disclinations in nematic layers.

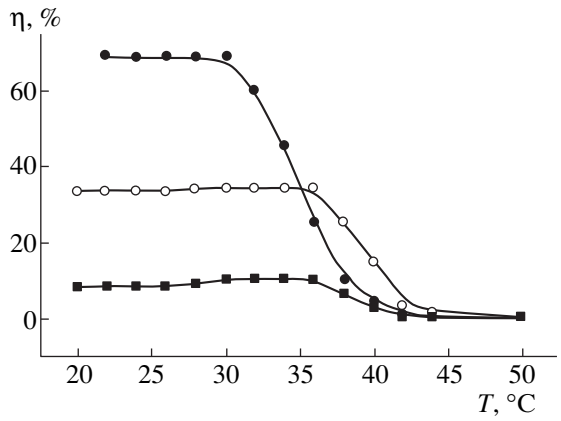


Fig. 7. Temperature dependences of the diffraction efficiency of a grating for SAM-BL036 mixtures; $\Lambda = 1.8$ μm. Radiation power: (■) 5 (○) , 30, and (●) 15 mW.

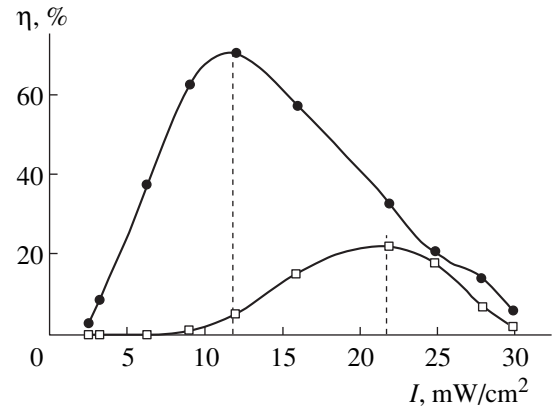


Fig. 8. Diffraction efficiency versus the radiation intensity for the mixture SAM-BL036; $\Lambda =$ (□) 0.31 and (●) 8 μm.

obtaining high diffraction efficiency) concentration of the nematic in the initial mixture, $C_N = 32\%$. For other concentrations in the range $C_N = 25$–40% the to diffraction efficiency was somewhat lower. This can be easily explained qualitatively. Specifically, for lower concentrations of the nematic, the width of the nematic stripes obtained was already equal to half the grating period, while for higher concentrations the stripes were wider than half the period, and in the limit of very high concentrations the stripes even overlapped. The concentration $C_N = 32\%$ corresponded experimentally to stripes whose width was equal to half the period, which is optimal for obtaining high diffraction efficiency.

It should be noted that, in contrast to mixtures of the first type, in the present experiments the irradiation temperature (65°C) was lower than the *N–I* transition temperature for a pure nematic (about 100°C). The problem is that irradiation at 100°C did not give the result: i.e., no gratings or subgratings were obtained. This is probably due to melting of the polymer at such temperatures. In this case the polymer chains acquire mobility and they is no reason for spatial modulation of the concentration (see [6]).

Thus, phase separation and the *I–N* transition in the polymerization process are possible in principle. This was observed experimentally above a certain critical value of the radiation intensity $I_{th}(\Lambda)$, which depended strongly on the period of grating. The texture of the grating was partially damaged by drop defects as well as by scattering isotropic inclusions present in the polymer layers (see Fig. 5e). The temperature dependences of the to diffraction efficiency of the gratings obtained are presented in Fig. 7 for the period $\Lambda = 1.8$ μm. The qualitative form of the dependences for other periods investigated is identical.

The radiation intensity dependences of the diffraction efficiency are presented in Fig. 8. It is evident that for each Λ there exists an optimal value of the radiation intensity, which depends on Λ. The dependences of the optimal intensity and of the corresponding value of the diffraction efficiency on the grating period are presented in Figs. 9a and 9b, respectively. It is evident that the diffraction efficiency reaches 70% even for short periods $\Lambda = 0.22$ μm, which correspond to Bragg reflection of green light at normal incidence; the diffraction efficiency is at least 20%.

The optimal intensity increases sharply at the grating period decreases. We also note that there exists a critical value of the latter, in this case $\Lambda = 0.19$ μm, below which it is impossible to obtain a grating even with the intensity exceeding the optimal value by a factor of 20 for $\Lambda = 0.22$ μm.

The longevity of the gratings obtained was quite poor. As a rule, a grating degraded after 40–50 h of storage at room temperature. However, the degradation processes themselves in this case are quite curious and have nothing in common with the degradation occurring in SAM-5CB mixtures. Microscopic observation

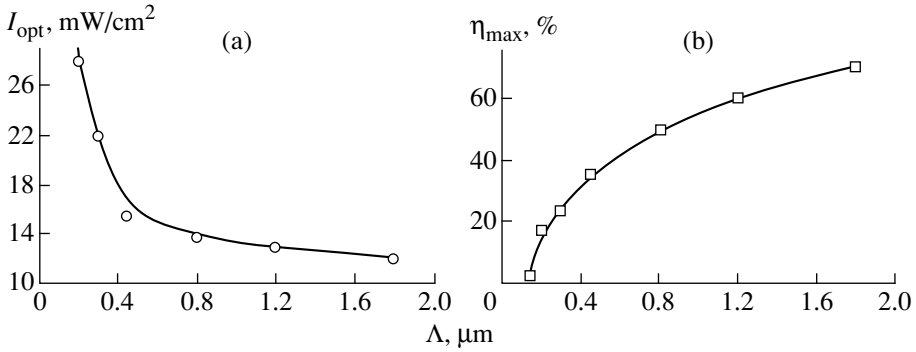


Fig. 9. Optimal writing intensity the (a) and maximum attainable diffraction efficiency (b) versus the grating period for the mixture SAM-BL036.

reveals formation and growth of nematic "lakes," localized at the nematic–composite interfaces.

On the other hand if an exposed sample was held at the exposure temperature for more than one week, no textural defects were observed after cooling, and the diffraction efficiency was the same as that obtained with immediate cooling after exposure.

It seems that in this case the degradation of the gratings has nothing in common with real physical–chemical processes occurring in the interior volume of a sample. Judging from everything, the degradation is due to the difference of the temperature expansion coefficients of the nematic and the polymer. This results in mechanical stresses in a cooled sample, causing the polymer layers to detach from the substrate and the nematic to flow into the gap that is formed. Thus, this degradation is most likely a technical defect of the mixture rather than a serious process that must be studied in detail.

3.3. NOA-65-5CB Mixtures

For mixtures of this type it should be noted, first and foremost, that in contrast to preceding mixtures the diffraction efficiency of the gratings, recorded at room temperature (with phase separation during polymerization and formation of PDLC), was not so bad. Consequently, in this case we performed comparative investigations for two cases: measurements at room and higher temperatures. The thickness of the samples was also 11 μm, and the exposure time needed to complete polymerization was slightly less than 7 min in this case. The nematic concentration in the initial mixture was 50% for all results presented below (which simply corresponds to the experimental optimal; see preceding section).

The morphologies of the gratings obtained at room temperature (T = 22°C) corresponded to the ordinary alternation of the layers of the polymer and PDLC, as in Fig. 3a. The only differences were that the characteristic drop size was much smaller (of the order of 0.5 μm) and the director in the nematic in the drops was oriented along the grating lines (70–80% of the drops according to a visual count).

Figure 10 shows the dependences of the diffraction efficiency on the radiation intensity with room-temperature exposure. It is evident that here the intensity is optimal, but its value does not depend on the period of the grating. The diffraction efficiencies obtained are moderate: 30% for Λ = 1.8 μm and less for shorter grating periods. Substantial losses (up to 20% of the power of probe beam) due to stochastic scattering are observed.

For high temperatures (T = 45°C, which is higher than the *N–I* transition temperature in pure 5CB) the analogous dependences are presented in Fig. 11. It is easy to see that the diffraction efficiencies are twice the values at room temperature. This is due primarily to the absence of scattering losses (the grating morphology is virtually identical to that presented in Fig. 3b). The optimal intensities turned out to be much higher than for the room-temperature case, and they increased as the grating period decreased. The dependences of the optimal intensity and the corresponding diffraction intensity are presented, respectively, in Figs. 12a and 12b for both

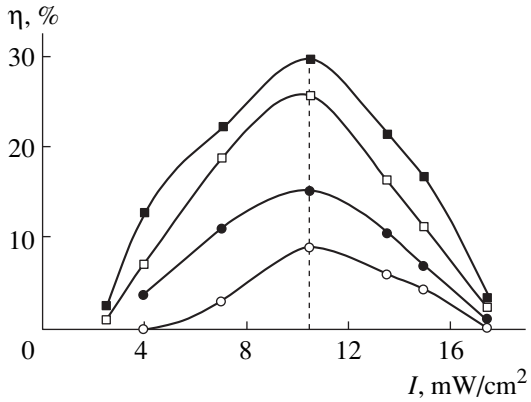


Fig. 10. Intensity dependence of the diffraction efficiency for NOA-5CB mixtures irradiated at T = 22°C, Λ = (■) 1.8, (□) 0.56, (●) 0.38, and (○) 0.22 μm.

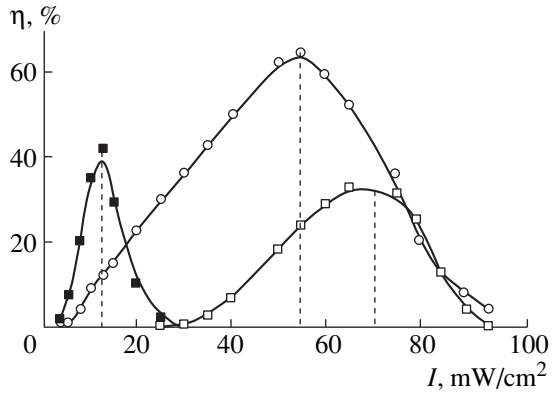


Fig. 11. Intensity dependence of the diffraction efficiency for NOA-5CB, $T = 45°C$, $\Lambda = (\blacksquare)$ 1.8, $(\circ)$ 0.38, and $(\circ)$ 0.22 μm.

exposure regimes. It is evident that the "hot" exposure regime, aside from everything else, substantially expands the range of admissible grating periods toward shorter values. Specifically, it is possible to obtain a grating with a period of 0.17 μm (compared with your 0.25 μm at room temperature). Even though this difference seems negligible, it is of central importance from the standpoint of the possibility of recording Bragg reflection gratings and holograms. The second value presented above (0.25 μm) corresponds to a reflection grating for $\lambda = 0.66$ μm (far red part of the spectrum, almost the infrared range); the first value (0.17 μm) corresponds to $\lambda = 0.51$ μm (green line of the argon laser).

The longevity of the samples in these experiments was much better than for the two preceding mixtures. The properties of the gratings remained unchanged for at least one month.

4. DISCUSSION AND CONCLUSIONS

We shall now formulate the basic qualitative results obtained.

(1) The technique described above makes it possible to obtain an actually new type of composite material—a spatially periodic sequence of submicron layers of a polymer and a liquid crystal. The texture of the grating formed is free of defects with scale of the order of the wavelength of the optical wave; this makes it possible to obtain, specifically, a diffraction grating whose efficiency is much higher than that of PDLC.

(2) The spatial period of the gratings formed can be substantially shorter than that of PDLC gratings in the same initial material. This makes it possible, specifically, to obtain reflective gratings and holograms.

(3) The nature of the processes leading to the formation of the above-described gratings lies not in the optically induced phase separation, as in for PDLC but rather in the induced spatially nonuniform polymerization of the mass transfer (diffusion) of the components of the initial mixture. Consequently, the texture of the grating obtained does not contain stochastic defects which are characteristic of random processes such as nucleation with phase separation.

(4) The dependence of the diffraction efficiency of the gratings obtained on the intensity of the writing radiation exhibits a peak whose magnitude and position depend on the period of the grating.

(5) The lower limit on the grating period below which it is impossible to write a grating for any intensity (by analogy to radio engineering we shall call it the "cut off period") is observed for each specific type of initial mixture used.

(6) Degradation processes occur in the structures described above, but these processes are quite simple and in all probability they can be prevented.

We propose the following scenario of the processes which are responsible for the formation of gratings. We believe that this scenario leads to a qualitative correspondence of the gratings to the observed properties. First, even though before mass transfer starts the nematic has a spatially uniform distribution, we assume that the processes can be described on the basis of the standard Fick diffusion.

However, this diffusion occurs not in the entire volume of the sample but only in its "active" part, which is

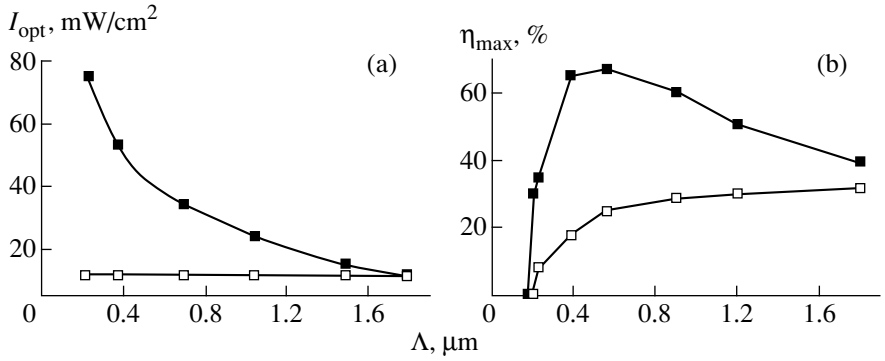


Fig. 12. Optimal intensity (a) and maximum attainable diffraction efficiency (b) versus the grating period Λ for the mixture NOA-5CB; $T = (\blacksquare)$ 45 and $(\square)$ 22°C.

not occupied by stationary polymer chains (the "passive" volume). We assume that only part of the nematic molecules, which approximately corresponds to the solubility of the nematic in the polymer, remains in the passive volume. The excess domestic is displaced into the active volume, where its concentration is spatially modulated and ordinary diffusion, tending to equalize the concentration, starts. Thus, the average concentration of the nematic over the entire volume (which is responsible for modulation of the index of refraction) is spatially modulated. Thermodynamically, this formulation of the problem is essentially analogous to osmotic processes, where the passive volume of the stationary polymer acts like a membrane. The corresponding thermodynamic analysis was performed in [6]. It was based on considerations of spatial uniformity of chemical potentials of the components in the final equilibrium state. This analysis is quite simple, but, unfortunately, it is fruitless from the standpoint of predictions, since many unknown parameters, such as the intermolecular interaction energy, are used.

Diffusion of the monomer, which likewise tends to equalize its concentration in the active volume, starts simultaneously with diffusion of the nematic. This enriches with the monomer the regions where the reaction rate is highest (these regions correspond to the maximum of the interference pattern of the radiation). This process presumes to possible reaction machines. The first one (slow polymerization) is a regime where the reaction is slow enough (for a given grating period) so that the monomer can diffuse before it is consumed in the reaction. In the second regime the reaction is quite rapid and there is not enough time for mass transfer of the monomer to occur.

At first glance the slow regime is preferable, since it gives additional modulation of the concentration of the reaction products, delivering the monomer to the actively reacting regions. However, the final judgment concerning this question requires a more detailed analysis, since it is necessary to determine what the above-indicated passive volume, whose "concentration" controls the mass-transfer processes, is.

A distinguishing feature of the passive volume is its constituent macromolecules are stationary. Consequently, not all reaction products, but only chains with sufficient length N, can be associated to the passive volume. On the other hand it is known [7] that for radical polymerization, which is typical for the systems considered, the chains formed are all the longer, the slower the reaction. Thus, we can arrive at the opposite conclusion. Specifically, the diffusion of the monomer removes the initial material from the regions where long chains are produced (corresponding to the minima of the interference pattern), thereby decreasing the final modulation of the nematic concentration. Consequently, the question of the preferability of one reaction regime over another requires a quantitative analysis.

We shall now examine qualitatively the fast regime, which is simpler. For very low intensities virtually all polymer chains—at the maxima and minima of the interference pattern—are long enough so that they can be associated to the passive volume. Consequently, its "concentration" is high but almost uniform, and therefore the modulation is weak. Conversely, for very high intensities only short chains are formed everywhere—the "concentration" of the passive volume is low and therefore the modulation is also weak. Thus, we arrive at the following conclusion, qualitatively corresponding to the experimental data, that a maximum is present in the dependence of the diffraction efficiency on the radiation intensity.

Finally, in the slow regime a situation where the diffusion time over a period becomes comparable to the lifetime of the radicals which nucleate chains can arise as the grating period decreases. This lifetime does not depend on the intensity [7]. In this case diffusion within a period of the grating equalizes the concentration of the radicals. This means that even though the initiating radiation is initially uniform (interference pattern), the reaction as such proceeds essentially uniformly in space without resulting in a grating. This argument agrees with the observed presence of a "cut off period" for a given mixture.

Thus, in all probability the propose scenario of "spatially limited diffusion" can explain the observed properties of the gratings obtained. We have constructed a quantitative formulation of the scenario, but it will be presented in a separate publication because the corresponding mathematical apparatus is quite complicated. We merely note that the results actually demonstrate qualitative agreement with experiment. Therefore there are grounds for believing that this model will be capable of predicting results in this field, whose completely empirical nature today [1–6] is severely holding back research.

ACKNOWLEDGMENTS

We are deeply grateful to L.M. Blinov and F. Simoni for fruitful discussions. This work was supported by the European Foundation for Regional Development as part of the project "Progetto Sud INFM," subproject FESR-UME.

REFERENCES

1. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, and T. J. Bunning, Chem. Mater. **5**, 1533 (1993).
2. R. L. Sutherland, V. P. Tondiglia, and L. V. Natarajan, Appl. Phys. Lett. **64**, 1074 (1994).
3. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, J. Nonlinear Opt. Phys. Mater. **5**, 89 (1996).
4. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, Opt. Lett. **20**, 1325 (1995).
5. D. Duca, A. V. Sukhov, and C. Umeton, Liq. Cryst. **26**, 931 (1999).
6. R. Caputo, A. V. Sukhov, and C. Umeton, J. Mol. Mater. **12**, 192 (1999).
7. P. W. Atkins, *Physical Chemistry* (Oxford Univ. Press, Oxford, 1987), p. 717.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1198–1203.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1384–1390.

SOLIDS
Structure

Nonlinear Dynamics of a Domain Wall in the Field of a Sound Wave Propagating in the Plane of the Wall

V. S. Gerasimchuk[a, *] and A. L. Sukstanskiĭ[b, **]

[a]*Donbass State Academy of Building and Architecture, Makeevka, 339023 Ukraine*
[b]*Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, Donetsk, 340114 Ukraine*
**e-mail: vsg@donace.dn.ua*
***e-mail: sukstan@purr.fti.ac.donetsk.ua*

Received March 22, 2000

Abstract—A theoretical model is proposed for describing the drift of a 180° domain wall in a weak ferromagnet in the field of elastic stresses produced by a strong sound wave propagating in the plane of the wall. The dependence of the drift velocity on the frequency, amplitude, and polarization of the sound wave is found. It is predicted that drift of a striped domain structure is possible. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The investigation of the influence of the magnetoelastic interaction on the static and dynamic properties of magnets is one of the classical problems of the physics of magnetically ordered media. The theoretical and experimental works devoted to this interaction study its effect on the spectrum of linear excitations of a magnet (spin waves), magnetoacoustic resonance, and so on. A substantial body of work on the influence of the magnetoelastic interaction on the nonlinear excitations of a magnet is devoted to the analysis of the phonon stopping of domain walls, the scattering of sound by a domain structure, and the mutual excitation of acoustic and domain vibrations [1–7]. The direct interaction of domain walls with an elastic stress field produced by an external sound wave has been much less studied, though this interaction results in a variety of interesting effects, specifically, the so-called drift motion of domain walls in an oscillating external field. The theory of such motion, based on the method of averaging when solving the approximate Slonczewski equations, is given in [2, 3]. A direct experimental observation of this effect was made in [8].

A more systematic approach to the theoretical description of drift of domain walls in the field of a sound wave, based on an analysis and direct solution of the nonlinear equations of motion for the magnetization vector (Landau–Lifshitz equations), has been proposed in [9]. In this work, just as in [2, 3], the simplest, from the standpoint of theoretical analysis, case of the propagation of a sound wave perpendicular to the plane of a domain wall is studied. At the same time, in direct experimental study of the interaction of ultrasound with a domain wall [8], directed drift of a domain wall has been observed in a different geometry, specifically, the propagation of an elastic wave in the plane of a domain wall.

In this paper, a theory is developed for the drift motion of a 180° domain wall in the field of a sound wave propagating in the plane of the wall for the example of a two-sublattice model of a weak ferromagnet. This theory makes it possible to describe, specifically, the magnetic subsystem of rare-earth orthoferrites.

2. EQUATIONS OF MOTION

We shall describe the nonlinear macroscopic dynamics of a two-sublattice weak ferromagnet on the basis of the Lagrangian density L, represented in terms of the unit vector of the antiferromagnet $\mathbf{l}$, $\mathbf{l}^2 = 1$. For a weak ferromagnet, such as rare-earth orthoferrites with characteristic symmetry $2_x^-2_z^-$ (the Cartesian x-, y-, and z-axes are oriented along the a-, b-, and c-axes, respectively, of the crystal), the Lagrangian density $L(\mathbf{l})$, taking account of the magnetoelastic interaction, can be written in the form

$$L(\mathbf{l}) = M_0^2 \times \left\{ \frac{\alpha}{2c^2}\dot{\mathbf{l}}^2 - \frac{\alpha}{2}(\nabla\mathbf{l})^2 - \frac{\beta_1}{2}l_z^2 - \frac{\beta_2}{2}l_y^2 - \gamma u_{ik} l_i l_k \right\}, \quad (1)$$

where the dot indicates a time derivative, M_0 is the modulus of the magnetization vector of the sublattices, $c = gM_0(\alpha/\beta)^{1/2}/2$ is the characteristic velocity, which corresponds with the minimum phase velocity of the spin waves, δ and α are, respectively, the uniform and nonuniform exchange interaction constants, g is the gyromagnetic ratio, β_1 and β_2 are the effective anisotropy constants, u_{ik} is the elastic deformations tensor, and γ is the magnetoelastic constant. In Eq. (1) we do not include a term describing the energy of the elastic subsystem as such, since in what follows we shall con-

sider the sound wave as a prescribed external field, neglecting the back effect of the magnetic subsystem on the elastic wave.

We now introduce the angular variables θ and φ, parameterizing the unit vector $\mathbf{l}$:

$$l_x + il_z = \sin\theta\exp(i\varphi), \quad l_y = \cos\theta \tag{2}$$

in terms of which the Lagrangian density (1) assumes the form

$$\begin{aligned} L\{\theta,\varphi\} = M_0^2\Bigg\{&\frac{\alpha}{2c^2}(\dot\theta^2+\dot\varphi^2\sin^2\theta) \\ &-\frac{\alpha}{2}[(\nabla\theta)^2+(\nabla\varphi)^2\sin^2\theta]-\frac{\beta_1}{2}\sin^2\theta\sin^2\varphi \\ &-\frac{\beta_2}{2}\cos^2\theta-\gamma[\sin 2\theta(u_{xy}\cos\varphi+u_{yz}\sin\varphi)+u_{yy}\cos^2\theta \\ &+\sin^2\theta(u_{xx}\cos^2\varphi+u_{xz}\sin 2\varphi+u_{zz}\sin^2\varphi)]\Bigg\}. \end{aligned} \tag{3}$$

The dynamic retardation of the domain wall, due to dissipative processes, is taken into account by the dissipative function Q:

$$Q = \frac{\lambda M_0}{2g}\dot{\mathbf{l}}^2 = \frac{\lambda M_0}{2g}(\dot\theta^2+\dot\varphi^2\sin^2\theta), \tag{4}$$

where λ is a phenomenological relaxation constant.

The equations of motion in terms of the angular variables θ and φ have the form

$$\begin{aligned} &\alpha\left(\Delta\theta-\frac{\alpha}{c^2}\ddot\theta\right)+\sin\theta\cos\theta \\ &\times\left[\alpha\left(\frac{1}{c^2}\dot\varphi^2-(\nabla\varphi)^2\right)-\beta_1\sin^2\varphi+\beta_2\right] \\ &-\gamma[\sin 2\theta(u_{xx}\cos^2\varphi+u_{xz}2\varphi+u_{zz}\sin^2\varphi-u_{yy}) \\ &+2\cos 2\theta(u_{xy}\cos\varphi+u_{yz}\sin\varphi)] = \frac{\lambda}{gM_0}\dot\theta, \end{aligned} \tag{5}$$

$$\begin{aligned} &\alpha\nabla[\sin^2\theta(\nabla\varphi)]-\frac{\alpha}{c^2}\frac{\partial}{\partial t}(\dot\varphi\sin^2\theta)-\beta_1\sin^2\theta\sin\varphi\cos\varphi \\ &+\gamma[\sin^2\theta(u_{xx}\sin 2\varphi-2u_{xz}\cos 2\varphi-u_{zz}\sin 2\varphi) \\ &+\sin 2\theta(u_{xy}\sin\varphi-u_{yz}\cos\varphi)] = \frac{\lambda}{gM_0}\dot\varphi\sin^2\theta. \end{aligned} \tag{6}$$

If $0 < \beta_1 < \beta_2$, then in a weak ferromagnet, in the absence of external fields, the orientation of $\mathbf{l}$ along the x-axis is the equilibrium orientation and far from the spin reorientation region a 180° domain wall with the $\mathbf{l}$ vector rotating in the (xz) plane is stable. This domain wall corresponds to $\theta = \theta_0 = \pi/2$, and the angular variable $\varphi = \varphi_0(y)$ satisfies the equation

$$\alpha\varphi_0'' - \beta_1\sin\varphi_0\cos\varphi_0 = 0 \tag{7}$$

(we assume that the magnetization distribution in the domain wall is nonuniform along the y-axis; a prime denotes differentiation with respect to the coordinate). A static 180° domain wall, in which the function $\varphi_0(y)$ satisfies the boundary conditions

$$\varphi_0(-\infty) = 0, \quad \varphi_0(+\infty) = \pi, \quad \varphi_0'(\pm\infty) = 0$$

is described by the relations

$$\begin{aligned} \varphi_0' &= \frac{1}{y_0}\sin\varphi_0 = \frac{1}{y_0}\cosh^{-1}\frac{y}{y_0}, \\ \cos\varphi_0(y) &= -\tanh\frac{y}{y_0}, \end{aligned} \tag{8}$$

where $y_0 = (\alpha/\beta_1)^{1/2}$ is the thickness of the domain wall.

3. PERTURBATION THEORY

To analyze the motion of a domain wall in the field of a sound wave we shall use, following [9–11], a version of perturbation theory for solitons, assuming the amplitude of the sound wave to be sufficiently small. We shall determine the collective variable $Y(\mathbf{r}_\perp, t)$ as the local coordinate of the center of the domain wall (i.e., the points $y = Y(\mathbf{r}_\perp, t)$, where $\varphi_0 = \pi/2$) at an arbitrary time t and we shall seek the solution of the equations of motion in the form

$$\begin{aligned} \theta(\mathbf{r},t) &= \pi/2 + \vartheta_1(\xi,\mathbf{r}_\perp,t) + \vartheta_2(\xi,\mathbf{r}_\perp,t) + \ldots, \\ \varphi(\mathbf{r},t) &= \varphi_0(\xi) + \psi_1(\xi,\mathbf{r}_\perp,t) + \psi_2(\xi,\mathbf{r}_\perp,t) + \ldots, \end{aligned} \tag{9}$$

where $\xi = y - Y(\mathbf{r}_\perp, t)$, $\mathbf{r}_\perp = (x, z)$, and the indices $n = 1, 2, \ldots$ indicate the order of smallness of the quantity with respect to the amplitude of the sound wave. The function $\varphi_0(\xi)$ has the same structure as the static solution $\varphi_0(y)$ (8), and the terms ϑ_n and ψ_n ($n = 1, 2, \ldots$) describe the distortion of the shape of a domain wall as a result of excitation of spin waves accompanying the interaction of the domain wall with a sound wave.

We note that when a sound wave propagates perpendicular to the plane of the domain wall [9] (i.e., for $k_x = k_z = 0$, $k_y \neq 0$) the flexural distortions of the domain wall, which are associated with the dependence of the position of the center of the wall on the transverse coordinate $\mathbf{r}_\perp$, are absent, $Y = Y(t)$, and the distortions of the form of the domain wall are due only to the excitation of spin waves. However, in the case considered in the present paper, where waves propagate in the plane of the domain wall ($k_y = 0$, $k_x \neq 0$, $k_z \neq 0$), the local position of the center of the domain wall depends not only on time but also on the transverse coordinate $\mathbf{r}_\perp$, $Y = Y(\mathbf{r}_\perp, t)$.

We shall determine the drift velocity of a domain wall as the average value of the instantaneous velocity $V(\mathbf{r}_\perp, t) = \dot{Y}(\mathbf{r}_\perp, t)$ over a period of the oscillations, $V_{\rm dr} = \overline{V(\mathbf{r}_\perp, t)}$ (the bar denotes averaging over a period of the sound wave).

Assuming the sound wave propagating parallel to the plane of the domain wall to be monochromatic with frequency ω, $\mathbf{u}(\mathbf{r}_\perp, t) = \mathbf{u}_0 \exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)$, and setting $Y = Y_1 + Y_2 + \ldots$, the equations of the first approximation in the amplitude $\mathbf{u}_0$ of the sound wave can be written in the form

$$\left(\hat{L} + \frac{1}{\omega_0^2}\frac{\partial^2}{\partial t^2} + \frac{\omega_r}{\omega_0^2}\frac{\partial}{\partial t}\right)\psi_1(\xi, \mathbf{r}_\perp, t) = y_0 \sin\varphi_0(\xi)$$

$$\times\left(-\Delta_\perp Y_1 + \frac{1}{c^2}\frac{\partial^2 Y_1}{\partial t_2} + \frac{\lambda}{\alpha g M_0}\frac{\partial Y_1}{\partial t}\right)$$

$$-\frac{i\gamma}{\beta_1}\exp[i(\mathbf{k}_\perp\mathbf{r}_\perp - \omega t)][(k_z u_{0z} - k_x u_{0x})\sin 2\varphi_0(\xi) \tag{10}$$

$$+ (k_z u_{0x} + k_x u_{oz})\cos 2\varphi_0(\xi)],$$

$$\left(\hat{L} + \sigma + \frac{1}{\omega_0^2}\frac{\partial^2}{\partial t^2} + \frac{\omega_r}{\omega_0^2}\frac{\partial}{\partial t}\right)\vartheta_1(\xi, \mathbf{r}_\perp, t) = \frac{i\gamma u_{0y}}{\beta_1} \tag{11}$$

$$\times \exp[i(\mathbf{k}_\perp\mathbf{r}_\perp - \omega t)][k_x \cos\varphi_0(\xi) + k_z \sin\varphi_0(\xi)],$$

where $\Delta_\perp = \partial^2/\partial x^2 + \partial^2/\partial z^2$, $\sigma = (\beta_2 - \beta_1)/\beta_1$, and $\omega_0 = c/y_0$ is the frequency of activation of the lower branch of the volume spin waves, $\omega_r = \lambda\delta g M_0/4$ is the characteristic relaxation frequency, $\mathbf{k}$ is the wave vector of the sound wave, $k = |\mathbf{k}| = \omega/s$, and s is the speed of the wave.

The operator $\hat{L}$ has the form of the Schrödinger operator with a nonreflective potential:

$$\hat{L} = -y_0^2\frac{d^2}{d\xi^2} + 1 - \frac{2}{\cosh^2(\xi/y_0)}. \tag{12}$$

The spectrum and wave functions of the operator $\hat{L}$ (12) are well known. The spectrum possesses one discrete level with eigenvalue $\lambda_0 = 0$, which corresponds to the localized wave function

$$f_0(\xi) = \frac{1}{\sqrt{2y_0}}\cosh^{-1}\frac{\xi}{y_0} \tag{13}$$

and a continuous spectrum $\lambda_p = 1 + p^2 y_0^2$, which corresponds to the eigenfunctions

$$f_p(\xi) = \frac{1}{b_p\sqrt{L}}\left(\tanh\frac{\xi}{y_0} - ipy_0\right)e^{ip\xi}, \tag{14}$$

where $b_p = \sqrt{1 + p^2 y_0^2}$ and L is the length of the crystal.

The functions $\{f_0, f_p\}$ form a complete orthonormal set, and it is natural to seek the solutions of the equations of the first approximation (10) and (11) in the form of an expansion in terms of this set:

$$\vartheta_1(\xi, \mathbf{r}_\perp, t)$$

$$= \mathrm{Re}\left\{\left[\sum_p c_p f_p(\xi) + c_0 f_0(\xi)\right]\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\right\}, \tag{15}$$

$$\psi_1(\xi, \mathbf{r}_\perp, t)$$

$$= \mathrm{Re}\left\{\left[\sum_p d_p f_p(\xi) + d_0 f_0(\xi)\right]\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\right\}. \tag{16}$$

The following important circumstance should be noted here. The equations of the first approximation (10) and (11) describe the excitation of linear spin waves against the background of a domain wall. The last term in the expansion in the function $\psi_1(\xi, \mathbf{r}_\perp, t)$ corresponds to a shear (goldstone) mode, i.e., the motion of a domain wall as a whole. However, the corresponding degree of freedom of the system has already been taken into account by introducing the collective coordinate $Y(\mathbf{r}_\perp, t)$ in the definition of the variable ξ. Consequently, the shear mode in the expansion (16) must be dropped, i.e, we must set $d_0 = 0$ (see the monograph [12] for a discussion of this question).

The coefficients in the expansions (15) and (16) are found in the standard manner by multiplying the right-hand side of Eqs. (10) and (11) by f_p^* or f_0^* and integrating over the variable ξ. For a monochromatic sound wave $\mathbf{u} = \mathbf{u}_0\cos(\omega t - \mathbf{k}\mathbf{r}_\perp)$ with frequency ω we obtain from Eqs. (10) and (11)

$$\psi_1(\xi, \mathbf{r}_\perp, t) = \frac{\gamma}{\beta_1}\mathrm{Re}\{[a_x\cos\varphi_0(\xi) + a_z\sin\varphi_0(\xi)]\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\}, \tag{17}$$

$$\vartheta_1(\xi, \mathbf{r}_\perp, t) = \frac{\gamma}{\beta_1} \times \mathrm{Re}\{[b_1 B(\xi) + b_2 G_1(\xi)]\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\}. \tag{18}$$

Here we have introduced the notation

$$a_x = \frac{ik_x u_{0y}}{1 + \sigma - q_1 - iq_2},\quad a_z = \frac{ik_z u_{0y}}{\sigma - q_1 - iq_2},$$

$$b_1 = iB(\xi)(k_z u_{0z} - k_x u_{0x}),$$

$$b_2 = -iG_1(\xi)(k_z u_{0x} + k_x u_{0z}),$$

$$B(\xi) = \frac{y_0}{2}\int_{-\infty}^{\infty}\frac{\tanh(\xi/y_0)\cos(p\xi) + py_0\sin(p\xi)}{(\lambda_p - q_1 - iq_2)\cosh(\pi p y_0/2)}dp,$$

$$G_1(\xi) = \frac{y_0}{2}\int_{-\infty}^{\infty}\frac{\tanh(\xi/y_0)\sin(p\xi) - py_0\cos(p\xi)}{(\lambda_p - q_1 - iq_2)\sinh(\pi p y_0/2)}dp,$$

where $q_1 = (\omega/\omega_0)^2$ and $q_2 = \omega\omega_r/\omega_0^2$.

The condition $d_0 = 0$ determines the equation for the velocity of the domain wall in an approximation linear in the field $Y_1(\mathbf{r}_\perp, t)$:

$$-\Delta_\perp Y_1 + \frac{1}{c^2}\frac{\partial^2 Y_1}{\partial t^2} + \frac{\lambda}{\alpha g M_0}\frac{\partial Y_1}{\partial t} = 0. \qquad (19)$$

Since the right-hand side of Eq. (19) is zero, Eq. (19) possesses only a trivial solution (we are interested only in forced solutions of the equations of motion). Thus, in an approximation linear in the wave amplitude sound does not give rise to motion of the domain wall, but rather it leads only to excitation of localized and nonlocalized spin waves, which is described by the expressions (17) and (18). We note that a similar result—absence of forced motion of a domain wall in the linear approximation—also occurs for propagation of a sound wave perpendicular to the plane of the wall [9].[1]

We shall now analyze the equations of the second approximation in the amplitude of the acoustic wave. Taking account of the solutions found above for the equations of the first approximation (10) and (11), the second-approximation equation for the function $\psi_2(\xi, t)$ can be put into the form

$$\hat{L}\psi_2(\xi, \mathbf{r}_\perp, t) = \left(-\Delta_\perp Y_2 + \frac{1}{c^2}\frac{\partial^2 Y_2}{\partial t^2} + \frac{\lambda}{\alpha g M_0}\frac{\partial Y_2}{\partial t}\right)$$

$$\times y_0\sin\varphi_0 + \left[\left(\frac{\partial Y_1}{\partial x}\right)^2 + \left(\frac{\partial Y_1}{\partial z}\right)^2 - \left(\frac{V_1}{c}\right)^2\right]\frac{\sin 2\varphi_0}{2}$$

$$+\frac{y_0^2}{c^2}V_1\dot{\psi}_1'(\xi, \mathbf{r}_\perp, t) - 2y_0\vartheta_1(\xi, \mathbf{r}_\perp, t)\vartheta_1'(\xi, \mathbf{r}_\perp, t)\sin\varphi_0$$

$$+\psi_1^2(\xi, \mathbf{r}_\perp, t)\sin 2\varphi_0 + \frac{\gamma}{\beta_1}\mathrm{Re}\{[2(k_x u_{0x} - k_z u_{0z}) \qquad (20)$$

$$\times\psi_1(\xi, \mathbf{r}_\perp, t)\cos 2\varphi_0 - (k_x u_{0y}\sin\varphi_0 - k_z u_{0y}\cos\varphi_0)$$

$$\times\vartheta_1(\xi, \mathbf{r}_\perp, t) + 2(k_z u_{0x} + k_x u_{0z})$$

$$\times\psi_1(\xi, \mathbf{r}_\perp, t)\sin 2\varphi_0]\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\}.$$

The second-approximation equation for the function $\vartheta_2(\xi, \mathbf{r}_\perp, t)$ does not contain a second-order term in the expansion of the collective coordinate Y_2, and consequently it will be of no interest to us below.

The solution of Eq. (20) can once again be sought as an expansion in the eigenfunctions of the operator $\hat{L}$, similar to Eq. (16). Just as for the first-approximation equations, we must require that the expansion coefficient $d_0^{(2)}$, corresponding to a shear mode, must vanish. Consequently, to calculate the velocity of the domain wall there is no need to find the complete solution of Eq. (20); it is sufficient to calculate the coefficient $d_0^{(2)}$ and equate it to zero. The result is an equation for the function $Y_2(\xi, \mathbf{r}_\perp, t)$, which, using the solutions (17) and (18), can be put into the form

$$\left(-\Delta_\perp Y_2 + \frac{1}{c^2}\frac{\partial^2 Y_2}{dt^2} + \frac{\lambda}{\alpha g M_0}\frac{\partial Y_2}{\partial t}\right) \qquad (21)$$

$$= N + N_1\exp(2i\omega t) + N_2\exp(-2i\omega t),$$

where

$$N = \mathrm{Re}\int_{-\infty}^{\infty} d\xi\sin\varphi_0\left\{-\frac{V_1 y_0^2}{c^2}\dot{\psi}' + 2y\vartheta_1\vartheta_1'\sin\varphi_0\right.$$

$$-\psi_1^2\sin 2\varphi_0 - [2(k_x u_{0x} - k_z u_{0z})\psi_1\cos 2\varphi_0$$

$$+2(k_z u_{0x} + k_x u_{0z})\psi_1\sin 2\varphi_0 - k_x u_{0y}\vartheta_1\sin\varphi_0 \qquad (22)$$

$$\left.+k_z u_{0y}\vartheta_1\cos\varphi_0]\left(\frac{i\gamma}{\beta_1}\right)\exp(i\mathbf{k}\mathbf{r}_\perp - i\omega t)\right\}.$$

The explicit form of the coefficients N_1 and N_2, whose structure is similar to that of Eq. (22) but is more complicated, is not presented here, since on subsequent averaging of the solution of Eq. (21) the terms with N_1 and N_2 vanish.

The Eq. (21) describes the dynamics of a domain wall in second-order perturbation theory in the amplitude of the sound wave.

4. DRIFT OF THE DOMAIN STRUCTURE

The relations (21) and (22) describe periodic oscillations of the wall and the drift motion of the domain boundary, the effect of interest to us. Integrating Eq. (21) and averaging the obtained solution over a period of the sound wave, we obtain the desired drift velocity of the domain wall $V_{dr} = \bar{V}_2 = \overline{\partial Y_2/\partial t}$:

$$V_{dr} = \mu_t(\omega)k_x k_z u_{0y}^2$$

$$+\mu_l(\omega)[k_x k_z(u_{0z}^2 - u_{0x}^2) + u_{0x}u_{0z}(k_z^2 - k_x^2)], \qquad (23)$$

[1] The experimentally observable oscillations of a domain wall in ferromagnets [8] under the action of a sound wave are due to the nonuniformity of the wall, specifically, the presence of Bloch lines in it. A sound wave gives rise to oscillations of Bloch lines, and the gyrotropic forces exerted by the latter result in oscillations of the entire domain wall.

where μ_t and μ_l are the nonlinear mobilities of the domain wall in the field of an acoustic wave:

$$\mu_t(\omega) = -\mu_0\left(\frac{q_2^2}{[(1+\sigma-q_1)^2+q_2^2][(\sigma-q_1)^2+q_2^2]}\right),$$
$$\mu_l(\omega) = \mu_0\frac{\pi\eta}{8}\left(\frac{1}{(1-q_1)^2+q_2^2}\right), \tag{24}$$

$\mu_0 = \nu_0\delta(\gamma M_0)^2/4\beta_1$, $\nu_0 = \pi y_0 g^2/4\omega_r$ is the characteristic nonlinear mobility of the domain wall in a weak ferromagnet in an ac magnetic field [11], and η is a numerical factor of the order of 10^{-3}.

We note that the characteristic frequency dependence of the nonlinear mobilities μ_t and μ_l, just as that of the characteristic quantities determining the drift of a domain wall in an oscillating magnetic field [11], is of a resonance character at the frequencies ω_0, $\omega_0(1+\sigma)^{1/2}$, and $\omega_0\sigma^{1/2}$. The first two frequencies correspond to the activation frequencies of volume spin waves in a weak ferromagnet, and the third frequency corresponds to the activation frequency of a spin wave localized on the domain wall.

For a sound wave propagating perpendicular to the plane of the domain wall [9] the nonlinear mobilities $\mu_{ij}(k)$ possess a somewhat different dependence on the frequency and the parameters of the weak ferromagnet.[2] In the long-wavelength approximation ($ky_0 \ll 1$)

$$\frac{\mu_{xx}}{\mu_0} = \frac{\pi}{2}\frac{ky_0q_2}{\sigma(1+\sigma)^2}, \quad \frac{\mu_{zz}}{\mu_0} = \frac{2ky_0q_2}{\sigma^2(1+\sigma)},$$
$$\frac{\mu_{xz}}{\mu_0} = \frac{(ky_0)^2}{\sigma(1+\sigma)}\left[\sigma\left(\frac{1}{2}+\frac{1}{\pi}+\frac{1}{3\pi(1+\sigma)}\right)-\frac{\pi}{3}\right], \tag{25}$$

i.e., $\mu_{ij} \propto \omega^2$, and in the short-wavelength approximation ($ky_0 \gg 1$)

$$\frac{\mu_{xx}}{\mu_0} = -\frac{8}{3\pi}\frac{q_2}{(ky_0)^5}, \quad \frac{\mu_{zz}}{\mu_0} = -\frac{4}{3\pi}\frac{q_2}{(ky_0)^5},$$
$$\frac{\mu_{xz}}{\mu_0} = -\left(\frac{1}{2}+\frac{4}{\pi}\right)\frac{1}{(ky_0)^4} \tag{26}$$

and the velocity $V_{\rm dr} \propto \omega^{-2}$.

We shall now examine in greater detail the dependence of the drift velocity on the polarization of the sound wave. We note that the frequency ω and the components of the vector **k** which appear in Eqs. (23) and (24) are not independent; they are related with one another by the standard dispersion law for sound waves $\omega = sk$, $k = |\mathbf{k}|$, where the sound speed s is different for longitudinal and transverse waves.

[2] The corresponding expressions for the nonlinear mobility for a wave propagating perpendicular to the plane of the domain wall, which are presented in [9], are not exact and, consequently, they differ from the expressions presented above.

Since the wave propagates in the plane of the wall, i.e., in the (xz) plane, we set $\mathbf{k} = (k_x, 0, k_z) = k(\cos\varphi_s, 0, \sin\varphi_s)$. The following polarizations of the sound wave are possible:

(1) A **transverse wave** with the displacement vector perpendicular to the plane of the wall. For it $\mathbf{u} = u_0(0, 1, 0)$, and the drift velocity at a prescribed frequency of sound can be represented as

$$V_{\rm dr} = \frac{1}{2}\mu_t(\omega)(k_tu_0)^2\sin 2\varphi_s, \tag{27}$$

where $k_t = \omega/s_t$ and s_t is the speed of transverse sound.

(2) A **transverse sound wave** with displacement vector lying in the plane of the wall, $\mathbf{u} = u_0(-\sin\varphi_s, 0, \cos\varphi_s)$. In this case the drift velocity is given by the expression

$$V_{\rm dr} = \frac{1}{2}\mu_l(\omega)(k_tu_0)^2\sin 4\varphi_s. \tag{28}$$

(3) A **longitudinal wave,** for which $\mathbf{u} = u_0(\cos\varphi_s, 0, \sin\varphi_s)$. In the field of this wave we have

$$V_{\rm dr} = -\frac{1}{2}\mu_l(\omega)(k_lu_0)^2\sin 4\varphi_s. \tag{29}$$

Here $k_l = \omega/s_l$, where s_l is the speed of longitudinal sound.

For comparison, we note that for a sound wave propagating perpendicular to the plane of the domain wall [9], the wall can drift only if the polarization of the wave is transverse.

We now present some numerical estimates, using the parameters of the well-studied weak ferromagnet $YFeO_3$ [13]: $M_0 \approx 10^3$ Oe, $y_0 = 10^{-6}$ cm, $g \approx 1.76 \times 10^7$ s^{-1} Oe^{-1}, $\gamma M_0^2 \approx 10^7$ erg/cm^3, $\omega_0 = c/y_0 \approx 2 \times 10^{12}$ s^{-1}, $\omega_r \approx 0.7 \times 10^{10}$ s^{-1}, and the nonlinear mobility of the domain wall in a magnetic field $\nu_0 \approx 3.5 \times 10^{-2}$ cm/(s Oe^2) [11]. It is found that for all reasonable values of the frequency $q_2 = (\omega\omega_r/\omega_0^2) \sim 10^{-15}\omega \ll 1$, and consequently the nonlinear mobility in a transverse wave $\mu_t \sim q_2^2$ in Eq. (24) is negligibly small at all frequencies except resonance frequencies. At the resonance frequencies $\omega_1 = \omega_0\sigma^{1/2}$ and $\omega_2 = \omega_0(1+\sigma)^{1/2}$, which coincide with the activation frequencies of the localized and the top volume spin-wave branches, respectively, for the maximum admissible value of the deformation tensor $ku_0 \sim 10^{-5}$ we obtain the following value for the drift velocity (27):

$$V_{\rm dr}(\omega \approx \omega_{1,2}) \sim \mu_0(ku_0)^2 \text{ cm/s} \approx 10^{-1} \text{ cm/s}.$$

In the field of a longitudinal sound wave outside the resonance region

$$V_{\rm dr} \sim \mu_l(ku_0)^2 \sim 10^{-4} \text{ cm/s}.$$

Near the activation frequency of the bottom branch of volume spin waves the nonlinear mobility μ_l possesses a "resonance–antiresonance" singularity. The drift velocity reaches the maximum value

$$V_{\rm dr}(\omega \approx \omega_0) \sim \eta\mu_0\left(\frac{\omega_0}{\omega_r}\right)^2 (ku_0)^2 \sim 1 \text{ cm/s}.$$

Comparing our results with the results obtained in [9] shows that in order of magnitude the drift of a domain wall is somewhat greater when the sound wave propagates perpendicular to the plane of the domain wall. However, if the wave propagates along the plane of the domain wall, drift is also possible in the field of a longitudinal sound wave. The highest value of the drift velocity should be expected in magnets possessing a high magnetoelastic coupling constant and small saturation magnetization, damping, and anisotropy.

Our results agree qualitatively with the theoretical calculations of [2, 3], where the interaction of a transverse sound wave with a domain wall in a single-sublattice ferromagnet was studied using an approach with an abbreviated description of the dynamics of the domain wall.

In conclusion, we note that in the field of a sound wave propagating parallel to the plane of a domain wall (as in the case of a sound wave propagating orthogonally to the plane of a domain wall [9]) drift motion of not only a single domain wall but also a plane-parallel domain structure is possible. For this, a definite matching of the direction of rotation of the vector **l** of the antiferromagnet in neighboring domain walls is necessary.

We now introduce the topological charge of a domain wall $l_z(\pm\infty) = \mp R$ and a parameter describing the direction of rotation of the vector **l** on the wall $l_x(y = 0) = \rho$, assuming the value +1 or –1 (the magnetization distribution on the 180° domain wall (8) studied above corresponds to the values $R = \rho = +1$). In general the drift velocity of a domain wall with given values of the parameters R and ρ is determined by the formula

$$V_{\rm dr} = \rho R\{\mu_t(\omega)k_x k_z u_{0y}^2 + \mu_l(\omega)[k_x k_z(u_{0z}^2 - u_{0x}^2) + u_{0x}u_{0z}(k_z^2 - k_x^2)]\}. \quad (30)$$

Since the neighboring domain walls possess opposite values of the topological charge R, for drift of a striped domain structure formed by such walls the parameters ρ on neighboring domain walls also must be different. In other words the vector **l** must have opposite orientations on neighboring domain walls and the direction of rotation must be the same (this question is discussed in greater detail in [9]). As one can see from Eq. (30), drift of a domain structure is possible in transverse and longitudinal sound waves propagating parallel to the plane of a domain wall.

REFERENCES

1. V. G. Bar'yakhtar and B. A. Ivanov, Fiz. Met. Metalloved. **39**, 478 (1975).
2. S. I. Denisov, Fiz. Tverd. Tela (Leningrad) **31** (11), 270 (1989) [Sov. Phys. Solid State **31**, 1992 (1989)].
3. Yu. I. Gorobets and S. I. Denisov, Ukr. Fiz. Zh. **35**, 271 (1990).
4. E. A. Turov and A. A. Lugovoĭ, Fiz. Met. Metalloved. **50**, 903 (1980).
5. G. M. Nedlin and R. Kh. Shapiro, Fiz. Tverd. Tela (Leningrad) **18**, 1696 (1976) [Sov. Phys. Solid State **18**, 985 (1976)].
6. R. J. Yeh, H. E. Wigen, and H. Dötch, Solid State Commun. **44**, 1183 (1982).
7. L. É. Gurevich and É. V. Liverts, Zh. Éksp. Teor. Fiz. **88**, 1257 (1985) [Sov. Phys. JETP **61**, 743 (1985)].
8. V. K. Vlasko-Vlasov and O. A. Tikhomirov, Fiz. Tverd. Tela (Leningrad) **32**, 1678 (1990) [Sov. Phys. Solid State **32**, 978 (1990)]; Fiz. Tverd. Tela (Leningrad) **33**, 3498 (1991) [Sov. Phys. Solid State **33**, 1964 (1991)].
9. V. S. Gerasimchuk and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **106**, 1146 (1994) [JETP **79**, 622 (1994)].
10. V. G. Bar'yakhtar, Yu. I. Gorobets, and S. I. Denisov, Zh. Éksp. Teor. Fiz. **98**, 1345 (1990) [Sov. Phys. JETP **71**, 751 (1990)].
11. V. S. Gerasimchuk and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **103**, 151 (1993) [JETP **76**, 82 (1993)].
12. R. Rajaraman, *Solitons and Instantons: an Introduction to Solitons and Instantons in Quantum Field Theory* (North-Holland, Amsterdam, 1982; Mir, Moscow, 1985).
13. V. G. Bar'yakhtar, B. A. Ivanov, and M. V. Chetkin, Usp. Fiz. Nauk **146**, 417 (1985) [Sov. Phys. Usp. **28**, 563 (1985)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1204–1212.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1391–1401.

SOLIDS
Structure

Two-Dimensional and Layered Structures in the Discrete ϕ^4 Model

V. V. Savkin* and A. N. Rubtsov

Moscow State University, Moscow, 119899 Russia
**e-mail: volodya@astral.ilc.msu.su*

Received March 31, 2000

Abstract—The properties of phase transitions in two-dimensional and layered systems are investigated on the basis of a discrete ϕ^4 model by numerical and analytical methods. The only parameter a of the discrete ϕ^4 model determines the behavior of the system and makes it possible to investigate phase transitions ranging from transitions of the displacement type ($a \longrightarrow +0$) to order–disorder type ($a \longrightarrow +\infty$). The behavior of a two-dimensional system is investigated in a wide range of values of the parameter a. The temperature dependences of the squared order parameter $\eta^2(T)$ and the phase transition temperature T_c as a function of the thickness N of the system are obtained for three characteristic values of the parameter a using the Monte Carlo method. The properties of phase transitions in the discrete ϕ^4 model are investigated on the basis of the mean-field approximation and the independent-mode approximation. The results obtained in the numerical experiments are compared with the analytical approximations. It is shown that the mean-field approximation qualitatively describes the behavior of the phase-transition temperature T_c as a function of the thickness N of the system for a wide range of values of the parameter a, and the independent-mode approximation describes quantitatively, to within 5%, the results of the numerical simulation for small values of a. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The so-called structural phase transitions form an important class of phase transitions in physical systems. The main property of these transitions is the change in the point symmetry group of the medium at the phase-transition point [1]. Among physical objects with structural phase transitions, ferromagnets and ferroelectrics have been investigated most widely both theoretically and experimentally.

Order–disorder phase transitions are observed in most ferromagnetic materials. Transitions of this type can be studied on the basis of the following considerations. Let an element of the unit cell have several (for example, two) possible positions. Then at a temperature above the transition point the magnetic moment can occupy with equal probability both possible positions, and at a temperature below the phase-transition point the average number of atoms in one state becomes greater than in the other state [2].

At the same time displacement- and mixed-type transitions are observed in ferroelectric materials. In these cases the position of the charged ions in a unit cell changes continuously. A displacement-type transition is a collective phenomenon, in which a lattice consisting of atoms of one kind is displaced relative to a lattice consisting of atoms of another kind [2]. In systems with displacement-type transitions phonons are well-substantiated quasiparticles, and the phase transition itself is due to the appearance of a soft mode in the structure—a mode whose frequency vanishes at the transition point. This criterion is convenient from the experimental standpoint because the phonon spectrum is relatively easy to measure [3, 4]. We note that in some ferroelectric materials the properties of the observed phase transitions are close to those of order–disorder transitions [5–7] or to mixed-type transitions, which strictly cannot be classified as order–disorder or displacement type transitions [8].

The properties of phase transitions in two-dimensional and layered ferromagnetic films are being widely investigated experimentally [9–11] and theoretically [12–16]. At the same time the question of the properties of two-dimensional and layered ferroelectric systems has not been adequately studied theoretically. Such systems have recently been obtained experimentally in Langmuir ferroelectric films based on the polymer P(VDF-TrFE 70:30) [17, 18]. Experimental investigations of the properties of thin Langmuir films are now being performed using dielectric measurements [17–19] and nonlinear-optical methods [20, 21]. For example, experiments investigating structures with various thicknesses are described in [17]. It can be concluded from the data obtained that the phase transition temperature decreases as the number of layers in the system decreases. These works show that a detailed investigation of the microscopic models describing phase transitions in thin ferroelectric films are needed.

One of the most often used microscopic models for describing phase transitions in ferroelectric systems is the so-called discrete ϕ^4 model [22]. The system studied in this model is a collection of classical anharmonic 2–4 oscillators with harmonic interaction with nearest

neighbors. Thus the potential energy of the lattice is written in the form [22, 23]

$$V = -\frac{A}{2}\sum_{n} x_n^2 + \frac{B}{4}\sum_{n} x_n^4 + \frac{C}{2}\sum_{n,\,n'} (x_n - x_{n'})^2 \sigma_{nn'}, \quad (1)$$

where x_n is the one-dimensional displacement of the nth element of the lattice; A, B, and C are parameters of the model; $\sigma_{nn'}$ becomes 1 for nearest neighbors and 0 in other cases. A phase transition exists in the system for any A, B, $C > 0$. In this model the quantity $\eta = \langle x_n \rangle$, which vanishes at temperatures above the phase transition point and is different from zero otherwise, is the order parameter.

Using appropriate scales for the energy and the coordinates, V can be written in a form containing only the parameter $a = A/C$, which completely determines the behavior of the system:

$$V = -\frac{a}{2}\sum_{n} x_n^2 + \frac{a}{4}\sum_{n} x_n^4 + \frac{1}{2}\sum_{n,\,n'} (x_n - x_{n'})^2 \sigma_{nn'} \quad (2)$$

(energy is measured in the units CA/B and the coordinate is measured in the units $\sqrt{A/B}$). For the system (2) the order parameter $\eta = \langle x_n \rangle$ at zero temperature is one. The value of the parameter a determines the type of transition: as $a \longrightarrow +0$ displacement-type phase transition occurs in the system and as $a \longrightarrow +\infty$ an order–disorder type transition occurs [24]. In the latter case the ϕ^4 model transforms into the Ising model. The dimensionless transition temperature is a function of the parameter a.

Universal properties of the ϕ^4 model are well known–the values of the critical exponents describing the behavior of the basic physical quantities near phase-transition points [22].

The phase-transition temperatures T_c in limiting cases can be obtained for systems with various dimensions from the Ising model ($a \longrightarrow +\infty$) [22, 25] and the approximation of independent phonon modes ($a \longrightarrow +0$) [22].

For small values of the parameter a the dependence $\eta^2(T)$ is nearly linear virtually up to T_c, i.e., it can be described using Landau's phenomenological theory. However, even for $a \approx 10$ the critical behavior $\eta(T) \propto (T_c - T)^\beta$ extends to temperatures comparable to T_c [26, 27]. For small values of a the discrete ϕ^4 model is a lattice analog of the field ϕ^4 model. In this context the behavior of the coupling constant for the 1D and 2D cases was investigated in [28, 29] numerically and analytically in the ϕ^4 model. In [30] the ϕ^4 model was investigated for a 2D lattice. Numerical simulation gave the dependences of the specific heat, energy, magnetization, and susceptibility as functions of two parameters of the model [30]. In addition the phase diagram was obtained in [30] by a numerical method. New applications of the Monte Carlo method have been studied for the 2D lattice of the ϕ^4 model [31, 32].

Table

Dimension	3D	2D	1D
$T_c(a \longrightarrow +0)$	2.64	Tens to zero	Does not exist
$T_c(a \longrightarrow +\infty)$	9.02	4.53	Does not exist

We note that, even though extensive investigations of the ϕ^4 model have been performed, the question of the properties of the model in a wide range of values of the parameter a (from $a \longrightarrow +0$ to $a \longrightarrow +\infty$) for 2D lattices remains open. The three-dimensional case was recently studied in [33]. There the discrete ϕ^4 model for a 3D lattice was investigated numerically and analytically. The Monte Carlo method was used to obtain the phase transition temperature T_c as a function of the order parameter a of the model, which in the limits $a \longrightarrow +\infty$ and $a \longrightarrow +0$ is identical to the well-known asymptotics of order–disorder and displacement transitions, respectively (see table). The simplest mean-field approximation was used as an analytical approximation to describe the numerical results, and an analytical scheme, which is a combination of the mean-field and independent-modes approximations, which describes to a high degree of accuracy the numerical results for a wide range of values of the model parameter a was proposed.

Layered systems have not been studied on the basis of the discrete ϕ^4 model. Such investigations have been confined only to the Ising model, in the first place because of the simplicity of the model and, in the second place, in view of their urgency because of the experiments being performed with thin ferromagnetic films [9, 10]. The Ising model is a typical microscopic model for describing order–disorder type transitions [1, 34]. The first investigations of layered systems in the Ising model were published quite a long time ago [35–38]. Numerical methods and the mean-field approximation were used in these works. In addition the effect of the interaction in the surface layers of the structures on the phase-transition temperature was investigated in the works indicated. However, there is still great interest in studying layered systems in the Ising model [12–16]. New analytical approximations are being proposed and discussed in detail [12, 13, 15], and the effect of various types of lattices and the properties of the spin (Ising or Heisenberg) on the phase-transition temperature is being investigated [16] and the behavior of the effective critical exponents in the entire temperature range (from 0 up to T_c) is being studied [14].

In the present paper the results of investigations of the properties of two-dimensional and layered structures by numerical and analytical methods on the basis of the discrete ϕ^4 model, for which the Ising model is one of the limits ($a \longrightarrow +\infty$), are presented. Attention is devoted primarily to the following: first, the study of the behavior of the main parameters for a wide range of values of the parameter a; second, the study of layered

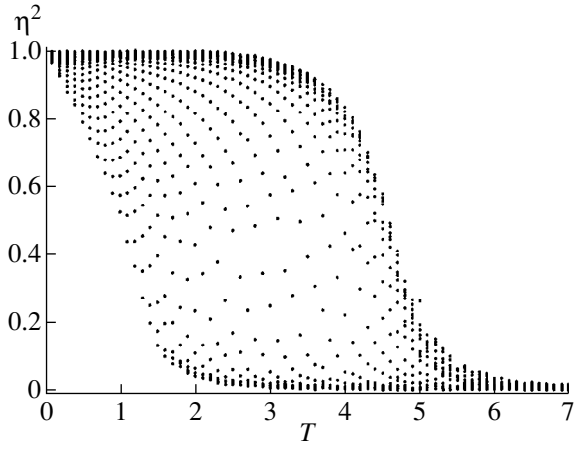


Fig. 1. Temperature dependence of the squared order parameter for a 2D lattice in a wide range of values of the parameter a of the discrete ϕ^4 model. The curves were obtained by numerical simulation. The extreme left-hand dependence corresponds to $a = 0.98$ and the extreme right-hand dependence corresponds to $a = 4000$. Between neighboring curves the parameter a differs by a factor of $\sqrt{2}$.

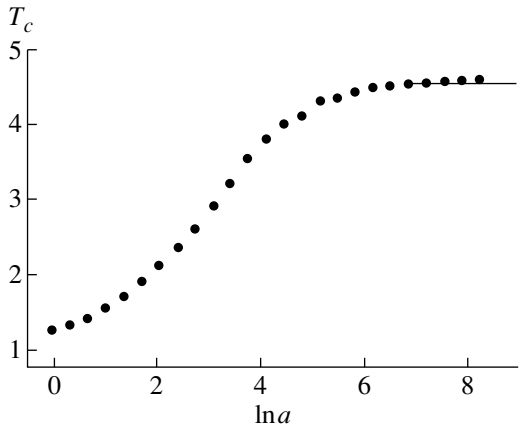


Fig. 2. Phase transition temperature T_c versus the parameter a of the discrete ϕ^4 model for a 2D lattice. The data were obtained from Fig. 1. The horizontal straight line corresponds to the transition temperature of the 2D Ising model: $T_c = 4.53$.

systems in the context of the discrete ϕ^4 model; and, third, comparison of the results of an investigation which were obtained by numerical and analytical methods. The main parameters of the model (the order parameter $\eta = \langle x_n \rangle$ and the phase-transition temperature T_c) are studied as a function of the number N of layers in the system under study for various values of the parameter a. The investigation is performed on the basis of Monte Carlo simulation (Section 2) and using very simple analytical mean-field and independent-modes approximations (Section 3). The analytical schemes presented are discussed as a method of describing the results obtained by numerical methods.

2. NUMERICAL SIMULATION

Monte Carlo simulation on the basis of the discrete ϕ^4 model was performed for various values of the parameter a, determining the type of phase transition, and for systems of various thicknesses. To this end it is convenient to use the expression in the form (2) for the potential energy.

Numerical experiments were performed on a $15 \times 15 \times N$ lattice of atoms, where N determines the thickness of the system; $N = 1$ corresponds to the 2D case. Periodic boundary conditions were used for the first two directions, and free boundary conditions were used for the third direction, determining the thickness of the system. The effect of the finiteness of the size of the atomic lattice in the Monte Carlo simulation, for example, for the 3D case, can be found in [33]. The number of averages for each lattice element far from a phase-transition point (with respect to temperature) was 5000, while for temperatures close to the transition temperatures the number of averages was taken to be 15000. For large values of the parameter a, with probability 0.1 the so-called "magic" displacements were used, which change the sign of the coordinate x_n, so that the system under study could reach thermodynamic equilibrium quite rapidly. Standard criteria were used to analyze the displacements x_n for each atom.

The squared order parameter was determined as

$$\eta^2 = \langle X_0^2 - X_1^2 \rangle, \tag{3}$$

where

$$X_k = \frac{1}{\sqrt{N^*}} \sum_n x_n \exp(ikn)$$

is the Fourier transform of the displacement x_n and N^* is the total number of atoms in the lattice. Taking account of the first harmonic in the expression for the squared order parameter makes it possible to decrease the magnitude of the fluctuations at high temperatures and near a phase transition.

The behavior of the order parameter for various values of the parameter a was investigated for a 2D lattice and for layered systems. For definiteness the temperature dependences of the squared order parameter for a 2D lattice are presented in Fig. 1. The model parameter a varies from 4000 (extreme right-hand dependence) to 0.98 (extreme left-hand dependence). Between neighboring dependences the parameter a differs by a factor of $\sqrt{2}$. The dependence of the phase transition temperature on the parameter a was determined from the dependences obtained (Fig. 2). The horizontal line shows the asymptotics for the 2D Ising model. The phase transition temperature T_c was determined as the inflection point of the function $\eta^2(T)$ or, which is the same thing, as the minimum of the first derivative. The correctness of this criterion can be checked by changing the lattice dimensions. If T_c does not change as the

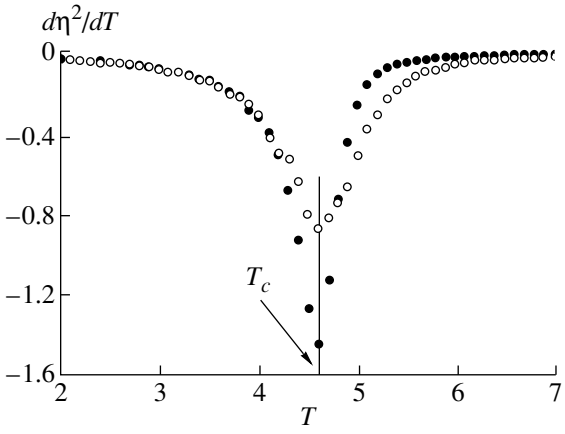


Fig. 3. Temperature dependences of the derivative of the squared order parameter, which were obtained using the Monte Carlo simulation for the 2D lattice. The parameter of the discrete ϕ^4 model $a = 4000$. Two dependences corresponding to different sizes of the atomic lattices are shown: open circles—15×15 atoms, filled circles—30×30 atoms. The phase transition occurs at the point where the derivative has a minimum value. The transition temperature determined in this manner (indicated by the arrow) is 4.6 and remains unchanged as the lattice size increases. The exact value of the transition temperature in the 2D Ising model is 4.53 [22].

size of the system increases, then the transition temperature determined in this manner has the correct value. Figure 3 shows results for a case close to the 2D Ising model ($a = 4000$). The transition temperature determined in this manner does not change as the lattice size is doubled and it agrees well with the known value of the transition temperature for the 2D Ising model. These assertions are valid for a wide range of values of the parameter a of the discrete ϕ^4 model.

Numerical experiments for layered structures were performed with systems with $N = 1, 2, 3, 4, 5$, and 6 layers. The dependence $T_c(a)$ for the 3D case has a monotonic form with two asymptotic forms for small (~1) and large (~3000) parameters a. The values of the temperatures T_c in the indicated asymptotic forms are 2.64 and 9.02, respectively (see table). The system for the 2D case (Fig. 2) also shows a similar behavior but with asymptotics 0 and 4.53 for small and large values of the parameters a, respectively (see table). The case of small parameters a ($a < 1$) has already been investigated (see, e.g., [30]). In the present work three characteristic values of the parameter a were chosen, close to the lower and upper asymptotics, corresponding to displacement and order–disorder type transitions, respectively, as well as for an intermediate case. The values considered for the parameter a are 0.98, 44.2, and 4000 (Fig. 2); the data for the 3D systems with such parameters are presented in [33].

The results of numerical experiments for three values of the parameter a and systems with various thicknesses are presented in Figs. 4–6. These figures show the temperature dependences of the squared order parameter. The dependences in Fig. 4 correspond to the case of displacement-type phase transitions and the dependences in Fig. 6 correspond to order–disorder transitions. The phase transition temperatures T_c were determined from the curves obtained in the manner described above. The results are presented in Fig. 7. This figure shows the dependences of the phase transition temperatures as a function of the number of layers in the system for the corresponding parameters a of the discrete ϕ^4 model (dark symbols). The transition temperatures indicated by the horizontal lines for the 3D case are taken from [33] for each parameter a considered.

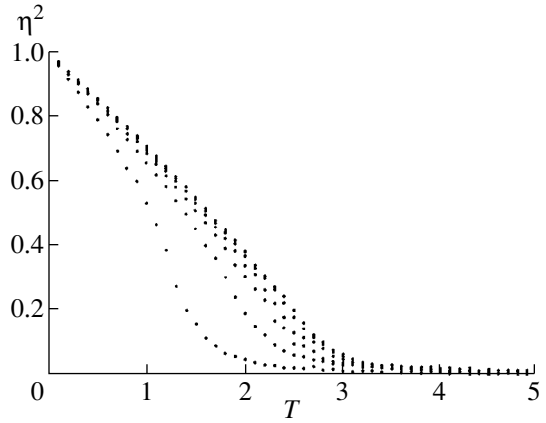


Fig. 4. Temperature dependences of the squared order parameter obtained by Monte Carlo simulation for layered systems. The parameter of the discrete ϕ^4 model $a = 0.98$. Six curves corresponding to system thickness $N = 1, 2, 3, 4, 5$, and 6 layers are shown. The extreme left-hand curve corresponds to a single-layer system and the extreme right-hand curve corresponds to a six-layer system.

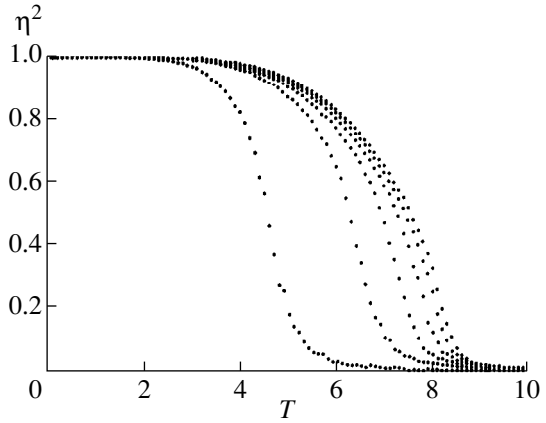


Fig. 5. Temperature dependences of the squared order parameter obtained by Monte Carlo simulation for layered systems. The parameter of the discrete ϕ^4 model $a = 44.2$. Six curves are shown in the system, corresponding to system thickness $N = 1, 2, 3, 4, 5$, and 6 layers. The extreme left-hand curve corresponds to a single-layer system, and the extreme right-hand curve corresponds to a six-layer system.

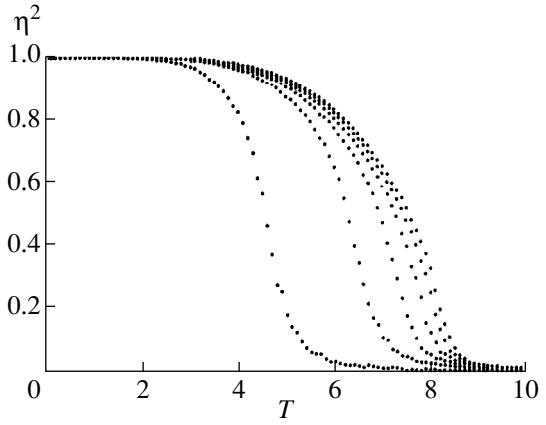


Fig. 6. Temperature dependences of the squared order parameter obtained by Monte Carlo simulation for layered systems. The parameter of the discrete ϕ^4 model $a = 4000$. Six curves are shown in the system, corresponding to system thickness $N = 1, 2, 3, 4, 5$, and 6 layers. The extreme left-hand curve corresponds to a single-layer system, and the extreme right-hand curve corresponds to a six-layer system.

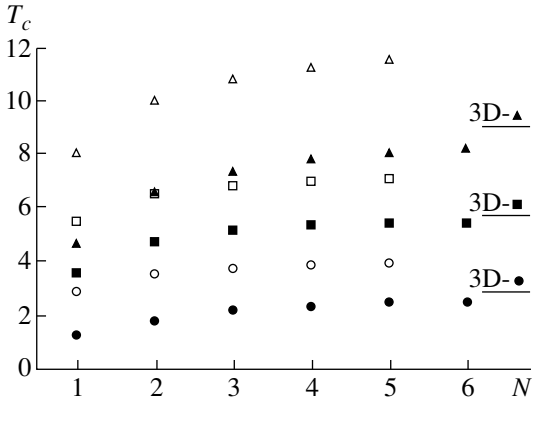


Fig. 7. Phase transition temperatures versus the system thickness. The curves obtained from the curves presented in Figs. 4–6 at the points of the minimum of the derivative (filled symbols) and from a calculation in the mean-field approximation (open symbols). The circles correspond to the model parameter $a = 0.98$, the squares correspond to $a = 44.2$, and the triangles correspond to $a = 4000$. The horizontal lines indicate that 3D limits for each of the three model parameters, taken from [33].

As shown above, the case $a = 4000$ corresponds to order–disorder transitions. This assertion can be substantiated as follows. In the first place the transition temperature for $a = 4000$ is identical to the known value of the transition temperature for the 2D Ising model (Fig. 2). In the second place, as noted in the introduction, it is sufficient to investigate layered systems in detail for the Ising model [14–16]. This makes it possible to compare with our results. Figure 8 shows three dependences

$$f(N) = 1 - \frac{T_c(N)}{T_c(\text{3D, Ising})}$$

two of which are taken from the literature [14, 39] while the third corresponds to data for $a = 4000$ from Fig. 7. It is evident that the results for $a = 4000$ agree, to a high degree of accuracy, with the results obtained previously for the Ising model. This makes it possible to talk in this case about order–disorder type phase transitions.

3. ANALYTIC APPROXIMATIONS

The mean-field and independent-modes approximations on the basis of the discrete ϕ^4 model are considered as the analytic approximations. These analytic schemes have been well studied for the 3D and 2D cases. Extensions of these analytic approximations to the case of layered systems, which we studied, in the discrete ϕ^4 model will be used below.

3.1. Mean-Field Approximation

In the mean-field approximation the real interaction of the particles is described by the mean field acting on each particle. This field is independent of the positions of the particle and its immediate environment. The crux of this analytic scheme is that the fluctuations of the resulting force acting on each particle are neglected [2].

We shall consider this approximation for the discrete ϕ^4 model. In this case the introduction of a mean field consists of replacing the sums over n and n' in the

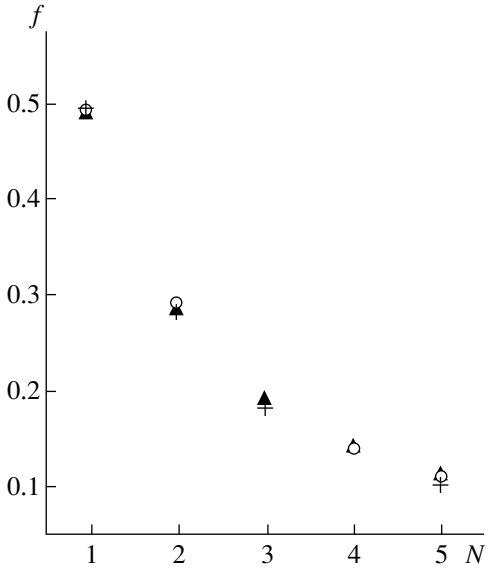


Fig. 8. The function $f(N) = 1 - \frac{T_c(N)}{T_c(\text{3D, Ising})}$ versus the thickness of the system N. Circles are data from [14], crosses are data from [39], and triangles are results obtained in the present work for the parameter a of the discrete ϕ^4 model, equal to 4000.

last term of the potential energy (2) by the expression $E\sum_n x_n$, where

$$E = 4d\eta \tag{4}$$

is the mean-field considered (d is the dimension of the system). The order parameter η can be calculated as an ordinary thermodynamic average:

$$\eta = \langle x_n \rangle = \frac{\int x \exp(-V(E, x)/T)dx}{\int \exp(-V(E, x)/T)dx}, \tag{5}$$

where

$$V(E, x) = \left(2d - \frac{a}{2}\right)x^2 + \frac{a}{4}x^4 - Ex. \tag{6}$$

The equations (4)–(6) determine the self-consistent system, whose solution is the temperature dependence of the order parameter $\eta(T)$ or the temperature dependence of the mean field $E(T)$. The phase transition temperature is determined at the point T_c, where $\eta(T_c) = 0$. Since near a phase-transition point the order parameter η assumes its minimum values, the transition temperature T_c can be obtained by expanding the right-hand side of Eq. (5) in a series in this small parameter. The dependences of the transition temperature on the parameter a $T_c(a)$ for the discrete ϕ^4 model are monotonically increasing curves with two asymptotic values for large and small parameters a as a function of temperature: 12 and 4 for the 3D case and 8 and 2.5 for the 2D case. These results can be compared with the exact results from the table. In the mean-field approximation a phase transition exists even in the one-dimensional case, while in reality the fluctuations destroy the order in any one-dimensional systems with short-range action [1]. It should be noted that the mean-field approximation, for example, for the 3D case of the discrete ϕ^4 model, repeats qualitatively correctly the dependence $T_c(a)$. However, this approximation gives a systematic identical quantitative error for all values of the parameter a.

Various variants of the mean-field approximation have been used in the literature for layered systems [15, 16]. We shall employ the following approximation for multilayer structures. We shall assume that the order parameter depends on the number N of layers. Then according to Eq. (5) the quantity $\langle x_n(n)\rangle$ can be expressed as

$$\langle x_n(n)\rangle = \frac{\int x \exp[-V(x_n(n), \langle x_n(n)\rangle, \langle x_n(n-1)\rangle, \langle x_n(n+1)\rangle)/T]dx}{\int \exp[-V(x_n(n), \langle x_n(n)\rangle, \langle x_n(n-1)\rangle, \langle x_n(n+1)\rangle)/T]dx}, \tag{7}$$

where the potential energy $V(x_n(n), \langle x_n(n)\rangle, \langle x_n(n-1)\rangle, \langle x_n(n+1)\rangle)$ is determined by the expression

$$\begin{aligned} &V(x_n(n), \langle x_n(n)\rangle, \langle x_n(n-1)\rangle, \langle x_n(n+1)\rangle) \\ &= \left(\alpha - \frac{a}{2}\right)x_n^2 + \frac{a}{4}x_n^4 - 8\langle x_n(n)\rangle x_n \\ &- 2(\langle x_n(n-1)\rangle + \langle x_n(n+1)\rangle)x_n. \end{aligned} \tag{8}$$

The constant α is determined according to the number of nearest neighbors for the layer with number n:

$$\alpha = \begin{cases} 5, & n = 1, N \\ 6, & n \neq 1, N. \end{cases} \tag{9}$$

Here it is assumed that $\langle x_n(n)\rangle = 0$, if $n > N$ or $n < 1$. For definiteness we shall write the system of equations for a lattice consisting of three layers:

$$\langle x_n(1)\rangle = \frac{\int x \exp[-V(x_n, \langle x_n(1)\rangle, \langle x_n(2)\rangle)/T]dx}{\int \exp[-V(x_n, \langle x_n(1)\rangle, \langle x_n(2)\rangle)/T]dx},$$

$$\begin{aligned} &\langle x_n(2)\rangle \\ &= \frac{\int x \exp[-V(x_n, \langle x_n(2)\rangle, \langle x_n(3)\rangle, \langle x_n(1)\rangle)/T]dx}{\int \exp[-V(x_n, \langle x_n(2)\rangle, \langle x_n(3)\rangle, \langle x_n(1)\rangle)/T]dx}, \end{aligned} \tag{10}$$

$$\langle x_n(3)\rangle = \frac{\int x \exp[-V(x_n, \langle x_n(3)\rangle, \langle x_n(2)\rangle)/T]dx}{\int \exp[-V(x_n, \langle x_n(3)\rangle, \langle x_n(2)\rangle)/T]dx},$$

where

$$\begin{aligned} V(x_n, \langle x_n(1)\rangle, \langle x_n(2)\rangle) &= \left(5 - \frac{a}{2}\right)x_n^2 + \frac{a}{4}x_n^4 \\ &- 8\langle x_n(1)\rangle x_n - 2\langle x_n(2)\rangle x_n, \\ V(x_n, \langle x_n(1)\rangle, \langle x_n(2)\rangle, \langle x_n(3)\rangle) &= \left(6 - \frac{a}{2}\right)x_n^2 + \frac{a}{4}x_n^4 \\ &- 8\langle x_n(2)\rangle x_n - 2\langle x_n(1)\rangle x_n - 2\langle x_n(3)\rangle x_n, \\ V(x_n, \langle x_n(2)\rangle, \langle x_n(3)\rangle) &= \left(5 - \frac{a}{2}\right)x_n^2 + \frac{a}{4}x_n^4 \\ &- 8\langle x_n(3)\rangle x_n - 2\langle x_n(2)\rangle x_n. \end{aligned} \tag{11}$$

If only the transition temperature needs to be determined, then since $\langle x_n(n)\rangle$ are small near the phase-transition point, the right-hand sides of the equations of the system can be expanded in a series in this small parameter. We shall determine the phase transition temperature by equation to zero the determinant of the system of equations obtained in this manner. From the N roots of this determinant we choose only the one root for which each component $\langle x_n(n)\rangle$ is zero. For three layers this determinant has the form

$$\begin{vmatrix} P(T) & R(T) & 0 \\ Q(T) & S(T) & Q(T) \\ 0 & R(T) & P(T) \end{vmatrix} = 0, \tag{12}$$

where

$$\begin{aligned} P(T) &= \int f_1(x)\left(1-\frac{8x^2}{T}\right)dx, \\ R(T) &= \int f_1(x)\left(-\frac{2x^2}{T}\right)dx, \\ Q(T) &= \int f_2(x)\left(-\frac{2x^2}{T}\right)dx, \\ S(T) &= \int f_2(x)\left(1-\frac{8x^2}{T}\right)dx. \end{aligned} \tag{13}$$

The functions $f_1(x)$ and $f_2(x)$ are defined as

$$f_1(x) = \exp\left\{-\left[\left(5-\frac{a}{2}\right)x^2+\frac{a}{4}x^4\right]\frac{1}{T}\right\},$$

$$f_2(x) = \exp\left\{-\left[\left(6-\frac{a}{2}\right)x^2+\frac{a}{4}x^4\right]\frac{1}{T}\right\}.$$

The results of calculations in the mean-field approximation are presented in Fig. 7. The phase-transition temperatures T_c were obtained as a function of the parameter a of the discrete ϕ^4 model for systems consisting of several layers (from one layer (the 2D case) up to five layers). The dependences of the transition temperatures on the thickness of the system were obtained from the indicated curves for three different values of the model parameter a, taken from numerical experiments (Fig. 7). The curves presented show that the mean-field approximation (open symbols) have the same qualitative behavior as the curves obtained from numerical experiments (filled symbols).

3.2. Independent-Modes Approximation

In the independent-modes approximation the system in the Fourier representation is treated as a collection of independent harmonic oscillators. It is assumed that a phase transition in this system occurs at the moment when oscillators with zero characteristic frequency appear (condensation of elementary excitations in the soft mode occurs) [22]. For this reason the independent-modes approximation is used primarily to describe displacement-type phase transitions. If this approximation is studied with respect to the discrete ϕ^4 model, then in this case only one point, corresponding to a displacement-type transition, $a \longrightarrow +0$, can be described correctly. The larger the value of the parameter a, the more strongly correlated the modes become and in consequence this approximation gives a worse description of the behavior of the system. In this sense the independent-modes approximation is related to Landau's phenomenological model.

To obtain an equation to determine the temperature dependence of the order parameter the basic equations of the independent-modes approximation must be used [22]:

$$\omega^2(q) = -1 + 4dF(q) + 3(I+\eta^2), \tag{14}$$

$$I = \frac{1}{(2\pi)^d}\int d^dq\frac{T}{\omega^2(q)}, \tag{15}$$

$$\eta^3 + (3I-1)\eta = 0. \tag{16}$$

Here Eq. (14) determines the spectrum of the system, Eq. (15) determines the average value of the squared Fourier-transform coordinate for the classical harmonic oscillator, Eq. (16) is the coupling equation for the order parameter and the quantity I, and

$$F(q) = 1-\frac{1}{d}\sum_{i=1}^{d}\cos q_i. \tag{17}$$

The equation for determining the dependence of the temperature on the order parameter η is obtained from the equations described above and has the following form:

$$T = \frac{1-\eta^2}{\dfrac{3}{(2\pi)^d}\displaystyle\int\left[4\left(d-\sum_{i=1}^{d}\cos q_i\right)+2\alpha\eta^2\right]^{-1}d^dq}. \tag{18}$$

In the 3D case a transition is observed at the point where the order parameter $\eta = 0$. In this case the transition temperature is 2.64 (see table). In the 2D case the integral in the denominator with $\eta = 0$ diverges logarithmically. This divergence indicates that the independent-modes approximation is inapplicable near the transition point, where correlations in the phonon modes with large occupation numbers can no longer be neglected.

The independent modes approximation can be extended to the case of layered systems as follows. In a layer of finite thickness the dispersion law for phonons is identical to the law for the 3D case, but the momentum component q_z, perpendicular to the plane of the layers, can assume only a discrete set of values (the z-axis is

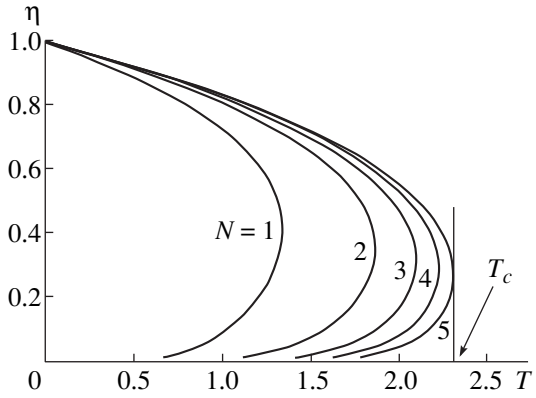


Fig. 9. Temperature dependences of the order parameter in the independent-modes approximation. The parameter of the discrete ϕ^4 model $a = 0.98$. Five curves corresponding to system thickness $N = 1, 2, 3, 4$, and 5 layers are shown. The transition temperatures are determined at the points of the maximum of the function $T(\eta)$. The transition temperature for a system consisting of five layers is marked in the figure.

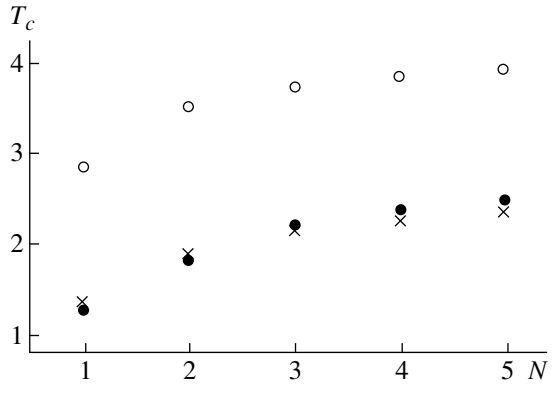


Fig. 10. Phase transition temperature versus the system thickness. The parameter of the discrete ϕ^4 model $a = 0.98$. Filled points are results of numerical simulation (the data were obtained from Fig. 7), open points are mean-field approximation (data obtained from Fig. 7), and crosses are independent modes approximation (data obtained from Fig. 9). The independent-modes approximation describes to within 5% the results obtained by numerical simulation.

normal to the surface). In a system of N layers with free boundary conditions $q_z = k\pi/N$, $k = 0, 1, \ldots, N-1$. These values can be obtained by solving the problem of a one-dimensional chain of atoms with a harmonic interaction with nearest neighbors [40]. Now, besides integrating over the components q_x and q_y, summation over q_z with weight $1/N$ must be performed in the formulas of the independent-modes approximation. Specifically, Eq. (18) becomes

$$T = (1-\eta^2)\left\{\frac{3}{(2\pi)^2 N}\sum_{q_z}\int\int\left[4\left(3-\cos q_z - \cos q_y - \cos q_z\right) + 2\alpha\eta^2\right]^{-1} dq_x dq_y\right\}^{-1}. \quad (19)$$

In the limiting cases $N = 1$ and $N \longrightarrow \infty$ we obtain the formulas for the 2D and 3D systems.

The temperature dependences of the order parameter for systems with various thicknesses can be determined from this equation. The investigations were performed for parameters $a < 1$. The results for the case $a = 0.98$ are presented in Fig. 9. As one can see in Fig. 9, the order parameter does not vanish for any nonzero temperature or for any finite N. As already mentioned, such behavior is a consequence of the inapplicability of the independent-modes approximation near the transition point. Actually, the same difficulties also arise for 3D systems, where the independent-modes approximation erroneously predicts a first-order transition [22]. However, as $a \longrightarrow +0$ the fluctuation region where the independent-modes approximation does not work is small, which makes it possible to determine the transition temperature. We determined the transition temperature for layered systems as the maximum of $T(\eta)$ (Fig. 9). The results of numerical experiments and calculations in the mean-field and independent-modes approximations for the parameter $a = 0.98$ are compared in Fig. 10. In this case the independent-modes approximation describes the results of the numerical experiments not only qualitatively but also quantitatively.

4. CONCLUSIONS

In the present paper the properties of phase transitions in two-dimensional and layered systems were investigated on the basis of the discrete ϕ^4 model by numerical and analytic methods. The discrete ϕ^4 model makes it possible to introduce a single model parameter a, which determines the type of phase transition (from displacement type transitions with $a \longrightarrow +0$ to order–disorder type transitions with $a \longrightarrow +\infty$). Two quantities were varied to study the properties of phase transitions in the discrete ϕ^4 model: the parameter a and the thickness N of the system.

A discrete ϕ^4 model was investigated, using the Monte Carlo method, for a wide range of values of the parameter a, specifically, for a 2D lattice. In addition the behavior of the order parameter as a function of temperature was investigated for three characteristic values of the parameters of the model a, equal to 0.98, 4000, and 44.2, corresponding to displacement and order–disorder type phase transitions as well as an intermediate case, respectively. The transition temperatures T_c were found, from the results obtained, as functions of the thickness N of the system (from one to six layers) for the three cases described, and the dependence of the transition temperature on the parameter a for the 2D system was also found. On the basis of these results, which are presented in Fig. 7, it can be concluded that as the number of layers increases, the tran-

sition temperature increases monotonically, and the system is strongly sensitive to the addition of one or two layers to the first layer, after which the dependence slowly approaches the 3D asymptotics.

The numerical results obtained were compared with calculations performed in the mean-field and independent-modes approximations. It was shown that for a wide range of values of the parameter a the mean-field approximation qualitatively correctly describes the dependences of the phase transition temperatures T_c on the thickness N of the system, which were obtained in the numerical experiments, though it overestimates T_c by 30–100%. For the parameter $a = 0.98$ the results of the numerical calculations were compared with the calculation performed in the independent-modes approximation. In this case the independent-modes approximation describes the Monte-Carlo curves not only qualitatively but also quantitatively, to within about 5%. However, as the parameter a increases, the increase in the correlations in the phonon modes makes this approximation inapplicable.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980), Part 1.
2. B. A. Levanyuk and A. P. Strukov, *Physical Foundations of Ferroelectric Phenomena in Crystals* (Nauka, Moscow, 1995).
3. D. Heiman and S. Ushioda, Phys. Rev. B **17**, 3616 (1978).
4. H. J. Bakker, S. Hunsche, and H. Kurz, Phys. Rev. B **48**, 9331 (1993).
5. Ken-ichi Noba and Yosuke Kayanuma, Phys. Rev. B **60**, 4418 (1999).
6. P. H. M. van Loosdrecht, M. M. Maior, S. B. Molnar, *et al.*, Phys. Rev. B **48**, 6014 (1993).
7. Yu. M. Vysochanskii, V. A. Stephanovich, A. A. Molnar, *et al.*, Phys. Rev. B **58**, 9119 (1998).
8. J. Hlinka, T. Janssen, and V. Dvorak, J. Phys.: Condens. Matter **11**, 3209 (1999).
9. W. Dürr, M. Taborelli, O. Paul, *et al.*, Phys. Rev. Lett. **62**, 206 (1989).
10. Yi. Li and K. Baberschke, Phys. Rev. Lett. **68**, 1208 (1992).
11. C. M. Schneider, P. Bressler, P. Schuster, *et al.*, Phys. Rev. Lett. **64**, 1059 (1990).
12. D. L. Lin, Hao Che, and Yu Xia, Phys. Rev. A **46**, 1805 (1992).
13. D. L. Lin, Hao Che, Wei Lai, *et al.*, Phys. Rev. E **49**, 2155 (1994).
14. P. Schilbe, S. Siebentritt, and K.-H. Rieder, Phys. Lett. A **216**, 20 (1996).
15. A. Lipowski and M. Suzuki, Physica A (Amsterdam) **198**, 227 (1993).
16. P. J. Jensen, H. Dreysse, and K. H. Bennemann, Surf. Sci. **270**, 627 (1992).
17. A. V. Bune, V. M. Fridkin, St. Ducharme, *et al.*, Nature **391**, 874 (1998).
18. St. Ducharme, A. V. Bune, L. M. Blinov, *et al.*, Phys. Rev. B **57**, 25 (1998).
19. S. Palto, L. Blinov, A. Bune, *et al.*, Ferroelectrics **184**, 127 (1996).
20. O. A. Aktsipetrov, T. V. Misuryaev, T. V. Murzina, *et al.*, Opt. Lett. **25**, 411 (2000).
21. O. A. Aktsipetrov, L. M. Blinov, V. M. Fridkin, *et al.*, Surf. Sci. **454**, 1016 (2000).
22. A. D. Bruce and R. A. Cowley, *Structural Phase Transition* (Taylor and Francis, London, 1981).
23. T. Schneider and E. Stoll, Phys. Rev. B **17**, 1302 (1978).
24. S. Aubry, J. Chem. Phys. **62**, 3217 (1975).
25. A. M. Ferrenberg and D. P. Landau, Phys. Rev. B **44**, 5081 (1991).
26. A. P. Giddy, M. T. Dove, and V. Heine, J. Phys.: Condens. Matter **1**, 8327 (1989).
27. T. Radescu, I. Etxebarria, and J. M. Pérez-Mato, J. Phys.: Condens. Matter **7**, 585 (1995).
28. W. Loinaz and R. S. Willey, Phys. Rev. D **58**, 076003 (1998).
29. F. Cooper, B. Freedman, and D. Preston, Nucl. Phys. B **210**, 210 (1982).
30. R. Toral and A. Chakrabarti, Phys. Rev. B **42**, 2445 (1990).
31. W. Janke and T. Sauer, J. Stat. Phys. **78**, 759 (1995).
32. W. Janke and T. Sauer, Phys. Rev. E **49**, 3475 (1994).
33. A. N. Rubtsov, J. Hlinka, and T. Janssen, Phys. Rev. E **61**, 126 (2000).
34. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* (Nauka, Moscow, 1982, 2nd ed.; Pergamon, Oxford, 1979).
35. K. Binder and P. C. Hohenberg, Phys. Rev. B **9**, 2194 (1974).
36. G. A. T. Allan, Phys. Rev. B **1**, 352 (1970).
37. M. E. Fisher, *Critical Phenomena*, Ed. by M. S. Green (Academic, New York, 1971).
38. C. Domb and J. L. Lebowitz, *Phase Transitions and Critical Phenomena* (Academic, London, 1983).
39. K. Binder, Thin Solid Films **20**, 367 (1974).
40. M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Theory of Waves* (Nauka, Moscow, 1979).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1213–1220.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1402–1410.

SOLIDS
Structure

Reversible Magnetization of a Thin Superconducting Film with Radiation Defects

A. N. Artemov

Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, Donetsk, 340114 Ukraine
e-mail: artemov@kinetic.ac.donetsk.ua

Received May 4, 2000

Abstract—The thermodynamic potential of a system of Peierls vortices in a thin superconducting film, containing radiation defects, in a perpendicular external magnetic field is calculated. The equilibrium temperature dependences of the densities of free vortices and vortices trapped by defects are found in the mean-field approximation for various magnitudes of the external field. It is shown that the equilibrium magnetization of a thin superconducting film exhibits the same features that were observed experimentally in the reversible magnetization of high-temperature superconductors. An asymptotic expression is obtained for the difference of the magnetizations of perfect and irradiated films. According to this expression, the difference depends on the pinning energy of a vortex on a defect and the density of defects. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The layered structure of high-temperature superconductors is responsible for their quasi-two-dimensional behavior, which is manifested in the thermodynamic and kinetic properties. According to modern ideas this behavior is due to the possibility of the existence of two-dimensional (2D) topological defects—magnetic vortices—in superconducting layers of these materials [1–3]. A system of such vortices can occupy individual and collective states.

In an individual state the magnetic energy of a vortex diverges logarithmically and is cut off by the size of the sample. In a superconductor such 2D vortices can occupy only bound states with finite energy. In the absence of an external field these are dipoles consisting of two vortices, arising as thermal fluctuations, with oppositely directed fluxes. In an external field they can also be present in the form of bound states forming Abrikosov vortices.

However, if the system contains a finite density of free 2D vortices, then as a result of Debye screening the energy of a vortex becomes finite and depends on the density of free vortices [4]. In this state the 2D vortices in a superconductor can form a gas of free vortices. The behavior of a system of vortices in these states is radically different. A transition of the system between states can occur as a Berezinskii–Kosterlitz–Thouless (BKT) phase transition [5, 6].

The reversible magnetization of high-temperature superconductors is one manifestation of the quasi-two-dimensional behavior of a system of vortices in a collective state. Experiments on the magnetization of high-temperature superconductors [7–11] have shown that in sufficiently high magnetic fields and at sufficiently high temperatures the magnetic moment of a sample is a function of the state. It has been observed that at a certain temperature T^* the magnetic moment of the sample does not depend on the magnitude of the external magnetic field. The temperature dependences of the magnetic moment $M(T)$ overlap at this point, and below it the magnetic moment of the sample is essentially a linear function of the temperature.

The theoretical interpretation of the experimental results [12–14] was based on the concept of 2D vortices. As a result, the basic reason for this behavior of high-temperature superconductors became clear. In ordinary superconductors the magnetic flux is produced by a system of Abrikosov vortices which form a rigid lattice. The configurational energy of the system of vortices makes the main contribution to the free energy and hence to the magnetization. The entropy part of the free energy is negligibly small.

The behavior of magnetic fluxes in high-temperature superconductors differs from that of ordinary superconductors. Neutron scattering experiments [15] and measurements of the Josephson plasma resonance [16, 17] show that in magnetic fields above 0.05 T there is no correlation between the positions of the vortices in different layers and the vortices do not form vortex lines. This suggests that the magnetic flux is produced by a high-entropy gas of 2D vortices. The configurational and entropy contributions of the system of vortices enter in the magnetization with different signs. At the point T^*, close to or equal to the BKT transition temperature, these contributions compensate one another, which is why the magnetic field is independent of the external field.

A thin superconducting film also demonstrates quasi-two-dimensional properties. In the Peierls limit [18], where the film thickness d is much less than the

1063-7761/00/9106-1213$20.00

London length λ, the behavior of a system of vortices in a film is essentially identical to that of 2D vortices. The main difference is due to the existence of a large but finite effective Peierls length $\Lambda = 2\lambda^2/d$. This length limits the logarithmic growth of the self-energy of a Peierls vortex, which is finite. A system of Peierls vortices, just as a system of 2D vortices, can occupy individual as well as collective states. However, since the energy of Peierls vortices is finite, they can occupy an individual state in a superconductor not only bound in dipoles but also as free vortices. Here the individual and collective states differ not qualitatively, as in a system of 2D vortices, but only quantitatively. In an individual state the Debye screening length $\delta > \Lambda$ and the energy of a vortex is essentially independent of the density of the free vortices. In the opposite limit the energy of a Peierls vortex is cut off by the Debye screening length δ and the behavior of a system of Peierls vortices is identical to that of a system of 2D vortices in a collective state.

Hence it follows that at sufficiently high temperatures and in a sufficiently strong magnetic field, when a system of Peierls vortices is in a collective state, its reversible magnetization should exhibit that same characteristic features as in a system of 2D vortices. The calculation of the thermodynamic properties of a system of Peierls vortices in an external magnetic field [19] has shown this to be true.

The penetration of radiation defects into high-temperature superconductors strongly affects the process of reversible magnetization of the samples [20, 21]. A characteristic value of the magnetic field B_ϕ appears; it is comparable to the defect density N_d/S in the sample, $B_\phi = N_d\phi_0/S$. In weak ($H_e \ll B_\phi$) and strong ($H_e \gg B_\phi$) external magnetic fields the magnetization decreases approximately as the logarithm of the external field with increasing field. In intermediate fields, near B_ϕ, the magnetization no longer decreases but increases, switching from the low- to the high-field asymptotic behavior.

The main reasons for such behavior of the magnetization were correctly understood in the works cited above. Since the energy of a vortex trapped by a defect is less than the energy of a free defect by the amount of the energy of a normal nucleus or, equivalently, by the pinning energy, these are the first positions to be filled when the sample is magnetized. Consequently, the low-field part of the magnetization curves is due primarily to trapped vortices. The equilibrium number of trapped vortices in a sample with a low magnetic field is always larger than the equilibrium number of free vortices in a perfect sample. This circumstance is responsible for the difference of the magnetizations of the irradiated and perfect samples, which depends on the pinning energy. In fields $H_e \gg B_\phi$ most defects are filled and the free vortices determine magnetization.

A more systematic theoretical analysis of the equilibrium magnetization of layered superconductors is performed in [22]. Here the entropy of the system and the correlation length of the system are taken into account in the London approximation. A drawback of this work, in our view, is that only vortices oriented in the direction of the field are taken into account. This approach rules out taking into account the dissociation of vortex dipoles under the influence of thermal fluctuations and limits the applicability of the theory to low temperatures, where the concentration of the vortices oriented in a direction opposite to the field is negligibly small.

In the present paper a theory that takes account of the influence of radiation defects on the equilibrium magnetization of a thin superconducting film is proposed. It takes account of the configurational energy and entropy of a system of free and Peierls vortices and Peierls vortices trapped by defects as well as the magnetic flux energy of the samples. The free energy of a system of Peierls vortices in an external magnetic field is obtained in Section 2. The equilibrium density of free and trapped vortices in an external field are examined in Section 3. The equilibrium magnetization of a Peierls sample with radiation defects is discussed in Section 4.

2. FREE ENERGY OF A SYSTEM OF PEIERLS VORTICES IN A MAGNETIC FIELD

Let us consider a thin superconducting film with area S, containing N_d radiation defects occupying random points $\mathbf{X}_\alpha$. We represent the defects as sections of the order of the superconductor coherence length ξ in size, where superconductivity is completely suppressed. At finite temperatures such a system can contain $N_{+(-)}$ free vortices and $N_{t+(t-)}$ vortices trapped by defects with different directions of the magnetic flux. The vortices can be created and annihilated in pairs, and they can also be trapped by and become detached from defects. To calculate the thermodynamic properties of such a system it is necessary to estimate its grand partition function.

The partition function of a system of vortices in a zero external magnetic field has been written down and estimated in [23]. In this case the chemical potentials of the vortices are zero and the equilibrium properties of the system are obtained by minimizing the free energy. However, if the system is in a nonzero magnetic field, then the equilibrium values of the chemical potentials are different from zero, and the equilibrium values of the densities of the vortices are determined by the conditions that the thermodynamic potential G is minimum:

$$\Xi = \exp\{-\beta G\} = \sum_{N_\pm, N_{t\pm}} \exp\{-\beta F + \beta(\mu_+ N_+ + \mu_- N_- + \mu_{t+} N_{t+} + \mu_{t-} N_{t-})\}$$

$$= \sum_{N_\pm, N_{t\pm}} \exp\{\beta(\mu_+ N_+ + \mu_- N_- + \mu_{t+} N_{t+} + \mu_{t-} N_{t-})\} \times (N_+! N_-! N_{t+}! N_{t-})^{-1} \frac{\exp\{-(N_+ + N_-)\beta E_0\}}{(N_d - N_{t+} - N_{t-})!} \times \mathrm{Sp} \exp\left\{-\frac{\beta}{2}\sum_{i,j} U(\mathbf{x}_i - \mathbf{x}_j) - \beta \sum_{i,\alpha} U(\mathbf{x}_i - \mathbf{X}_\alpha) - \frac{\beta}{2}\sum_{\alpha,\gamma} U(\mathbf{X}_\alpha - \mathbf{X}_\gamma)\right\}. \quad (1)$$

The operation Sp denotes summation over all possible configurations of the system

$$\mathrm{Sp} = \left(\int \frac{d\mathbf{x}}{\pi\xi^2}\right)^{N_+ + N_-} \left(\sum_{\mathbf{x}}\right)^{N_{t+} + N_{t-}}.$$

Here F is the free energy of a system with a fixed number of vortices, μ_i are the chemical potentials of the vortices, $\beta = 1/T$, E_0 is the energy of the normal nucleus of a vortex equal to the pinning energy on radiation defects, and $U(\mathbf{X})$ is the interaction energy of Peierls vortices, free vortices, and vortices trapped by defects. The summation in the exponential extends independently over both indices, and terms with the same indices give the self-energy of the vortices. The quantity $N_d - N_{t+} - N_{t-}$ is the number of free defects. We neglect their interaction energy with vortices when estimating the configuration part of the energy, since it decreases rapidly with distance [24], but we include their contribution to the entropy of the system.

Before estimating the free energy of the system, we shall determine the equilibrium values of the chemical potentials. In a superconductor the free energy of a system of Peierls vortices in an external magnetic field is a function of temperature T and the induction, determined in this case as $B = S^{-1}\int d\mathbf{x} h(\mathbf{x}) = \phi_0(N_+ + N_{t+} - N_- - N_{t-})/S$. The equilibrium of a system of vortices in an external field occurs when the Gibbs potential G reaches a minimum value. Ordinarily, it is determined as a function of the internal field H_i, which is identical to the external field H_e only for a long cylinder:

$$G = F - \frac{BH_i}{4\pi}.$$

We cannot use such a geometry to calculate the magnetization of a thin film. Consequently, we shall consider a film as a flattened ellipsoid of revolution of height d and diameter $D \gg d$. In a magnetic field directed along the axis of rotation, the external field H_e is related with H_i and the induction B as follows [25]: $(1 - n)H_i = H_e - nB$, where $n = 1 - d/\pi D$ is a demagnetizing factor. Using a Legendre transformation we can obtain the Gibbs energy for a system as a function of the external field H_e [26]:

$$G = F - \frac{1}{4\pi}\frac{BH_e}{1-n}. \quad (2)$$

The equilibrium distribution of magnetic fluxes in the system is determined by minimizing the Gibbs energy [27]:

$$\frac{\partial F}{\partial B} = \frac{1}{4\pi}\frac{H_e}{1-n}.$$

Since we wish to determine the potential G, following from Eq. (1), and noting that $\partial/\partial N_\pm = \partial/\partial N_{t\pm} = \pm(\phi_0/S)\partial/\partial B$, we find a relation between the equilibrium chemical potentials of the system and the external field:

$$\frac{\partial F}{\partial N_\pm} = \frac{\partial F}{\partial N_{t\pm}} = \mu_\pm = \mu_{t\pm} = \pm\frac{1}{4\pi}\frac{\phi_0}{S}\frac{H_e}{1-n}. \quad (3)$$

We shall now estimate the free energy of a system of vortices. We expand the logarithm of the partition function (1) in a series. This series contains only irreducibly connected diagrams [28]. The coordinates of the vortices correspond to the vertices of the diagrams, and the bonds correspond to the interaction energy. Integration over the entire area S of the film is performed over the coordinates of the free vortices. If vortices trapped by defects are present in the system, then the diagrams corresponding to them and the configuration energy now depend on their coordinates.

We shall assume that the radiation defects in the samples occupy fixed random positions. The fluctuations of the coordinates of the defects, associated with thermal vibrations of the crystal lattice, can be neglected, since they are negligibly small compared with the fluctuations of the order parameter of the superconductor, i.e., the fluctuations of the coordinates of the free vortices. Thus, the free energy must be averaged over an ensemble of defect configurations. We shall assume that the defects are distributed over the entire area of the film with equal probability. We shall perform the averaging by integrating the free energy over the coordinates of all vortices, with weight $1/S$, trapped by defects. The result is that the trapped vortices make the same contribution to the average free energy as do the fluctuating free vortices. The only difference is that the energy of the nuclei of the trapped vortices is zero.

This result seems to us to be entirely natural, since the statistical averaging over an ensemble of fixed defect configurations is equivalent to time averaging [29], which must be performed in the case of fluctuating vortices. Only the probability of realization of various configurations is important. Here we adopted the hypothesis that the configurations are equally probable for free vortices and for defects, which is what determines the answer obtained.

We now separate the part of the free energy of the system that is due to the nonzero total topological charge associated with the fact that in an external magnetic field the numbers of vortices oriented parallel (N_+, N_{t+}) and antiparallel (N_-, N_{t-}) to the field are not equal to one another. This is the energy of the magnetic flux in a superconductor. It does not depend on the configuration of the flux and can be taken into account similarly to the way the positive background is taken into account in a one-component model of a plasma [28]:

$$\frac{1}{2}\sum_{i\neq j}\int\frac{d\mathbf{r}_i}{V}\frac{d\mathbf{r}_j}{V}U(\mathbf{x}_i-\mathbf{x}_j) = V\frac{\phi_0^2(n_+ + n_{t+} - n_- - n_{t-})^2}{8\pi(1-n)} = V\frac{B^2}{8\pi}\frac{1}{1-n}. \quad (4)$$

The summation in this expression extends independently over the coordinates of all free and trapped vortices, and the integration extends over the entire volume V of the sample.

It is now easy to estimate the configuration energy. The sequence of ring diagrams [28] turns out to be the leading sequence for the long-range interaction of particles, which the interaction of Peierls vortices is [26]. This expression corresponds precisely to the phenomenological Debye–Hückel approach and is actually its microscopic substantiation. The microscopic approach to the calculation of the free energy makes it possible to decrease the number of phenomenological parameters in the theory, which in our case is the energy E_0 of the nucleus of a vortex and the coherence length $\xi(T)$. In the Debye–Hückel approach we required another relation between the probabilities of dissociation and recombination of vortex dipoles. The expression for the configuration energy of a system of Peierls vortices in the ring approximation has been obtained in [26]. We need only substitute, together with the number of free vortices $N_\pm$, the sum of the numbers of free and trapped vortices $N_\pm + N_{t\pm}$.

Thus, we now have all the constituent parts of the free energy of a system of Peierls vortices, specifically, the entropy part, obtained from the factorial factors in Eq. (1), the configuration energy, and the magnetic-flux energy (4). Introducing the dimensionless parameters $n_i = \pi\xi^2 N_i/S$ and $f = \pi\xi^2\beta F/S$ we obtain

$$\begin{aligned} f &= n_+(\ln n_+ - 1) + n_-(\ln n_- - 1) + n_{t+}(\ln n_{t+} - 1) \\ &+ n_{t-}(\ln n_{t-} - 1) + (n_d - n_{t+} - n_{t-}) \\ &\times(\ln(n_d - n_{t+} - n_{t-}) - 1) + p(n_+ + n_- + n_{t+} + n_{t-}) \\ &\times[1 - \ln 8p(n_+ + n_- + n_{t+} + n_{t-})] \\ &+ \frac{1}{4\Lambda^2}W[32p\Lambda^2(n_+ + n_- + n_{t+} + n_{t-})] \\ &+ p(n_+ + n_-)e_0 + 4p\kappa^2\frac{(n_+ + n_{t+} - n_- - n_{t-})^2}{1-n}. \end{aligned} \quad (5)$$

All lengths are given in units of the coherence length ξ, $p = \phi_0^2/16\pi^2\Lambda T$, $\Lambda = \lambda^2/2d$ is the effective Peierls length, $\kappa = \lambda/\xi$, $pe_0 = \beta E_0$, and

$$W[x] = \frac{1}{2}\ln\frac{x}{4} + \sqrt{|1-x|}\begin{cases}\arctan\dfrac{1}{\sqrt{x-1}} - \dfrac{\pi}{2}, & x \geq 1 \\ \dfrac{1}{2}\ln\dfrac{1+\sqrt{1-x}}{1-\sqrt{1-x}}, & x \leq 1.\end{cases}$$

We have used dimensionless quantities to make the quite complicated expressions for the free energy and the equations of equilibrium more compact. We shall require notation for the external magnetic field $h_e = \pi\xi^2 H_e/\phi_0$ and the magnetic moment of the sample $m = \pi\xi^2 4\pi M/\phi_0$. With this method of obtaining dimensionless quantities, except for the characteristic temperature dependence, the temperature dependence of the coherence length $\xi(T)$ is obtained. Sometimes this is convenient, just as for the concentration of vortices when the dimensionless concentration determined in this manner is simply the fraction of positions occupied by vortices among all possible positions at a given temperature. For the magnetic field or moment this is not so convenient. Consequently, in such cases, when the dependence $\xi(T)$ is ruled out, we shall label the corresponding quantity with the index 0. Thus, the quantity h_e depends on temperature and h_{e0} does not.

3. THE EQUILIBRIUM STATE

The equilibrium concentrations of vortices in the system are given by the conditions (3), which lead to the system of equations

$$\begin{aligned} &\ln n_\pm - p\ln 8p(n_+ + n_- + n_{t+} + n_{t-}) \\ &+ 8pW'[32p\Lambda^2(n_+ + n_- + n_{t+} + n_{t-})] \\ &\pm 8p\kappa^2\frac{(n_+ + n_{t+} - n_- - n_{t-} - h_e)}{1-n} + pe_0 = 0, \\ &\ln n_{t\pm} - p\ln 8p(n_+ + n_- + n_{t+} + n_{t-}) \\ &+ 8pW'[32p\Lambda^2(n_+ + n_- + n_{t+} + n_{t-})] \\ &\pm 8p\kappa^2\frac{(n_+ + n_{t+} - n_- - n_{t-} - h_e)}{1-n} \\ &- \ln(n_d - n_{t+} - n_{t-}) = 0. \end{aligned} \quad (6)$$

Subtracting the second equation of the system (6) from first one and the fourth equation from the second one, we obtain a relation between the concentrations of vortices oriented parallel and antiparallel to the field:

$$\ln\frac{n_+}{n_-} = \ln\frac{n_{t+}}{n_{t-}} = -16p\kappa^2 m \quad (7)$$

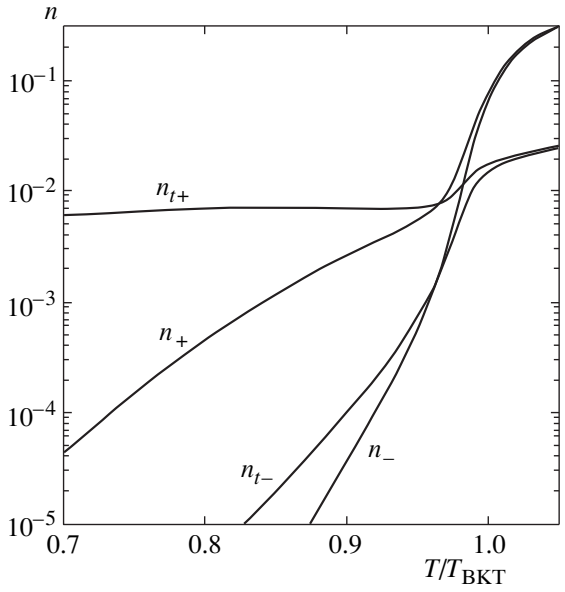


Fig. 1. Equilibrium concentrations of free vortices and vortices trapped by defects versus the temperature in an external field $h_{e0} = 4 \times 10^{-3}$ with defect density $n_{d0} = 0.013$.

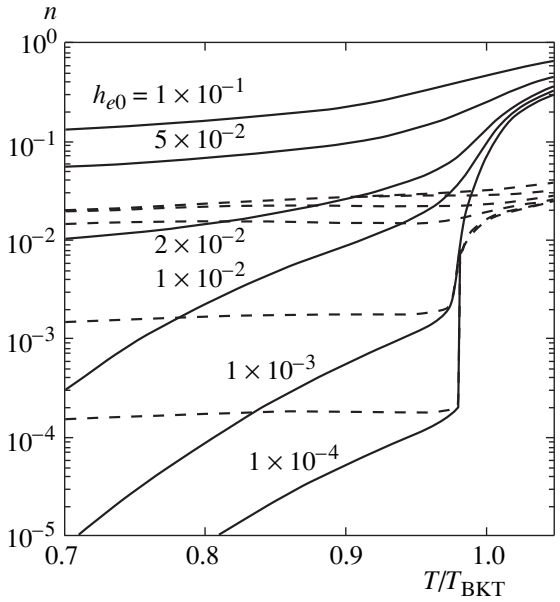


Fig. 2. Equilibrium concentrations of free vortices and defect-trapped vortices oriented along the field versus the temperature for various values of the external magnetic field with defect density $n_{d0} = 0.013$.

with the magnetic moment of the sample determined by the expression

$$4\pi M = \frac{B - H_e}{1 - n}.$$

The ratios of the densities of free and trapped vortices are equal because of the two subsystems are in chemical contact with one another. From the conditions of equilibrium (6) follows another relation between the concentrations of the free and trapped vortices

$$n_{t\pm} = \frac{n_\pm n_d}{n_+ + n_- + e^{-pe_0}}. \qquad (8)$$

This expression shows that the concentration $n_{t\pm}$ of trapped vortices tends to 0, if the total concentration of free vortices decreases more rapidly than the exponential function in the denominator. This behavior is characteristic for an individual state of a system of vortices, when the Debye screening interaction length δ is greater than the effective length Λ and obtains at comparatively low temperatures and for low magnitudes of the external field, falling outside the range of reversible magnetization.

However, if the concentration of free vortices decreases more slowly than e^{-pe_0}, then the trapped vortices tend to fill all defects. This is observed in a collective state of a system of vortices ($\delta < \Lambda$), which makes the process of magnetizing a superconductor reversible. In what follows, we shall be interested only in this range of temperatures and fields.

The equilibrium concentrations of the vortices as a function of temperature for various values of the external magnetic field, which were obtained by solving numerically the system of Eqs. (6), are shown in Figs. 1 and 2. The temperature behavior of the concentrations of the free vortices and the vortices trapped by defects is qualitatively different in low ($H_e \ll B_\phi$) and high ($H_e \gg B_\phi$) external magnetic fields. Figure 1 shows the temperature dependences of the vortex concentrations in a system with a low magnetic field $h_{e0} = 4 \times 10^{-4}$ with defect density $n_{d0} = 0.013$. For temperatures $T > T_{BKT}$ the concentration of vortices increases rapidly as a result of the Kosterlitz–Thouless instability. The concentration of free vortices here is higher than the concentration of trapped vortices, since the latter is limited by the number of defects.

In the temperature range $T < T_{BKT}$ the concentration of free vortices at first exceeds that of the trapped vortices. As the temperature decreases, the thermal fluctuations resulting in the dissociation of vortex dipoles become weaker and the concentration of free vortices rapidly becomes less than that of the trapped vortices. At these temperatures the concentration of vortices trapped by defects changes very little. These vortices are primarily the ones that form the flux of the magnetic field through the sample in weak external fields, since this state of the system of vortices is energetically more favorable.

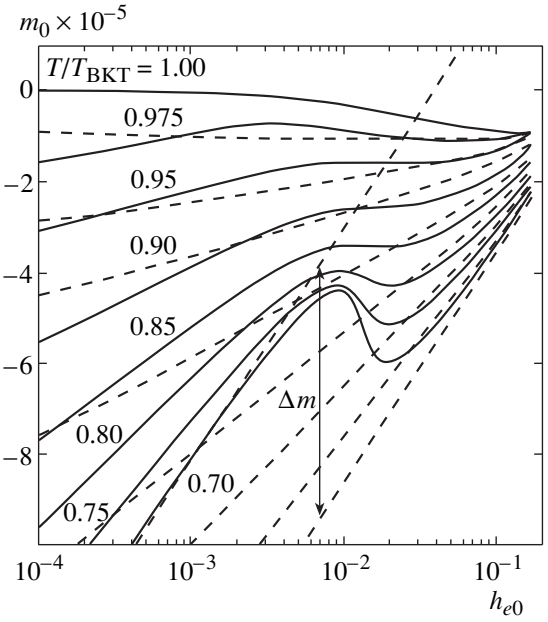


Fig. 3. Magnetic moment of a perfect (dashed lines) and irradiated (solid lines) sample of a thin superconducting film versus the external magnetic field for various temperatures.

In high external fields the trapped vortices can no longer give an equilibrium magnetic flux, and consequently the concentration of free vortices increases. In Fig. 2 the solid lines show the concentration of free vortices and the dashed lines show the concentration of trapped vortices, oriented along the field. It is clearly seen how the ratio between the concentrations of the free and trapped vortices changes as the external field increases. The concentrations of vortices oriented opposite to the field, for $T < T_{BKT}$, is always much less than the concentration of vortices oriented along the field, as shown in Fig. 1.

The numerical calculations were performed for a model film with the parameters $\kappa = 50$, $\Lambda/\xi = 500$, $n = 0.99$, and $e_0 = 3$ and with defect density $n_{d0} = 0.013$.

4. EQUILIBRIUM MAGNETIZATION OF THE SAMPLE

The equilibrium dependences of the magnetic moment of a sample on the external magnetic field for various values of the temperature are shown in Fig. 3 (solid lines) for a sample with defect density $n_{d0} = 0.013$ and for an unirradiated sample (dashed lines). All the basic features of the magnetization curves, which are observed in experiments on single-crystal samples of high-temperature superconductors, are clearly seen here [20, 21]. In weak, $H_e \ll B_\phi$ ($h_e \ll n_d$), and strong, $H_e \gg B_\phi$ ($h_e \gg n_d$), fields the magnetization of the sample decreases as $\ln H_e$, which is especially clearly seen for low temperatures, $T < 0.85T_{BKT}$. For $H_e \approx B_\phi$ the decrease of the magnetization is replaced by an increase of magnetization, and then once again by a decrease, switching from the low- to the high-field asymptotic behavior.

This behavior of the magnetization is due to the coexistence of two subsystems of magnetic vortices in the sample—free and trapped. In low fields and at low temperatures, when the total number of vortices is less than the number of defects, the concentration of trapped vortices is higher than that of free vortices, since their energy is lower. Consequently, the magnetic flux in this sample is primarily due to the vortices trapped by defects. In high fields, $H_e \gg B_\phi$, virtually all defects are occupied and most of the flux is due to free vortices. Since their energy is higher, it is more difficult for free vortices to enter the sample than the trapped vortices. As a result, the magnetization of the sample is greater by the amount ΔM than the magnetization that the trapped vortices would give.

The quantity ΔM can be easily estimated for low temperatures, where $n_+ \gg n_-$. Using this smallness, an asymptotic expression, in good agreement with the numerical solution for temperatures $T < 0.85T_{BKT}$, can be obtained:

$$8p\kappa^2 m = \ln\left(1 + \frac{n_d}{n_+ + e^{-pe_0}}\right) - (1-p)\ln n_+ + p(\ln 8p - e_0). \tag{9}$$

The asymptotic value of the concentration of free vortices oriented along the field is given by the expression

$$n_+ = -\frac{n_d + e^{-pe_0} - h_e}{2} + \sqrt{\frac{(n_d + e^{-pe_0} - h_e)^2}{4} + h_e e^{-pe_0}}. \tag{10}$$

The concentrations n_- and $n_{t\pm}$ can be estimated according to Eqs. (7) and (8). To estimate ΔM, we shall find the high- and low-field asymptotic forms of the expression (9). For large values of the external field, $H_e \gg B_\phi$, additional smallnesses $h_e \gg n_d$, e^{-pe_0}, appear in the problem and the expression (10) simplifies to $n_+ \approx h_e$. In this limit the magnetization has the form

$$8p\kappa^2 m_{high} \approx -(1-p)\ln h_e + p(\ln 8p - e_0). \tag{11}$$

Physically, the expression for the magnetization has been simplified because we neglected the contribution of trapped vortices to the magnetic flux through the sample. At the temperature $T = 0.7T_{BKT}$ the dependence $m(h_e)$, determined by the expression (11), is identical to the expression found numerically for a perfect film (the lower dashed line in Fig. 3). It lies somewhat below the magnetization curve of the irradiated sample, since in

the latter a part of the magnetic flux is due to the trapped vortices.

In the opposite limit of a weak external field it can be assumed that $h_d \gg n_e$, e^{-pe_0}, and $n_+ \approx h_e e^{-pe_0}/n_d$. In this limit the expression (9) becomes

$$8p\kappa^2 m_{\text{low}} \approx -(1-p)\ln h_e + p\ln 8p + \ln n_d. \qquad (12)$$

In this case we neglected the contribution of free vortices to the magnetization. The dependence $m(h_e)$ in this limit is shown in the figure for the temperature $T = 0.7T_{\text{BKT}}$ by the straight dashed line, close to the numerical solution for low values of the external field.

Thus, we obtain for the difference of the magnetizations the expression

$$\Delta m = m_{\text{low}} - m_{\text{high}} = \frac{1}{8p\kappa^2}(pe_0 + \ln n_d). \qquad (13)$$

As follows from the expression obtained, the difference of the magnetization is determined by two factors: the density of defects and the energy of the normal nucleus of a vortex or the pinning energy. In dimensional notation this expression is

$$\Delta M = \frac{1}{\phi_0 d}\left[U_{\text{pin}} + T\ln\left(\frac{B_\phi}{2H_{c2}}\right)\right],$$

where $U_{\text{pin}} = E_0$ is the pinning energy of a vortex on a defect and $H_{c2} = \phi_0/(2\pi\xi^2)$ is the second critical field.

As temperature increases, thermal fluctuations begin to make a substantial contribution to the behavior of the system of vortices. Dissociation and recombination of vortex dipoles are activated in the system. The equilibrium concentrations of the vortices oriented parallel and antiparallel increase, which causes the dependences $m(h_e)$ to deviate from the asymptotic form (9).

As one can see in Fig. 3, the temperature T^*, at which the magnetization does not depend on the external magnetic field, for a perfect sample is somewhat less than the temperature T_{BKT}. For a sample with radiation defects this possibility obtains for $T = T_{\text{BKT}}$ only for a weak external fields, $H_e \ll B_\phi$, as in the experiments of [20, 21].

5. CONCLUSIONS

In the present paper the effect of radiation defects on the equilibrium reversible magnetization of a thin superconducting field containing radiation defects was investigated. The nature of the reversible behavior of magnetic fluxes is not discussed here. In order to understand it, the thermodynamic properties of a system of vortices in the superconductor, which interact with the edges of the sample and with the Meissner current, must be studied. This is a much more complicated problem.

As follows from everything presented in this paper, the reversible magnetization of a Peierls film with radiation defects possesses the same features as the magnetization of high-temperature superconductors. The theoretical study of such an object is interesting from two points of view. In the first place this is a real object, which can be investigated experimentally. On the other hand it is a theoretical model, which is simpler than a layered superconductor, since problems due to Josephson coupling between layers and to the contribution of bound states of 2D vortices in different layers to the thermodynamics do not arise. In our view the effect of these two factors on the magnetic properties of layered superconductors is small and will not produce any qualitative differences from the properties of a thin film.

ACKNOWLEDGMENTS

This work was supported by the Ukrainian Ministry of Education and Science (project 2M/71-2000).

REFERENCES

1. S. N. Artemenko and A. N. Kruglov, Phys. Lett. A **143**, 485 (1990).
2. A. I. Buzdin and D. Feinberg, J. Phys. (Paris) **51**, 1971 (1990).
3. J. R. Clem, Phys. Rev. B **43**, 7837 (1991).
4. P. Minnhagen, Rev. Mod. Phys. **59**, 1001 (1987).
5. V. L. Berezinskii, Zh. Éksp. Teor. Fiz. **61**, 1144 (1971) [Sov. Phys. JETP **34**, 610 (1972)].
6. J. M. Kosterlitz and D. G. Thouless, J. Phys. C **6**, 1181 (1973).
7. P. H. Kes, C. J. van der Beek, M. P. Maley, *et al.*, Phys. Rev. Lett. **67**, 2383 (1991).
8. K. Kadowaki, Physica C (Amsterdam) **185–189**, 2249 (1991).
9. Q. Li, M. Suenaga, D. K. Finnemore, *et al.*, Phys. Rev. B **46**, 3195 (1992).
10. F. Zuo, D. Vacaru, H. M. Duan, and A. M. Hermann, Phys. Rev. B **47**, 8327 (1993).
11. Sung-Ik Lee, Mun-Seog Kim, Jim-Nam Park, *et al.*, Czech. J. Phys. **46**, 1781 (1996).
12. L. N. Bulaevskiĭ, M. Ledvij, and V. G. Kogan, Phys. Rev. Lett. **68**, 3773 (1992).
13. Z. Tešanović, L. Xing, L. N. Bulaevskii, *et al.*, Phys. Rev. Lett. **69**, 3563 (1992).
14. A. Yu. Martynovich and A. N. Artemov, Phys. Rev. B **56**, 14 827 (1997).
15. R. Cubitt and E. M. Forgan, Nature **365**, 407 (1993).
16. O. K. C. Tsui, N. P. Ong, and J. B. Peterson, Phys. Rev. Lett. **76**, 819 (1996).
17. L. N. Bulaevskii, V. L. Pokrovsky, and M. P. Maley, Phys. Rev. Lett. **76**, 1719 (1996).
18. J. Pearl, Appl. Phys. Lett. **5**, 65 (1964).
19. A. N. Artemov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 460 (1998) [JETP Lett. **68**, 492 (1998)].

20. Qiang Li, Y. Fukumoto, Y. Zhu, *et al.*, Phys. Rev. B **54**, R788 (1996).

21. C. J. van der Beek, M. Konczykowski, T. W. Li, *et al.*, Phys. Rev. B **54**, R792 (1996).

22. L. N. Bulaevskii, V. M. Vinokur, and M. P. Maley, Phys. Rev. Lett. **77**, 936 (1996).

23. A. N. Artemov, Zh. Éksp. Teor. Fiz. **116**, 1081 (1999) [JETP **89**, 577 (1999)].

24. J. R. Chen and T. Y. Tseng, Phys. Rev. B **54**, 502 (1996).

25. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).

26. V. N. Ryzhov and E. E. Tareyeva, Phys. Rev. B **48**, 12907 (1993).

27. P. G. de Gennes, *Superconductivity of Metals and Alloys* (Benjamin, New York, 1966; Mir, Moscow, 1968).

28. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978).

29. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, Oxford, 1980), Part 1.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1221–1229.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1411–1420.

SOLIDS
Structure

Antiferromagnetic Resonance Spectrum in $LaMnO_3$: Interrelation of the Orbital Structure and the Magnetic Properties

L. E. Gonchar'*, A. E. Nikiforov, and S. E. Popov

Ural State University, Yekaterinburg, 620083 Russia
**e-mail: lyudmila.gonchar@usu.ru*

Received May 10, 2000

Abstract—The relation between the orbital ordering and magnetic structure of the crystal $LaMnO_3$ is investigated. The dependence of the exchange parameters on the angle Φ of the orbital structure is determined. When the isotropic exchange interaction and the single-ion anisotropy, which depends on the angle Φ and the rotational distortions, are introduced into the spin Hamiltonian, a four-sublattice structure (A_X, F_Y, G_Z) is obtained with orientation of the magnetic moments of the sublattices near the long axis of the orthorhombic cell of the crystal in the basal plane of the crystal ($A_X \gg G_Z, F_Y$). The effect of a magnetic field on the magnetic structure and the antiferromagnetic resonance spectrum are investigated taking account of the nonequivalent, anisotropic, orbitally-dependent g tensors. The spin-flop and spin-flip transition fields are calculated. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The discovery of colossal magnetoresistance in compounds based on $LaMnO_3$ rekindled interest in this compound and led to a large number of investigations devoted to this crystal.

For $LaMnO_3$ strong coupling of the spin, charge, and orbital degrees of freedom is characteristic. It is also important to determine the characteristic features of this relationship for the parent compound $LaMnO_3$ in order to explain the properties of doped compounds.

The existence of orbital structure in $LaMnO_3$ is due to the nontrivial magnetic properties of this compound. The orbital structure of Jahn–Teller compounds has been discussed in a number of theoretical works [1, 2]. Recent investigations of resonance X-ray scattering [3] have confirmed the existence of orbital structure in lanthanum manganite. The appearance of orbital order was attributed to the exchange mechanism and electronic-vibrational interaction. The first model is now commonly used in theoretical works [4, 5] in connection with the idea of coupling of the orbital and magnetic subsystems and the possibility of calculating the coupled spectra of the orbital and magnetic excitations. However, the second model seems to us to be more justified because the electronic-vibrational coupling is quite strong in this crystal. The experimental data of [3], which confirm that the orbital structure is also preserved after magnetic order is disrupted, right up to the transition into the *O** phase without Jahn–Teller distortions, also attest to this.

The magnetic structure of lanthanum manganite has still not been unequivocally explained. The experimental investigations [6, 7] have shown that below the temperature T_N = 140 K the crystal possesses *A*-type magnetic structure (see Fig. 1) with the magnetic moments of the sublattices oriented along the long orthorhombic axis in the basal plane *O*' of the crystal structure (symmetry group *Pnma*). A qualitative analysis based on the Goodenough–Kanamori rules [1] makes it possible to explain the ferromagnetic sign of the isotropic exchange interaction in the basal plane; the orientation of the magnetic moments along the orthorhombic axis was attributed to the single-ion anisotropy of rank 4 [6]. The weak ferromagnetic moment along the *Y*-axis [6] (group *Pnma*) is usually attributed to antisymmetric exchange.

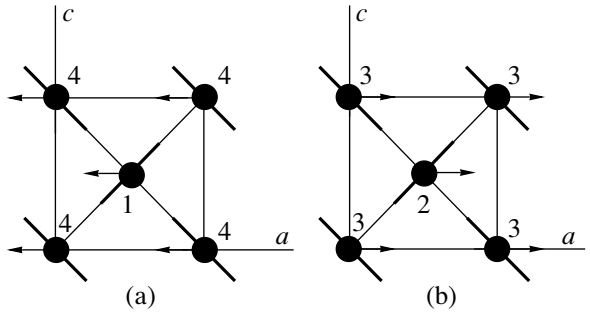


Fig. 1. Orbital and magnetic structures of pure manganite in two neighboring planes along the *Y*-axis (parts a and b). The oxygen and lanthanum ions are omitted. The arrows indicate the main directions of the magnetic sublattices. The thick lines denote the orientations of the e_g orbitals. The numbers enumerate the magnetic sublattices.

There now have appeared a number of neutron-scattering investigations [8, 9] of the spectrum of spin waves in $LaMnO_3$. These experiments make it possible to determine the exchange integrals in this compound; this makes it possible to construct the dependence of the exchange parameters on the orbital structure.

Until recently it was believed that it is impossible to investigate the resonance properties of $LaMnO_3$ because the spectrum contains a large gap [8, 9]. Nonetheless, measurements of the temperature [10] and field [11] dependences of the antiferromagnetic resonance frequencies in the submillimeter range have appeared recently; this made it possible to determine certain characteristic features of the magnetic excitations spectrum, which are accessible to measurement using neutron scattering.

On account of their high accuracy resonance methods for investigating magnets have always been regarded as more informative than diffraction methods. In the Jahn–Teller magnets, where the magnetic interaction is strongly anisotropic and there is a large number of different weak effects, which is associated with the orbital structure of these compounds, resonance makes it possible to investigate the characteristics of the magnetic structure to a high degree of accuracy. Advances in the technique of resonance investigations have made it possible to investigate magnets whose spectrum contains a large gap. Thus, it is now possible to study the resonance properties of $LaMnO_3$, and this is certainly of interest.

From the standpoint of a theoretical explanation of the spin-wave and antiferromagnetic resonance spectra, the authors of [8–11] adhere to phenomenological models on the basis of the experimentally known magnetic structure of $LaMnO_3$. They employ a two-sublattive magnet model with single-ion anisotropy, which aligns the magnetic moments precisely along the *X*-axis (group *Pnma*). This is sufficient to describe the dispersion properties of spin waves [8, 9]. Antisymmetric exchange with vector directed along the *Z*-axis and isotropic Zeeman interaction were added in order to describe the field dependence of the antiferromagnetic resonance [10, 11]. This model described the dependences well, but it does not agree with symmetry considerations and cannot be explained from the microscopic standpoint.

In the present work calculations of the magnetic structure and antiferromagnetic resonance spectrum were performed taking account of the orbitally dependent magnetic interactions. Our model consists of the following:

strong electron-vibrational interaction is responsible for the cooperative distortions Q_θ and Q_ε of the oxygen octahedra and for establishment of orbital ordering in the crystal;

the crystal possesses a regular structure and is completely stoichiometric;

the only form of the exchange interaction in the crystal is superexchange;

the exchange and Zeeman interactions, as well as the single-ion anisotropy, depend on the orbital structure; and

anisotropic exchange is neglected; the entire dependence on the rotational distortions is contained in the exchange and Zeeman interactions as well as in the single-ion anisotropy.

This model assumes that the magnetic structure is determined by the orbital ordering. Consequently, the dependences of all magnetic interactions on the parameter characterizing the orbital structure are constructed. In this model the spin-flop and spin-flip transitions, as well as the behavior of the magnetization of the sublattices in an external magnetic field, were calculated. In addition, the characteristic features of the field dependences of the antiferromagnetic resonance frequencies and the total magnetization were explained.

2. CRYSTAL AND ORBITAL STRUCTURES

The $LaMnO_3$ crystal possesses a distorted perovskite structure [6, 7, 12, 13]. At low temperatures this substance is in an orthorhombic *O*' phase, and at higher temperatures or with weak doping it transforms into a quasicubic phase *O**. The space group *Pnma* (D_{2h}^{16}) corresponds to both phases.

The orthorhombic distortion of an ideal perovskite lattice can be represented in the form of three basic distortions [14, 15].

(1) A distortion of the *R* type is a rotation of the oxygen octahedra around the $[110]_p$ axis of an ideal perovskite with cell doubling along all three axes ($\{k_{13}\}\tau_9$ (C_1C_10) in Kovalev's notation [16] or ($\varphi\varphi0$) in the notation of [17]). The angle $\varphi = 12°$ [18].

(2) The *M* type distortion is a rotation of oxygen octahedra around the $[001]_p$ axis with cell doubling along two axes ($\{k_{11}\}\tau_3$ ($00C_2$) in the notations of [16] or (00ψ) according to [17]). The angle $\psi = 10°$ [18].

(3) The ε type distortion (Q_ε) describes an *e*-type deformation of the crystal octahedron with doubling along two axes ($\{k_{11}\}\tau_5$in the notations of [16] and with the choice of the ray $[1/2\ 1/2\ 0]_p$). Distortions of this type are illustrated in Fig. 2.

The corresponding basic distortions are accompanied by adjustment of the lattice.

The orbitally degenerate ground state ^5E of the manganese ions is split as a result of the cooperative Jahn–Teller effect. The orbital state with the wave function

$$\psi_n = \sin\frac{\Phi_n}{2}\varphi_{n\theta} + \cos\frac{\Phi_n}{2}\varphi_{n\varepsilon}, \qquad (1)$$

where (φ_θ, φ_ε) are the basis functions for the *E* level and the angle Φ satisfies (see Fig. 1)

$$\Phi_1 = \Phi_2 = -\Phi_3 = \Phi_4 = \Phi, \qquad (2)$$

is established on each manganese ion. The value of this angle is expressed in terms of the distortion in the crystal:

$$\sin\Phi = \frac{Q_\varepsilon}{\sqrt{Q_\theta^2 + Q_\varepsilon^2}}, \quad \cos\Phi = \frac{Q_\theta}{\sqrt{Q_\theta^2 + Q_\varepsilon^2}}. \qquad (3)$$

Here $\Phi = 107°$ [18].

The orbital structure of manganite is shown in Fig. 1.

3. ORBITAL DEPENDENCE OF THE MAGNETIC INTERACTIONS

3.1. Isotropic Exchange

Taking account of the metal–ligand distances r_b and r_{ac} and the superexchange coupling angles φ_b and φ_{ac}, the authors of [19] obtained for pairs of interaction ions along the z_p axis and in the basal plane the dependences for the parameters of the isotropic exchange of manganese ions in a manganite crystal with structure D_{2h}^{16} on the orbital structure ($J_{12} = J_{34} = J_b$, $J_{14} = J_{23} = J_{ac}$ in accordance with the notations used in Fig. 1):

$$J_b = \frac{J_0\cos^2\varphi_b}{r_b^{10}}F_b(\Phi), \qquad J_{ac} = \frac{J_0\cos\varphi_{ac}}{r_{ac}^{10}}F_{ac}(\Phi), \qquad (4)$$

where

$$F_b(\Phi) = [1 + 2\alpha\cos\Phi + \beta\cos^2\Phi], \qquad F_{ac}(\Phi) = \left[1 - \alpha\cos\Phi + \beta\left(\cos^2\Phi - \frac{3}{4}\right)\right], \qquad (5)$$

J_0, α, and β are parameters that depend on the type of magnetic ions in a pair and the type of intermediate ion (they can be determined from the experimental data), and Φ is the angle of the Jahn–Teller distortions. The parameters in the relation (5) are determined in [19]: $J_0 = 1.24 \times 10^3$ meV Å^{10}, $\alpha = 0.9$, $\beta = 4.9$. The dependence of the exchange integrals on the Jahn–Teller distortion angle is presented in Fig. 3. The structural parameters for $LaMnO_3$ were taken from [18] and are $r_b = 1.96$ Å (the distance between the manganese and oxygen ions along the *b* axis), $r_{ac} = 2.04$ Å (the average of the length of the long and short distances between the manganese and oxygen ions in the *ac* plane), $\varphi_b = 155.1°$ (the superexchange binding angle along the *b*-axis), and $\varphi_{ac} = 153.8°$ (the superexchange bond angle in the *ac* plane).

It is obvious that the presence of a ferromagnetic exchange interaction in the basal plane is due to the presence of a definite orbital structure, since the orbitally independent part of the exchange is antiferromagnetic.

A characteristic feature of the dependence (5) is a possible change in magnitude and sign of the exchange parameter in the *ac* plane (the minus sign denotes a ferromagnetic interaction), which is due to the change in the magnetic structure with the crystal structure remaining unchanged (D_{2h}^{16}). This distinguishes the Jahn–Teller compounds from other magnetic dielectrics.

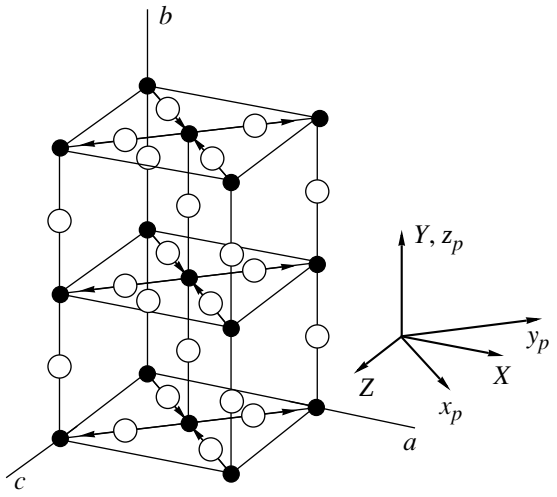


Fig. 2. Fragment of a $LaMnO_3$ cell with ε-type distortion. The La^{3+} ions are omitted; (*X*, *Y*, *Z*) is orthorhobic coordinate system; (x_p, y_p, z_p) is quasicubic coordinate system.

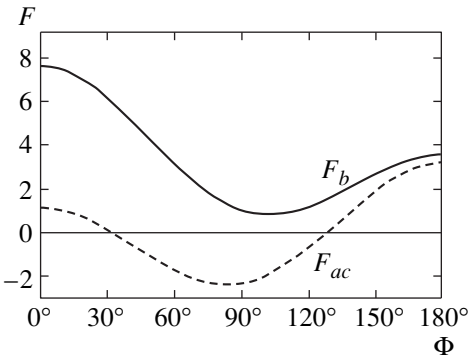


Fig. 3. Angular part of the exchange integrals along the *b* axis (F_b) and in the basal plane (F_{ac}) versus the Jahn–Teller distortion angle (angle of orbital structure) Φ.

3.2. Single-Ion Anisotropy and the Zeeman Interaction

The single-ion anisotropy appears in second-order perturbation theory with respect to the spin-orbital interaction. From microscopic calculations using perturbation theory $P \approx -\zeta_{3d}^2/(12\Delta)$, where ζ_{3d} is the spin-orbit interaction constant and Δ is the energy gap between the ^5E ground state and the $^3T_1$ excited state of the manganese ion. Taking $\zeta_{3d} = 17.1$ meV [6] and $\Delta = 169.5$ meV gives $P = -0.15$ meV.

An expression for the single-ion anisotropy is obtained in [19] in the local axes of the oxygen octahedra. It depends on the orbital ordering angle:

$$H_{an}^{(n)} = D_n S_{nz_1}^2 + E_n(S_{nx_1}^2 - S_{ny_1}^2), \quad (6)$$

where

$$D_n = 3P\cos\Phi_n, \quad (7)$$

$$E_n = \sqrt{3}P\sin\Phi_n. \quad (8)$$

The values of D_n are positive ($\cos\Phi_i < 0$, $P < 0$) and are the same for all magnetic ions in a cell, and E_n changes sign on switching to a neighboring magnetic ion in the basal plane.

The quantity P can be determined according to the magnitude of the gap for the lower branch of the spin waves $\Delta E = 2.7$ meV [8]: $P = -0.13$ meV [19]. For values of ΔE from [10, 11], $P \approx -0.1$ meV. In subsequent calculations we shall employ this value. In the section devoted to the antiferromagnetic resonance spectrum we shall examine the determination of the constant P in greater detail.

On account of the Jahn–Teller effect the g tensors of neighboring ions are not equivalent [3]. This greatly complicates the magnetic structure of the compound in an external nonzero magnetic field. The components of the g tensors also depend on the angle Φ. We shall write down the Zeeman interaction operator for the corresponding symmetry [20]:

$$\hat{H}_{\text{Zeem}}^{(n)} = g_1\mu_B(\mathbf{S}_n \cdot \mathbf{H}) + \frac{1}{2}g_2\mu_B[(3S_{nz_p}H_{z_p} - \mathbf{S}_n \cdot \mathbf{H})X_{n\theta} + \sqrt{3}(S_{nx_p}H_{x_p} - S_{ny_p}H_{y_p})X_{n\varepsilon}], \quad (9)$$

where $X_{\theta,\varepsilon}$ are orbital operators which act in the space of the functions (φ_θ, φ_ε) of the ground-state E term:

$$X_\theta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X_\varepsilon = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (10)$$

We now average it over the wave functions (1) and obtain an effective spin Hamiltonian in local axes of the oxygen octahedra

$$\hat{H}_{\text{Zeem}} = -\mu_B\sum_i(\mathbf{H} \cdot g^i \cdot \mathbf{S}_i), \quad (11)$$

where

$$g^1 = g^2 = \begin{pmatrix} g_1 + g_2\cos\left(\Phi - \frac{2\pi}{3}\right) & 0 & 0 \\ 0 & g_1 + g_2\cos\left(\Phi + \frac{2\pi}{3}\right) & 0 \\ 0 & 0 & g_1 + g_2\cos\Phi \end{pmatrix},$$
$$g^3 = g^4 = \begin{pmatrix} g_1 + g_2\cos\left(\Phi + \frac{2\pi}{3}\right) & 0 & 0 \\ 0 & g_1 + g_2\cos\left(\Phi - \frac{2\pi}{3}\right) & 0 \\ 0 & 0 & g_1 + g_2\cos\Phi \end{pmatrix}. \quad (12)$$

The parameters g_1 and g_2 were estimated in second-order perturbation theory to be $g_1 = 1.89$ and $g_2 = 0.047$ (for $\zeta_{3d} = 17.1$ meV and $\Delta = 169.5$ meV [6]).

4. RESULTS AND DISCUSSION

The following Hamiltonian was used to calculate the magnetic properties of lanthanum manganite:

$$\hat{H} = J_b \sum_{\substack{m>n \\ \text{interplane}}} (\mathbf{S}_m \cdot \mathbf{S}_n) + J_{ac} \sum_{\substack{m>n \\ \text{in plane}}} (\mathbf{S}_m \cdot \mathbf{S}_n) + \sum_n \hat{H}_{an}^{(n)} + \sum_n \hat{H}_{\text{Zeem}}^{(n)}. \quad (13)$$

The rotational distortions of oxygen octahedra are taken into account not only in the exchange Hamiltonian (4) but also in all other terms. For spin variables, a transformation is made in the terms of the single-ion anisotropy and Zeeman interaction from a local coordinate system tied to axes of the octahedron into a general orthorhombic system using the matrix

$$M(\varphi, \psi) = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos^2\frac{\varphi}{2} & \sin^2\frac{\varphi}{2} & -\frac{\sin\varphi}{\sqrt{2}} \\ -\sin^2\frac{\varphi}{2} & \cos^2\frac{\varphi}{2} & \frac{\sin\varphi}{\sqrt{2}} \\ -\frac{\sin\varphi}{\sqrt{2}} & \frac{\sin\varphi}{\sqrt{2}} & \cos\varphi \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}. \quad (14)$$

The correspondence between the rotation angles and the orbital structure in a primitive magnetic cell is given as follows:

$$\begin{aligned} &\mathrm{Mn}_1(\varphi, \psi, \Phi), \mathrm{Mn}_2(-\varphi, \psi, \Phi), \\ &\mathrm{Mn}_3(\varphi, -\psi, -\Phi), \mathrm{Mn}_4(-\varphi, -\psi, -\Phi). \end{aligned} \quad (15)$$

After appropriate transformations the Hamiltonian (13) can be rewritten as the energy of the magnetic subsystem in terms of the basal vectors of the magnetic structure for four sublattices (in the nearest-neighbors approximation):

$$\begin{aligned} E = \frac{N}{4}\{ & J_b(\mathbf{F}^2 - \mathbf{G}^2 - \mathbf{A}^2 + \mathbf{C}^2) \\ & + 2J_{ac}(\mathbf{F}^2 - \mathbf{G}^2 + \mathbf{A}^2 - \mathbf{C}^2) + a(F_X^2 + G_X^2 + A_X^2 + C_X^2) \\ & + b(F_Y^2 + G_Y^2 + A_Y^2 + C_Y^2) + c(F_Z^2 + G_Z^2 + A_Z^2 + C_Z^2) \\ & + d(F_X A_Y + F_Y A_X + G_X C_Y + G_Y C_X) \\ & + e(F_X C_Z + F_Z C_X + G_X A_Z + A_X G_Z) \\ & + f(F_Y G_Z + F_Z G_Y + A_Z C_Y + A_Y C_Z) \\ & + 4\mu_B[H_X(g_{11}F_X + g_{21}A_X) \\ & + g_{22}H_Y F_Y + H_Z(g_{13}F_Z + g_{23}A_Z)]\}, \end{aligned} \quad (16)$$

where N is the number of magnetic ions in the sublattice, $\mathbf{F} = \mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3 + \mathbf{S}_4$, $\mathbf{G} = \mathbf{S}_1 - \mathbf{S}_2 + \mathbf{S}_3 - \mathbf{S}_4$, $\mathbf{A} = \mathbf{S}_1 - \mathbf{S}_2 - \mathbf{S}_3 + \mathbf{S}_4$, and $\mathbf{C} = \mathbf{S}_1 + \mathbf{S}_2 - \mathbf{S}_3 - \mathbf{S}_4$ are the basal vectors of the four-sublattice magnetic structure [21], J_b and J_{ac} are the exchange integrals in $LaMnO_3$ (J_b = 1.16 meV, J_{ac} = –1.66 meV [9], J_b = 1.21 MeV, J_{ac} = –1.67 meV [8]), and the single-ion anisotropy and Zeeman interaction parameters are expressed in terms of the structural parameters as follows:

$$\begin{aligned} a &= \sqrt{3}P\sin\Phi\sin 2\psi, \\ b &= P(3\cos\Phi\cos^2\varphi - \sqrt{3}\sin\Phi\sin^2\varphi\sin 2\psi), \\ c &= P(3\cos\Phi\sin^2\varphi - \sqrt{3}\sin\Phi\cos^2\varphi\sin 2\psi), \\ d &= -2\sqrt{3}P\sin\Phi\cos 2\psi\sin\varphi, \\ e &= 2\sqrt{3}P\sin\Phi\cos 2\psi\cos\varphi, \\ f &= P\sin 2\varphi(3\cos\Phi + \sqrt{3}\sin\Phi\sin 2\psi), \end{aligned} \quad (17)$$

$$\begin{aligned} g_{11} &= g_1\cos 2\psi - \frac{1}{2}g_2\cos\Phi\cos\psi, \\ g_{21} &= \frac{\sqrt{3}}{2}g_2\sin\Phi\sin\psi, \\ g_{22} &= \cos\varphi(g_1 + g_2\cos\Phi), \\ g_{13} &= \cos\varphi\cos\psi\left(g_1 - \frac{1}{2}g_2\cos\Phi\right), \\ g_{23} &= -\frac{\sqrt{3}}{4}g_2\sin\Phi\sin 2\psi. \end{aligned} \quad (18)$$

The following values of the parameters enumerated above were used for quantitative description of the properties of $LaMnO_3$: J_b = 1.16 meV, J_{ac} = –1.66 meV [9], P = –0.1 meV, a = –0.06 meV, b = 0.09 meV, c = 0.06 meV, d = 0.06 meV, e = –0.3 meV, f = 0.01 meV, g_{11} = 1.865, g_{21} = 0.005, g_{22} = 1.83, g_{13} = 1.825, and g_{23} = –0.005 (calculated using Eqs. (17) and (18) and the parameters φ, ψ, and Φ of the crystal and orbital structures).

4.1. Magnetic Structure in an External Magnetic Field

Minimizing the magnetic energy in the model (16) leads to a four-sublattice magnetic structure, which possesses a noncollinear form and is classified as (A_X, 0, G_Z) (see Fig. 4a) neglecting the rotational distortions and as (A_X, F_Y, G_Z) (see Figs. 4a, 4b) taking these distortions into account.

The expression (16) makes it possible to determine immediately the main components of the magnetic

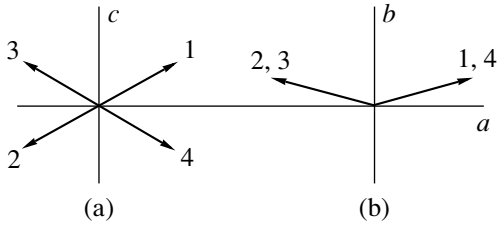


Fig. 4. Magnetic structure of $LaMnO_3$: (a) A_X and G_Z components; (b) A_X and F_Y components.

structure. Comparing the coefficients of the squared components of the basal vectors of the magnetic structure *F*, *G*, *A*, and *C* shows that the components A_X^2 and A_Z^2 have the most negative coefficients: $(-J_b + 2J_{ac} + a)$ and $(-J_b + 2J_{ac} + c)$. It is obvious that for $LaMnO_3$ A_X is the main component of the magnetic structure.

It is easy to see that in the absence of rotational distortions in the crystal the coefficients of the energy (16) $a = c = d = f = 0$, whence follows that in this case the structures $(A_X, 0, G_Z)$ and (A_X, F_Y, G_Z) become equivalent. This agrees with symmetry considerations, since in the absence of rotational distortions the symmetry of the crystal is tetragonal (D_{4h}^5) and the *X*- and *Z*-axes are equivalent.

On account of the presence of orbital structure the easy magnetization axis of the crystal is singled out ($a \sim \sin\Phi$) and nonlinear components of the magnetic structure appear: G_Z $(e \sim \sin\Phi)$ and F_Y $(d \sim \sin\Phi)$. The F_Y component appears also because the orthorhombic nature of the crystal is taken into account ($d \sim \sin\Phi$, $f \sim \sin 2\varphi$). The corresponding components of the magnetic structure can be written as

$$A_X \approx 4S,$$
$$F_Y \approx -4S\frac{ef + 2d(a - c + 4J_{ac})}{f^2 - 4(a - c + 4J_{ac})(a - b - 2J_b)}, \quad (19)$$
$$G_Z \approx -4S\frac{2e(a - b - 2J_b) + df}{f^2 - 4(a - c + 4J_{ac})(a - b - 2J_b)}.$$

Thus, in our calculations the values of the components of the magnetic moments on one manganese ion are $\mu_X \approx 3.7\mu_B$, $\mu_Y \approx 0.05\mu_B$, and $\mu_Z \approx 0.09\mu_B$ (assuming the magnetic moment of the Mn^{3+} ion to be $3.7\mu_B$ [18]). According to the experimental data, $\mu_Y \approx 0.1\mu_B$ [6]. The order of magnitude of μ_Y agrees with experiment even without including in the Hamiltonian the antisymmetric Dzyaloshinskiĭ–Moriya exchange, which we neglected.

It follows from the expression for the energy (16) that the single-ion anisotropy contributes to the terms responsible for the noncollinearity of the structure, specifically, the weak ferromagnetism. As a result of the large anisotropy, these terms are large enough to neglect the antisymmetric exchange.

For orthoferrites, where the oxygen environment of a magnetic iron ion is close to an ideal octahedron and the single-ion anisotropy, second-order in the spin, is negligibly small, the antisymmetric Dzyaloshinskiĭ–Moriya exchange plays a decisive role in the formation of the weakly ferromagnetic structure. In manganites, on account of the cooperative Jahn–Teller effect and the strong spin-orbit interaction, giving rise to a large single-ion anisotropy, the role of antisymmetric exchange in the formation of magnetic structure can be assumed to be secondary. In some calculations, including the present work, it can be rejected, since the role of orbital and crystal structures in the formation of the magnetic order is obvious even without this interaction. However, in the *O** phase the single-ion anisotropy is small, and then the role of antisymmetric exchange becomes appreciable. According to experimental investigations [22] this interaction is approximately the same in both phases.

Nonetheless, in most works (e.g., [8–11]) the contribution to the energy, responsible for the coupling of the A_X and G_Z components, was neglected. However, the constant *e*, determining the magnitude of this contribution and depending primarily on the orbital structure, in our model is quite large and must necessarily be included in the analysis.

In [6], where the magnetic structure of $LaMnO_3$ in the initial model was studied, the orientation of the magnetic moments along the long axis was explained by including the anisotropic terms which are of fourth-order in the spin variables. In our model the easy axis is singled out because of the orbital structure of the crystal.

The behavior of the magnetic structure of an easy-axis antiferromagnet in an external magnetic field is ordinarily studied in a simple model. Neglecting the orbital structure and the rotational distortions the Zeeman contribution to the energy has the form $-\mu_B g(\mathbf{F} \cdot \mathbf{H})$. In this case the magnetic structure in a field directed along the easy axis remains unchanged up to some magnitude of the field (H_{c1}), after which "flipping" of the magnetic sublattices (spin-flop transition) perpendicular to the field occurs. As the external field increases further, the direction of the magnetic moments gradually approaches the direction of the field until saturation occurs (spin-flip transition at H_{c2}). In this approach the behavior of the nonlinear components is not studied.

In our model the behavior of the magnetic structure is more complicated. This, in turn, is due to the fact that the nonlinear components of the magnetic structure are taken into account. The rotational distortions make the interaction of **F** and **H** anisotropic. The orbital structure changes by the magnitude of the components of the *g* tensors and in the presence of rotational distortions adds to this interaction antiferromagnetic components of the structure, specifically, A_X and A_Z. Consequently, we present below the results of numerical calculations.

Let us consider an external magnetic field applied parallel to the easy magnetization axis. For $H < H_{c1}$ the directions of the projections of the magnetic moments of the sublattices on the basal plane remain unchanged. The total ferromagnetic moment (*Y* component) increases depending on the field in a manner so that in the sublattices parallel to the external field the *Y* component increases and in other sublattices it decreases (see Fig. 5). In other words for $H < H_{c1}$ the F_Y and the A_Y components of the magnetic structure increase. The remaining components change very little.

When the field reaches the value H_{c1} the direction of the total ferromagnetic moment changes. In this mag-

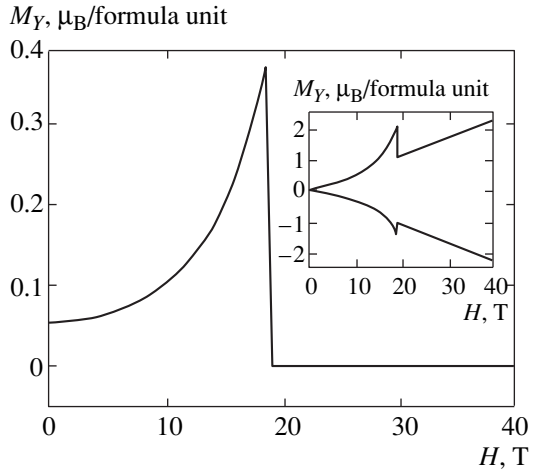


Fig. 5. Total Y component of the magnetic structure μ_Y versus the external magnetic field H. Inset: Y component of the magnetic sublattices versus the external magnetic field H.

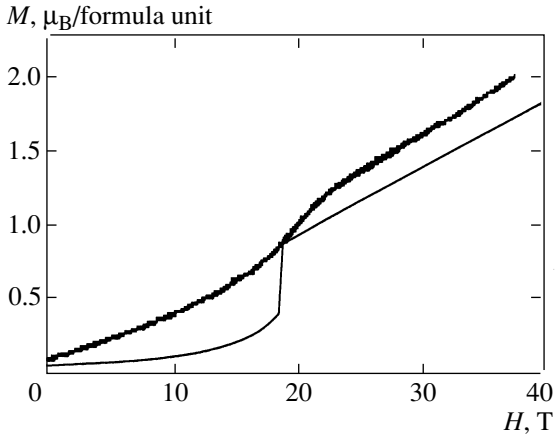


Fig. 6. Total magnetization (per formula unit) μ versus the external magnetic field H. Thick line is the result of measurements from [11].

netic field a spin-flop transition occurs from A_X into A_Z and from F_Y into F_X. The component A_Y continues to increase with the field. The component A_Y begins to decrease for $H > H_{c2}$, when A_Z reaches zero. When the field reaches the value H'_{c2} the magnetization curve tends to saturation. Complete saturation is not attainable in our model because of the anisotropy of the g tensors. The critical fields were $H_{c1} = 19$ T, $H_{c2} = 52.5$ T, and $H'_{c2} \sim 100$ T. The value of the field H_{c1} agrees with the value found experimentally in [11] and is 21 T. The values of the other critical fields have not been measured.

Thus the behavior of the total magnetization of the crystal measured in [11] and calculated in our work (see Fig. 6) can be explained. The computed dependence is qualitatively identical to the characteristic features of the experimental curve: the nonlinear growth up to H_{c1}, a jump at $H = H_{c1}$, linear dependence of the magnetization on the field for $H > H_{c1}$.

The nonlinearity of the growth of μ is associated with the growth of the F_Y component of the magnetic structure. The jump in μ near a spin-flop transition is due to the change in the direction of the ferromagnetic component of the total moment. A quantitative divergence is noticeable in fields below the spin-flop transition field. In [11] this behavior was explained by the presence of domains (twins) in the crystal, for which the external magnetic field is directed along the Y-axis.

4.2. Spin Waves and Antiferromagnetic Resonance

The dispersion dependences of magnons were investigated for the energy (16) in the linear spin-wave approximation. As a result of the four-sublattice model of the magnetic structure, the spin-wave spectrum has four branches. The spectrum is divided into two bands, each having two closely spaced branches.

For $k = 0$ the energy spectrum of the magnetic excitations can be observed with the aid of antiferromagnetic resonance (see Fig. 7).

Without an external magnetic field the lower branches of the spectrum are split slightly and separated by an energy gap. The formation of such a gap is characteristic for an easy-axis antiferromagnet. In a two-sublattice model the gap width is determined by the square root of the product of the exchange parameter and the single-ion anisotropy. In our model the gap width is determined by the square root of a complicated combination of the exchange interaction and single-ion anisotropy parameters. Neglecting the rotational distor-

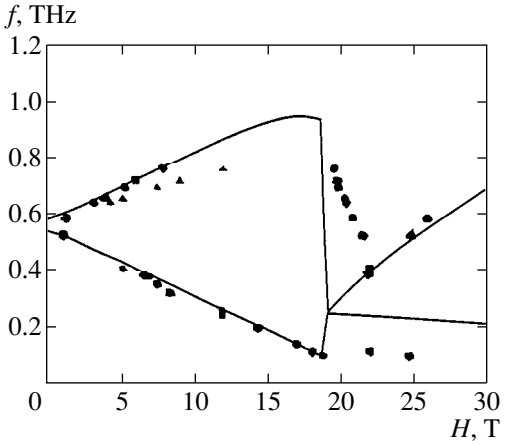


Fig. 7. Antiferromagnetic resonance frequencies f versus the external magnetic field H. Solid line is this calculation, points are the results of [11].

tions the expressions for the squared energies of the antiferromagnetic resonance as a function of the parameters J_b, J_{ac}, b, and e can be written as follows:

$$(\Delta E_1)^2 = \frac{\hbar\gamma S}{2}\left(\frac{e^2}{J_{ac}} + 16J_{ac}\right) \times\left(-\frac{e^2}{J_{ac}} + 4(b - 4J_{ac} + 2J_b)\right),$$
$$(\Delta E_2)^2 = \frac{\hbar\gamma S}{2}\left[\left(-\frac{e^2}{2J_{ac}} + 4b\right)\left(\frac{e^2}{J_{ac}} - 8J_b\right) - 2e^2\left(2 + \frac{b}{J_{ac}}\right)\left(-2 + \frac{J_b}{J_{ac}}\right)\right]. \quad (20)$$

The expressions for the energies taking into account the rotational distortions are even more complicated, and for this reason only the numerical calculations are presented below.

It is easy to see that the main parameters determining the width of the gap in the antiferromagnetic resonance spectrum are P, J_b, and J_{ac}. In contrast to "standard" phenomenological models the gap width depends not only on the antiferromagnetic exchange. Ferromagnetic intraplanar exchange also makes an important contribution to the gap width at the Γ point of the magnetic Brillouin zone. This is due to the influence of orbital structure.

To estimate the parameter P we compared the numerically determined quantity $(\Delta E_1 + \Delta E_2)/2$ with the gap width obtained in [8, 9, 11] and with the average of two energies obtained in [10]. Since the dependence of the splitting δE at $\mathbf{H} = 0$ on the parameters (17) is quite complicated, we did not adjust P to this splitting, especially since the structural parameters on which the constants (17) depend were assumed to be given.

According to our calculations the gap width is ΔE = 2.33 meV, and the splitting of the lower branches was δE = 0.18 meV. Our calculations agree well with the experimental values of these quantities (ΔE = 2.7 [8], 2.6 [9], 2.28 [11], 2.18 meV; δE = 0.12 meV [10]). The main indication that the model is adequate is the agreement between the splitting and the experiment value [10], since this parameter is obtained by using structural data.

In the models used in [8, 9] the formation of a gap is explained well, but the splitting of the branches cannot be explained. In the models of [10, 11] this splitting is attributed to the presence of the Dzyaloshinskiĭ–Moriya interaction. However, in these works the values of the anisotropic parameters are introduced phenomenologically and are not tied to the crystal and orbital structures. The direction of the antisymmetric exchange vector, taking account of the crystal symmetry, has a more complicated form than that assumed in [9, 10]. This is shown, for example, in [22, 23]. The two-sublattice model is inadequate here, since the change in the bond angle between the manganese ions, located in the same plane, must also be taken into account, which ordinarily is not done.

In our model the lifting of the degeneracy for the branches of the spectrum of an easy-axis magnet is associated with the specific form of the single-ion anisotropy, specifically, the presence of the term $e(A_X G_Z + A_Z G_X)$, which depends on the orbital structure. The increase in the number of sublattices, which is associated with the orbital structure, gives rise not only to doubling of the magnetic cell but also lifting of the degeneracy of the pairs of branches at the Γ point. When the rotational distortions are taken into account in the Hamiltonian the quantity δE decreases.

In addition, the field dependence of the antiferromagnetic resonance frequencies with $\mathbf{H} \parallel \mathbf{a}$ was taken into account. The behavior of the field dependence of the frequencies is characteristic for an easy-axis antiferromagnet. As the field in the direction of the easy magnetization axis increases, the two bottom branches of the spectrum diverge even more. At $H = H_{c1}$ a sharp jump associated with the spin-flop transition is observed. In ordinary antiferromagnets the lower branch possesses zero energy at this field. This is not so in our model. This effect is due to the fact that the magnetic structure is not strictly collinear, and the g tensors are anisotropic. As a result, the magnetic field cannot be directed strictly along the direction of the easy magnetization axis. The special features of the spin-flop transition in $LaMnO_3$ and the absence of complete saturation are associated with this.

5. CONCLUSIONS

In summary, in the present work a simple model describing the dispersion [8, 9] and field [10, 11] dependences of the spin-wave energies and the field dependence of the total magnetization [11] was presented. The model included the orbital-dependent interactions: exchange, single-ion anisotropy, Zeeman interaction; the orbital structure is assumed to be fixed and independent of the magnetic interactions. There is no anisotropic exchange in the model, and all rotational distortions are taken into account in other interactions.

As a result of the simplicity of the model the dependence of the magnetic structure on the orbital and crystal orderings can be clearly traced. For the same reason it is possible to give a qualitative explanation of the formation of the magnetic structure [19] and its behavior in an external magnetic field.

ACKNOWLEDGMENTS

This work was supported in part by the CRDF REC-005 program.

REFERENCES

1. J. B. Goodenough, Phys. Rev. **100**, 564 (1955).
2. K. I. Kugel' and D. I. Khomskiĭ, Usp. Fiz. Nauk **136**, 621 (1982) [Sov. Phys. Usp. **25**, 231 (1982)].

3. Y. Murakami, J. P. Hill, D. Gibbs, *et al.*, Phys. Rev. Lett. **81**, 582 (1998).
4. G. Khaliullin and V. Oudovenko, Phys. Rev. B **56**, R14243 (1997).
5. L. F. Feiner and A. M. Oles, Phys. Rev. B **59**, 3295 (1999).
6. G. Matsumoto, J. Phys. Soc. Jpn. **29**, 606 (1970).
7. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
8. F. Moussa, M. Hennion, J. Rodríguez-Carvajal, *et al.*, Phys. Rev. B **54**, 15149 (1996).
9. K. Hirota, N. Kaneko, A. Nishizawa, and Y. Endoh, J. Phys. Soc. Jpn. **65**, 3736 (1996).
10. V. Yu. Ivanov, V. D. Travkin, A. A. Mukhin, *et al.*, J. Appl. Phys. **83**, 7180 (1998).
11. S. Mitsudo, K. Hirano, H. Nojiri, *et al.*, J. Magn. Magn. Mater. **177–181**, 877 (1998).
12. P. Norby, I. G. K. Andersen, E. K. Andersen, and N. H. Andersen, J. Solid State Chem. **119**, 191 (1995).
13. J. Rodríguez-Carvajal, M. Hennion, F. Moussa, *et al.*, Phys. Rev. B **57**, R3189 (1998).
14. V. E. Naĭsh, Fiz. Met. Metalloved. **85**, 5 (1998).
15. A. E. Nikiforov, S. É. Popov, and S. Yu. Shashkin, Fiz. Met. Metalloved. **87**, 16 (1999).
16. O. V. Kovalev, *Irreducible and Induced Representations and Corepresentations of Fedorov Groups* (Nauka, Moscow, 1986).
17. K. S. Aleksandrov, A. T. Anistratov, B. V. Beznosikov, and N. V. Fedoseeva, *Phase Transition in Crystals of Halogen Compounds* (Nauka, Novosibirsk, 1981).
18. Q. Huang, A. Santoro, J. W. Lynn, *et al.*, Phys. Rev. B **55**, 14987 (1997).
19. L. É. Gonchar' and A. E. Nikiforov, Fiz. Tverd. Tela (St. Petersburg) **42**, 1038 (2000) [Phys. Solid State **42**, 1070 (2000)].
20. V. Ya. Mitrofanov, A. E. Nikiforov, and V. I. Cherepanov, *Spectroscopy of Exchangegeable Complexes in Ionic Crystals* (Nauka, Moscow, 1985).
21. A. K. Zvezdin, V. M. Matveev, A. A. Mukhin, and A. I. Popov, *Rare-Earth Ions in Magnetic-Ordered Crystals* (Nauka, Moscow, 1985).
22. D. L. Huber, G. Alejandro, A. Caneiro, *et al.*, Phys. Rev. B **60**, 12155 (1999).
23. I. Solovyev, N. Hamada, and K. Terakura, Phys. Rev. Lett. **76**, 4825 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1230–1234.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1421–1425.

SOLIDS
Structure

Experimental Observation of Bose Condensation in High-Temperature Superconductors

N. P. Seregin* and P. P. Seregin
St. Petersburg State Technical University, St. Petersburg, 195251 Russia
**e-mail: seregin@tuexph.stu.neva.ru*
Received May 30, 2000

Abstract—Emission Mössbauer spectroscopy on ^{67}Cu (^{67}Zn) and ^{67}Ga (^{67}Zn) isotopes was used to show that for the superconductors $Nd_{1.85}Ce_{0.15}CuO_4$, $La_{1.85}Sr_{0.15}CuO_4$, and $Tl_2Ba_2CaCuO_8$ in the temperature range $T > T_c$ the temperature dependence of the center of gravity S of the Mössbauer spectrum is determined by the second-order Doppler shift, while in the range $T < T_c$ the Bose condensation of Cooper pairs influences the value of S (here T_c is the superconducting transition temperature). The spatial nonuniformity produced in the electron density by a Bose condensate of Cooper pairs was observed for $La_{1.85}Sr_{0.15}CuO_4$. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Superconductivity is due to the appearance of Cooper pairs and the formation of a Bose condensate, which is described by a single coherent wave function [1]. This means that the electron density distribution at the lattice sites of a superconductor should be different at temperatures above and below T_c—the superconducting transition temperature.

Since the isomeric shift *I.S.* of the Mössbauer spectrum is determined by the electron density on the nuclei investigated, it is in principle possible to observe the formation of Cooper pairs and a Bose condensate by measuring the temperature dependence of the center of gravity S of the Mössbauer spectrum of superconductors. The temperature dependence of S at constant pressure P is determined by three terms [2]:

$$\left(\frac{\delta S}{\delta T}\right)_P = \left(\frac{\delta I.S.}{\delta \ln V}\right)_T \left(\frac{\delta \ln V}{\delta T}\right)_P + \left(\frac{\delta D}{\delta T}\right)_P + \left(\frac{\delta I.S.}{\delta T}\right)_V. \quad (1)$$

The first-term in Eq. (1) is the dependence of the isomeric shift *I.S.* on the volume V, the second term describes the influence of the second-order Doppler shift, *I.S.*, and in the Debye the approximation it has the form [2]

$$\left(\frac{\delta D}{\delta T}\right)_P = -\left(\frac{3k_B E_0}{2Mc^2}\right) F\left(\frac{T}{\Theta}\right), \quad (2)$$

where k_B is the Boltzmann constant, E_0 is the isomeric transition energy, M is the mass of the probe nucleus, c is the velocity of light in vacuum, Θ is the divine temperature, and $F(T/\Theta)$ is the Debye function. Finally, the third term in Eq. (1) describes the temperature dependence of the isomeric shift *I.S.* at constant volume. The appearance of this term is due to the change in the electron density at the Mössbauer nuclei, and this effect is expected when the matrix transforms into the superconducting state.

However, attempts to observe the formation of Cooper pairs and a Bose condensate by measuring the temperature dependence of the center of gravity S of the Mössbauer spectrum of ^{119}Sn for the classical superconductor Nb_3Sn were unsuccessful [3]: the observed dependence $S(T)$ was described satisfactorily by the second-order Doppler shift and no features in the behavior of $S(T)$ which could be attributed to the change in the isomeric shift were observed near T_c. Similarly, no anomalous change in S of the Mössbauer spectrum of the impurity atoms ^{57}Fe in high-temperature superconductors was observed ($YBa_2Cu_3O_{7-x}$ [4], $(Bi,Pb)_2SrCu_3O_{10}$ [5], $(Tl,Pb)Sr_2Ca_2Cu_3O_y$ [6]). These facts are explained by the small value of $\Delta_{I.S.}/2G$ (here $\Delta_{I.S.}$ is the maximum achievable difference of the isomeric shift of the Mössbauer spectrum in the ordinary and superconducting phases, $G = \hbar/\tau_0$ is the natural width of the nuclear level, and τ_0 is the average lifetime of the nuclear level), which for Mössbauer spectroscopy on the isotopes ^{57}Fe and ^{119}Sn does not exceed 6.

The observation of Cooper pairs by Mössbauer spectroscopy should be most favorable for high-temperature superconductors (which have the minimum Cooper-correlation scale), if a probe for which $\Delta_{I.S.} \gg 10$ is used. The choice of objects for investigation should also take account of the need to introduce a Mössbauer probe at the lattice sites. All these conditions are satisfied for the Mössbauer probe ^{67}Zn in copper metal-oxide lattices when using the emission variant of Mössbauer spectroscopy on the isotopes ^{67}Cu (^{67}Zn): $\Delta_{I.S.}/2G \approx 200$ for ^{67}Zn and the parent isotope ^{67}Cu can be intro-

duced at the copper sites during synthesis, so that the daughter isotope ^{67}Zn also occupies the copper sites in the lattice [7]. However, if the emission variant of spectroscopy on the isotope ^{67}Cu (^{67}Zn) is used, then for copper metal oxides containing rare-earth metals the parent isotope ^{67}Ga occupies the rare-earth metal sites [8], and it becomes possible to investigate the spatial nonuniformity produced in the electron density by the Bose condensate of Cooper pairs. Finally, we note that the Mössbauer probe ^{67}Zn is a dielectronic center with negative correlation energy [9]—its charge state can be changed only by transferring simultaneously two electrons, and the electron pair localized on the center possesses zero total angular momentum, orbital angular momentum and spin. On the other hand, according to be BCS model, for $T < T_c$ electrons with opposite momentum are paired, so that the total angular momentum, orbital angular momentum, and spin of a Cooper pair are also zero. It is the combination of these factors that should be favorable for observing Bose condensation using a ^{67}Zn probe.

In the present paper the results of such investigations are presented for a ^{67}Zn probe in $Nd_{1.85}Ce_{0.15}CuO_4$, $La_{1.85}Sr_{0.15}CuO_4$, and $Tl_2Ba_2CaCu_2O_8$ lattices. Corprous oxide Cu_2O was chosen as the control, for which a transition into the superconducting state is not observed.

2. EXPERIMENTAL RESULTS AND DISCUSSION

The the Mössbauer sources $Nd_{1.85}Ce_{0.15}CuO_4$, $La_{1.85}Sr_{0.15}CuO_4$, $Tl_2Ba_2CaCu_2O_8$, and $^{67}Cu_2O$ were prepared by the diffusing radioactive carrier-free ^{67}Cu into polycrystalline samples $Nd_{1.85}Ce_{0.15}CuO_4$ (T_c = 22 K), $La_{1.85}Sr_{0.15}CuO_4$ (T_c = 37 K), $Tl_2Ba_2CaCu_2O_8$ (T_c = 60 K), and Cu_2O in evacuated quartz ampuls at 450°C for 2 h in an oxygen stream. For $La_{1.85}Sr_{0.15}CuO_4$ diffusion of radioactive carrier-free ^{67}Ga was also conducted under similar conditions. No appreciable change in T_c was observed for the control samples.

The Mössbauer spectra were obtained using the absorber ^{67}ZnS. The absorber temperature was 10 ± 2 K for all samples, while the source temperature could vary from 10 ± 1 to 80 ± 1 K.

The Mössbauer spectra of all ceramics in the chosen temperature range consisted of well-resolved quadrupole triplets, whose isomeric shift corresponds to $^{67}Zn^{2+}$ ions. The typical spectra for the ceramic $La_{1.85}Sr_{0.15}CuO_4$ are displayed in Fig. 1. For the ^{67}Cu (^{67}Zn) spectra it was assumed that in the diffusion process the parent atoms ^{67}Cu occupy the copper sites (the data from [10–12] attest to this), and therefore the probe $^{67}Zn^{2+}$, which forms after ^{67}Cu decays, occupies the copper sites. For the spectra of $La_{1.85}Sr_{0.15}CuO_4$: ^{67}Ga it was assumed that as a result of diffusion doping the parent ^{67}Ga atoms occupy lanthanum sites and therefore the probe $^{67}Zn^{2+}$, formed after ^{67}Ga decays, also occupies the lanthanum sites. This is supported by the fact that in the C–V_{zz} diagram constructed in [8] on the basis of emission Mössbauer spectroscopy data obtained on the isotope ^{67}Cu (^{67}Zn) for the compounds $RBa_2Cu_3O_7$: ^{67}Ga (Fig. 2) the point for $La_{1.85}Sr_{0.15}CuO_4$: ^{67}Ga falls on a straight line plotted assuming that the parent ^{67}Ga atoms (and therefore the daughter probes $^{67}Zn^{2+}$) occupy the rare-earth metal sites R (here C is the quadrupole interaction constant for the $^{67}Zn^{2+}$ probe, determined from the experimental spectrum, and V_{zz} is the principal component of the tensor of the crystal gradients of the electric field, calculated according to the point-charge model).

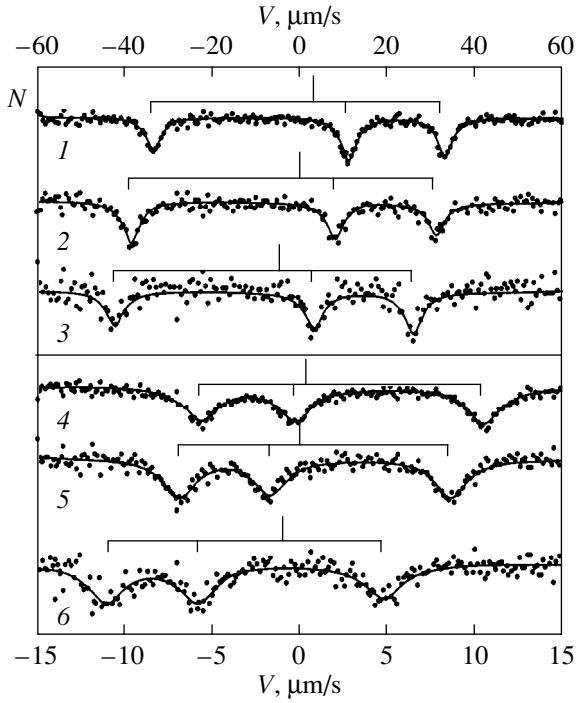


Fig. 1. Mössbauer spectrum of the compounds $La_{1.85}Sr_{0.15}{}^{67}CuO_4$ (*1–3*) and $La_{1.85}Sr_{0.15}CuO_4$: ^{67}Ga (*4–6*) for various source temperatures (absorber temperature 10 ± 1 K, N is the count rate in arbitrary units): 10 K (*1*); 11 K(*4*); 37 K (*2*, *5*); 70 K (*3*); 75 K (*6*).

It was found that the quadrupole interaction constants C for all ceramics are essentially temperature-independent. Since for the Zn^{2+} probe the electric-field gradient at the ^{67}Zn nuclei is produced only by the lattice ions [9–12] and the variations of the lattice constants in the temperature range 4.8–80 K are negligibly small [13, 14], the temperature independence of C is not unexpected.

The temperature dependences of the center of gravity S of the spectrum, which were measured relative to its value at T_c, differ substantially for the control and superconducting materials (as an example, Fig. 3 displays such curves for $La_{1.85}Sr_{0.15}CuO_4$ and Cu_2O),

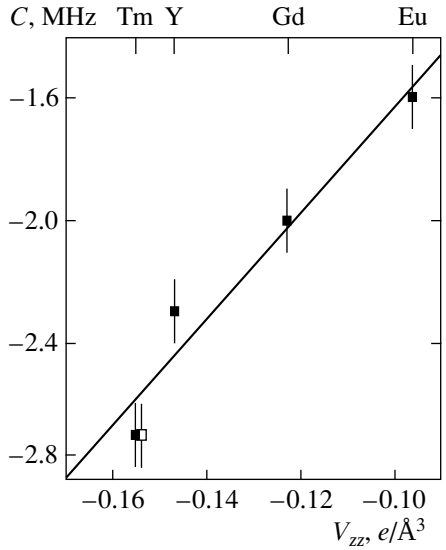


Fig. 2. Quadrupole interaction constant C for rare-earth metal sites (the experimental data, obtained by emission Mössbauer spectroscopy of ^{67}Ga (^{67}Zn) versus the principal components of the tensor of the crystal gradient of the electric field V_{zz} at the same sites (the computational results were obtained in the point-charge approximation) for $RBa_2Cu_3O_7$, R = Y, Eu, Gd, Tm (small filled squares) [8]. The large open square represents our data for lanthanum sites in the $La_{1.85}Sr_{0.15}CuO_4$ lattice.

though no sharp jumps in S are observed at a transition through T_c in any of the compounds.

The temperature dependence of S is determined by the expression (1) and, as calculations show [15], the first term in Eq. (1) for the case ^{67}Zn can be neglected, since for the temperature range chosen it does not exceed 0.03 μm/s and, and in the temperature range 10–80 K no structural phase transitions are observed in any of the compounds investigated [16].

The second term in Eq. (1) describes the effect of the second-order Doppler shift. As one can see from Fig. 3, the experimental data for the control sample Cu_2O in chosen temperature range are satisfactorily described by the dependence (2), drawn for $\Theta \approx 400$ K. For superconducting samples the experimental data for temperatures $T > T_c$ are also described by the dependence (2) plotted for $\Theta \approx 360$ ($Nd_{1.85}Ce_{0.15}CuO_4$), 400 ($La_{1.85}Sr_{0.15}CuO_4$), and 260 K ($Tl_2Ba_2CaCu_2O_8$) (according to specific-heat measurements the Debye temperatures for Nd_2CuO_4, $La_{2-x}Sr_xCuO_4$, and $Tl_2Ba_2CaCu_2O_8$ are, respectively, 300 [17], 420 [18], and 270 K [18]).

Finally, the third term in the expression (1) describes the temperature dependence of the isomeric shift *I.S.*, and this term appears because the electron density changes at the ^{67}Zn nuclei. The quantity *I.S.* at a given temperature T can be found as the difference

$$[I.S.]_T = S_T - D_T$$

(here S_T and D_T are, respectively, the center of gravity of the spectrum and the Doppler shift at temperature T). The fact that $[I.S.]_T$ increases with decreasing temperature in the range $T < T_c$ shows that the electron density increases on the ^{67}Zn nuclei as result of the appearance of Cooper pairs and their Bose condensation. The limiting

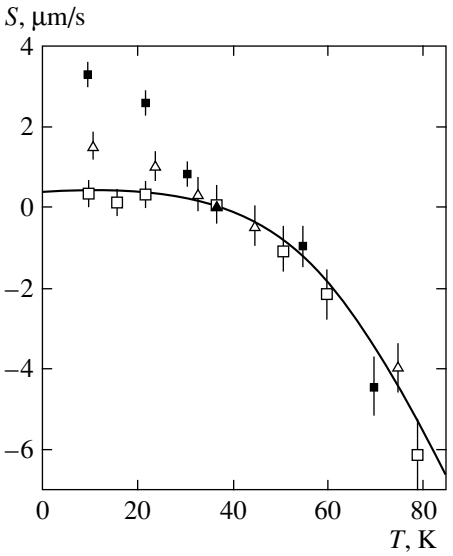


Fig. 3. Temperature dependences of the center of gravity S of the Mössbauer spectrum of ^{67}Zn, measured relative to its value at 37 K, for $La_{1.85}Sr_{0.15}{}^{67}CuO_4$ (■), $La_{1.85}Sr_{0.15}CuO_4$: ^{67}Ga (△), and $^{67}Cu_2O$ (□). The theoretical temperature dependence of S is presented for the second-order Doppler shift with $\Theta = 400$ K.

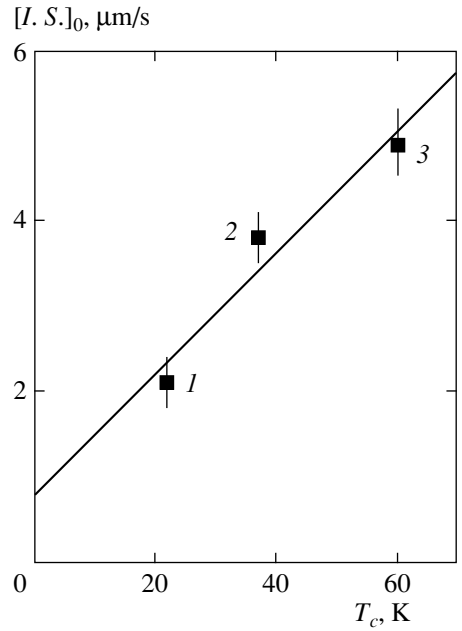


Fig. 4. $[I.S.]_0$ versus T_c. The symbols represent the compounds $Nd_{1.85}Ce_{0.15}{}^{67}CuO_4$ (*1*), $La_{1.85}Sr_{0.15}{}^{67}CuO_4$ (*2*), and $Tl_2Ba_2Ca^{67}Cu_2O_8$ (*3*).

values of *I.S.* in the limit $T \longrightarrow 0$, $[I.S.]_0 = S_0 - D_0$, should depend on the size of the Cooper pairs, i.e., on T_c. The validity of this conclusion is illustrated in Fig. 4, which shows the dependence of $[I.S.]_0$ on T_c: as T_c increases (i.e., as the Cooper-correlation radius decreases), the quantity $[I.S.]_0$ increases; this reflects the fact that the electron density on the $^{67}Zn^{2+}$ nuclei increases. Ii is significant that the quantity $[I.S.]_0 = 1.5 \pm 0.3$ μm/s for $^{67}Zn^{2+}$ centers at the lanthanum sites is much smaller than $[I.S.]_0 = 3.8 \pm 0.3$ μm/s for $^{67}Zn^{2+}$ centers at the copper sites (see Fig. 3). Evidently, this is a consequence of the spatial nonuniformity produced in the electron density by the Bose condensate of Cooper pairs.

Thus, it has been established that for the superconductors $Nd_{1.85}Ce_{0.15}CuO_4$, $La_{1.85}Sr_{0.15}CuO_4$, and $Tl_2Ba_2CaCu_2O_8$ the temperature dependence of *S* in the range $T > T_c$ is determined by the second-order Doppler shift, while in the temperature range *S* the appearance of Cooper pairs and their Bose condensation also influence the value of *S*. As the temperature decreases, the influence of the indicated process on *S* increases, since the Bose-condensate fraction increases with decreasing temperature.

In general, the temperature dependence of the effective density of superfluid electrons, ρ(*T*), can be written [1] as

$$\rho(T) = 1 - \frac{2\beta E_F}{k_F^5}\int_0^\infty \frac{k^4 \exp(\beta E_k)}{[\exp(\beta E_k) + 1]^2} dk,$$

where $E_F = k_F^2/2m$ is the Fermi energy, *m* is the particle mass, *k* is the wave number, k_F is the wave number at the fermi surface, E_k is the energy of the *k* state, and β is the reciprocal of the binding energy of the superfluid component.

On the other hand it should be expected that

$$\rho(T) \propto [I.S.]_T/[I.S.]_0.$$

Consequently, the theoretical dependence of ρ on the parameter $x = 1.76(k_B T/\Delta)$ ($\Delta = 3.06 k_B \sqrt{T_c(T_c - T)}$ is the energy gap in the elementary exultations spectrum of the superconductor), taken from [1], is presented in Fig. 5 together with our data on the dependence of $[I.S.]_T/[I.S.]_0$ on the parameter *x* for the emission spectroscopy of ^{67}Cu (^{67}Zn). It is evident that the agreement between the calculated and the experimental values is satisfactory. In other words Mössbauer spectroscopy on the isotope ^{67}Zn is an effective method for investigating the formation of Cooper pairs and their Bose condensation in high-temperature superconductors.

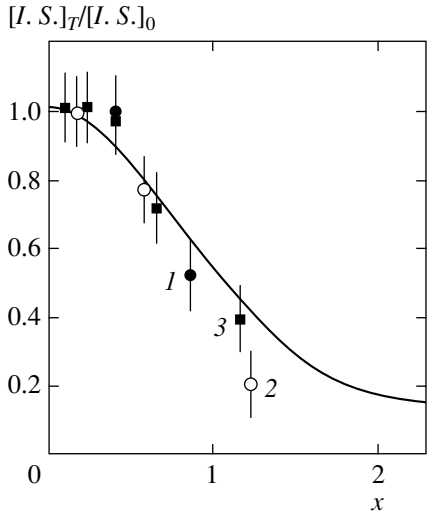


Fig. 5. $[I.S.]_T/[I.S.]_0$ versus the parameter $x = 1.76(k_B T/\Delta)$. The solid curve shows the theoretical dependence of the effective density of superfluid electrons on the parameter *x*. The symbols represent the compounds $Nd_{1.85}Ce_{0.15}{}^{67}CuO_4$ (*1*), $La_{1.85}Sr_{0.15}{}^{67}CuO_4$ (*2*), and $Tl_2Ba_2Ca^{67}Cu_2O_8$ (*3*).

3. CONCLUSIONS

It was shown for the compounds $Nd_{1.85}Ce_{0.15}CuO_4$, $La_{1.85}Sr_{0.15}CuO_4$, and $Tl_2Ba_2CaCu_2O_8$ by emission Mössbauer spectroscopy on the isotopes ^{67}Cu (^{67}Zn) and ^{67}Ga (^{67}Zn) that the transition into the superconducting state is accompanied by a redistribution of the electronic density of the crystal, and emission Mössbauer spectroscopy of ^{67}Zn is an effective method for investigating Bose condensation of Cooper pairs.

REFERENCES

1. J. R. Schrieffer, *Theory of Superconductivity* (Benjamin, New York, 1964; Nauka, Moscow, 1970).
2. D. L. Nagy, in *Mössbauer Spectroscopy of Frozen Solutions*, Ed. by A. Vértes and D. L. Nagy (Akadémiai Kiadó, Budapest, 1990; Mir, Moscow, 1998).
3. J. S. Shier and R. D. Taylor, Phys. Rev. **174**, 346 (1968).
4. Y. Wu, S. Pradhan, and P. Boolchand, Phys. Rev. Lett. **67**, 3184 (1991).
5. Th. Sinnemann, R. Job, and M. Rosenberg, Phys. Rev. B **45**, 4941 (1992).
6. Yun-Bo Wang, Guo-Hui Cao, Yang Li, *et al.*, Physica C (Amsterdam) **282–287**, 1087 (1997).
7. F. S. Nasredinov, N. P. Seregin, and P. P. Seregin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 632 (1999) [JETP Lett. **70**, 641 (1999)].
8. V. F. Masterov, F. S. Nasredinov, N. P. Seregin, and P. P. Seregin, J. Phys.: Condens. Matter **7**, 2345 (1995).
9. F. S. Nasredinov, N. P. Seregin, and P. P. Seregin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 275 (2000) [Semiconductors **34**, 269 (2000)].

10. N. P. Seregin, F. S. Nasredinov, V. F. Masterov, and P. P. Seregin, Solid State Commun. **87**, 345 (1993).
11. N. P. Seregin, F. S. Nasredinov, V. F. Masterov, *et al.*, Zh. Éksp. Teor. Fiz. **100**, 2038 (1991) [Sov. Phys. JETP **73**, 1129 (1991)].
12. V. F. Masterov, F. S. Nasredinov, N. P. Seregin, and P. P. Seregin, Zh. Éksp. Teor. Fiz. **114**, 1079 (1998) [JETP **87**, 588 (1998)].
13. W. Sadowski, H. Hagemann, M. Francois, *et al.*, Physica C (Amsterdam) **170**, 103 (1990).
14. M. Braden, P. Schweiss, C. Heger, *et al.*, Physica C (Amsterdam) **223**, 396 (1994).
15. M. Steiner, W. Potzel, C. Schafer, *et al.*, Phys. Rev. B **41**, 1750 (1990).
16. K. Yvon and M. Francois, Z. Phys. B **76**, 415 (1989).
17. A. Tigheza, R. Kuentzler, G. Pourroy, *et al.*, Physica B (Amsterdam) **165–166**, 1331 (1990).
18. H. M. Ledbetter, S. A. Kim, and R. B. Goldfarb, Phys. Rev. B **39**, 9689 (1989).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1235–1241.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1426–1433.

SOLIDS
Electronic Properties

Features of the Dissipative Electron Transport Near the Boundary in the System Al(*N*)-In(*S*)

Y. N. Chiang*, S. N. Gritsenko, and O. G. Shevchenko

Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of Ukraine, Kharkov, 61164 Ukraine
**e-mail: chiang@ilt.kharkov.ua*

Received June 1, 2000

Abstract—The temperature behavior of the coherent electron transport that arises in the presence of Andreev reflections is studied near the boundary of a hybrid system consisting of aluminum (*N*) and indium (*S*). The qualitative change of the temperature behavior upon changes in the electron mean free path l_{el} as a result of the elastic deformation of the sample is observed for the first time on the same sample, at probes mounted in the normal region at fixed distances and from the boundary. As the temperature is lowered, the measured effective resistance decreases for $l_{el} \ll L_{1,2}$ and increases in a certain temperature region in which $l_{el} \sim L_{1,2}$. It is shown that phenomena of this kind correspond to quantum-interference features in the scattering of coherent electronic and Andreev hole excitations on elastic centers. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

This paper is a continuation of research on electron transport in very pure normal metals in contact with a superconductor. The use of such metals in systems is of fundamental importance, since for them the dissipative transport in the temperature region corresponding to elastic scattering is substantially coherent (determined by the scattering of coherent electronic excitations and hole (Andreev) excitations) at macroscopic distances from the boundary. This, of course, is due to the fact that in pure metals the ballistic mean free path of the electrons reaches macroscopic values, and the spectrum of Andreev excitations is always resolved at ballistic distances from an boundary. In such a situation it becomes possible to carry out experimental studies of the fundamental problems associated with quantum interference of coherent excitations and their dissipative contribution to the conductivity. This is the main difference between the experimental setup studied here and the study of coherent transport in mesoscopic samples [1, 2], in which this transport is only a small correction to the diffusion contribution on account of the extremely short ballistic mean free paths of the electrons (several orders of magnitude smaller than for pure metals). Studies [3–5] of systems containing pure metals (Cu–Sn, Pb(*N*)–Pb(*S*), Sn(*N*)–Sn(*S*)) have confirmed the theoretical expectations of an unusual manifestation of quantum interference of coherent electrons in the conductivity [6–8] under conditions such that the scattering of coherent excitations on elastic centers is predominant, as in pure metals at distances of up to several millimeters from an boundary. Previously we have studied hybrid systems in which a contact between a normal metal and a superconductor was either formed by melting [3, 4] or arose as a boundary between normal and superconducting phases in the intermediate domain state of a type-I superconductor in a magnetic field [5]. In the present paper we investigate the features of the coherent transport in a system with yet another type of contact—a contact of a purely mechanical type.

2. EXPERIMENTAL PROCEDURE AND RESULTS

The inset in Fig. 1 shows the construction of the sample holder with a hybrid system consisting of single-crystal aluminum and indium (shaded). The mean

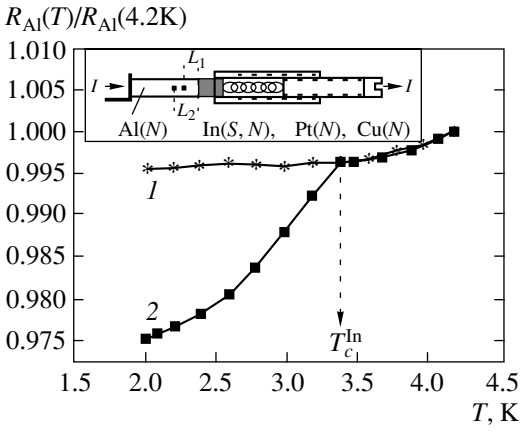


Fig. 1. Temperature dependence of the effective resistance (normalized to the value at $T = 4.2$ K) of aluminum at a short elastic mean free path $l_{el} \approx 0.2$ μm and $l_{inel} > L_2$, $\bar{L} \gg l_{el}$; (*1*) at an *NN* boundary, (*2*) at an *NS* boundary for $T < T_c^{In}$. The lines connect the experimental points. The inset shows the construction of the sample holder with the hybrid *NS* system.

1063-7761/00/9106-1235$20.00 © 2000 MAIK "Nauka/Interperiodica"

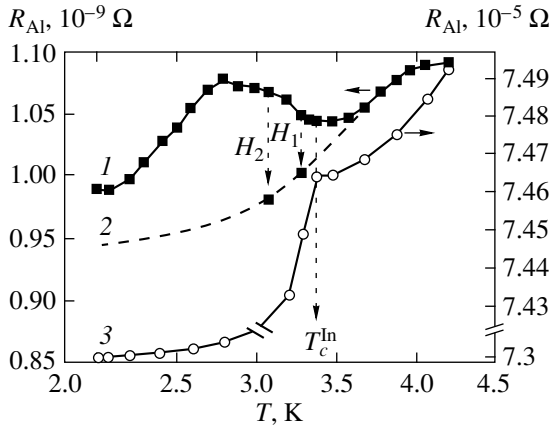


Fig. 2. Temperature dependence of the unnormalized effective resistance of aluminum: (*1*) for a large elastic mean free path m and $l_{el} \approx 100$ μm with the *NS* boundary at $T < T_c^{In}$; (*2*) the same, with an *NS* boundary (the points on this curve were obtained with the superconducting state of the indium destroyed by an external magnetic field); (*3*) the unnormalized curve *2* from Fig. 1. The lines connect the experimental points.

free path in the system was varied by applying uniaxial compression with a spring and screw mechanism. Aluminum sample had a cross section of mm and a total length of around 1 cm. Copper measuring probes were spot welded to the sample at distances of $L_1 \approx 0.1$ mm and $L_2 \approx 0.5$ mm from the contact with the indium. The resistance ratio (room-to-helium temperature) measured at these probes prior to compression and after compression had values of 14490 ($l_{el} \approx 0.1$ mm) and 19 ($l_{el} \approx 2 \times 10^{-4}$ mm), respectively. The first cooling to helium temperature of the stressed sample was done immediately after its mechanical compression. After a cycle of measurements the entire structure containing the sample was held at room temperature for a week. The subsequent measurements of the of aluminum (≈14020) revealed a practically total recovery of this ratio to the above-indicated value for the initial unstressed state, with the absolute values of the resistance of the sample at helium and room temperatures decreasing by four and two orders of magnitude, respectively. This means that the deformation of the sample did not exceed the limits of elasticity and that the change in the mean free path occurred mainly on account of a change in the concentration of point defects of the vacancy type, which, as can be seen from the value of after the stress was relieved, have a "self-healing" capability. The stress was relieved through a loss of elasticity of the spring after thermocycling and the plastic flow of the indium spacer between the aluminum and the spring.

The sample and holder were placed in a small superconducting solenoid with an inner diameter of 8 mm in order to permit changing the state of the boundary by destroying the superconducting state of the indium.

The resistance of the layer of normal metal with a volume of the order of l3 at $l = l_{el} \sim 100$ m is 10^{-8}–10^{-9} Ω. Therefore, at reasonable measuring currents (≤1 A), investigation of resistive effects of this order requires a resolution of the potential U differences at the level of $\delta U = (10^{-11}$–$10^{-12})$ V or better. Such a resolution was achieved by employing a null method of measuring potential differences with the use of a superconducting modulator [9].

Curve *1* (asterisks) in Fig. 1 shows the temperature dependence of the normal effective resistance measured at probes and for a contact between aluminum and a metal that maintains a normal state over the entire temperature interval investigated. The contact is either mechanical (with copper) or welded (with platinum). The form of curve *1* is the same for both types of boundaries and does not depend on the quality of the contact. It is seen that the change of the resistance with temperature over the temperature interval shown is a fraction of a percent.

A completely different picture is observed in the contact of a sample with indium when the sample is elastically compressed until the mean free path is decreased to $\sim 2 \times 10^{-4}$ mm (curve *2*, squares). Up until the transition of indium to the superconducting state, the temperature-dependent part of the resistance of the aluminum between the measuring probes and behaves the same as for any other type of normal boundary, maintaining a practically constant absolute value independently of the value of the total resistance, i.e., of the elastic mean free path of the electrons; this situation is described to good accuracy by the Matthiessen rule. However, after the transition of the indium to the superconducting state (at the critical temperature T_c^{In}) and the formation of an boundary, the conductivity of the aluminum region bounded by the probes increases strongly as the temperature is lowered. The change in resistance of this region as the temperature is reduced from $T = T_c^{In}$ to the lower boundary of the measurement interval is around 2% of the value of the resistance at $T = T_c^{In}$ rather than the value measured at a normal boundary, i.e., the change is nearly 100 times larger.

Curve *1* in Fig. 2 shows the temperature dependence of the renormalized effective resistance of the aluminum region between the same pair of probes after recovery of the elastic mean free path of the electrons to its initial value in the unstressed sample, $l_{el} \sim 100$ μm. The two separate points are the values of the resistance of the aluminum when the superconducting state of the indium is destroyed by the field of the solenoid, equal to the critical fields of indium, 49 and 140 G, at the respective temperatures. (At lower temperatures, i.e., at higher critical fields, the magnetoresistive effective distorts the values of the normal resistance of aluminum and for this reason are not used here.) These values, together with the part of curve *1* at $T > T_c^{In}$, can give an

approximate idea of the temperature dependence of the resistance of the region $[L_1; L_2]$ of aluminum in the absence of an boundary over the entire temperature interval of the measurements (dashed curve *2*). For comparison, we also show in this figure the temperature dependence of the unnormalized resistance for the stressed sample (curve *3*). It follows from the form of curve *1* that for an elastic mean free path of the electrons comparable to the extent of the normal region from the site of contact with the superconducting indium, the effective resistance of the region manifests unusual nonmonotonic behavior as the temperature is lowered, initially increasing after the formation of the boundary and then decreasing at temperatures lying below approximately $0.8T_c^{\mathrm{In}}$.

3. DISCUSSION OF THE RESULTS

The results obtained in the experiment give a clear idea of the features of the interaction of coherent electrons with elastic scattering centers and permit making qualitative and quantitative estimates of a number of important microscopic parameters pertaining to coherent transport in the presence of an boundary. Although some of these features have been reported in our previous papers [3, 4], the corresponding data were obtained either on different samples or with a replacement of the probes or with the current applied through a point boundary, when the ratio of the mean free path not only to the thickness but also to the width of the normal layer is important and the pattern of current spreading is undetermined (sample Sp4 of [4]), so that it is difficult to make quantitative estimates. The experiment of the present study is free from these limitations.

3.1. Theory

We shall show below that the phenomena observed in this study are qualitatively and quantitatively incorporated in ideas about the features of the interaction of coherent excitations which are due to Andreev reflections at an boundary and which are responsible for the transport in the normal metal in a region adjacent to the boundary on a scale of the phase coherence length, which is comparable to the inelastic mean free path of the coherent excitations l_{inel}. In view of the lack of a theoretical treatment of the temperature behavior of the coherent transport in metals layers with a thickness of the order of l_{inel}, in this Subsection we shall use these ideas to obtain a suitable analytical expression for describing the temperature behavior of the conductance of such a layer as a whole. We shall show that in the case when a layer of normal metal of thickness $\bar{L} \gg l_{\mathrm{el}}$ contains a large number of elastic centers, which scatter excitations with a loss of momentum but not of the phase of the wave function ($l_{\mathrm{inel}} \geq \bar{L}$), the conductance of this layer can increase as the temperature is lowered on account of the thermal expansion of the region of phase coherence (an increase in l_{inel}) and, as a consequence, a decrease in the fraction of dissipative coherent trajectories (with a length that depends on the number of Andreev reflections at the boundary), on which the phase of the wave functions relaxes in the equilibrium region.

In Subsection 3.2.2 we shall show that when the thickness of the layer of normal metal adjacent to the boundary is comparable to the distance between elastic centers ($l_{\mathrm{inel}} > \bar{L} \sim l_{\mathrm{el}}$) and with the thermal length $\lambda_T \approx \hbar v_{\mathrm{F}}/k_{\mathrm{B}}T$ (v_{F} is the Fermi velocity and k_{B} is Boltzmann's constant), in addition to the aforementioned behavior a temperature behavior of the opposite character should also appear, due to the independent dissipative contribution of coherent trajectories which is not related to the position of the equilibrium region (to l_{inel}), i.e., trajectories on which the phase of the wave functions does not relax. The contribution of this group of trajectories is related to the size λ_T of the thermal layer, and it is the increase in the size of this thermal layer as the temperature is lowered that determines the temperature behavior of this contribution. The competition of these two quantum mechanisms of dissipation near the boundary under these conditions is analyzed for the experimental results obtained.

Previously we have studied in detail the latter of these dissipation mechanisms [4] and have shown that it corresponds to the fundamental conclusion of the theory that there is a twofold increase in the scattering cross section of an elastic center for an electron undergoing coherent Andreev reflection at the boundary [6]. In view of this circumstance, with allowance for the multiple coherent scattering on all the impurities in a layer of thickness λ_T adjacent to the boundary, the resistance of the metal measured at a distance from the boundary should include the following correction to the value that obtains in the absence $\bar{L}$ of an boundary ($\bar{L} \geq l_{\mathrm{el}} > \lambda_T$, $\bar{L}^{-1} \simeq \bar{L}_1^{-1} - \bar{L}_2^{-1}$ for $L_i \sim l_{\mathrm{el}}$ [4]):

$$\frac{\delta R_N}{R_N} = \frac{l_{\mathrm{el}}}{\bar{L}}\bar{r}, \tag{1}$$

where $\bar{r}$ is the effective scattering probability for electrons undergoing Andreev reflection in the layer λ_T as a whole [8]. An analysis of $\bar{r}$ showed [4] that expression (1) correctly describes the experimentally observed growth of the resistance of layers of normal metal located at a distance of the order of the elastic mean free path from the boundary when the temperature is lowered below the temperature of the onset of the superconducting state of the metal in contact with the normal metal.

However, as was made clear in a study [4] of the hybrid system Cu–Sn and is clearly confirmed by the results of the present study, the resistance of layers adjoining the boundary and having thicknesses satisfy-

ing the condition $l_{el} \ll \bar{L} < l_{inel}$ can exhibit temperature behavior that is directly opposite, decreasing strongly as the temperature is lowered (Figs. 1 and 2). The problem of the temperature dependence of the excess conductivity due to coherent scattering under the indicated conditions has not yet been analyzed theoretically, but this can be done in terms of the Landauer concept (which has been confirmed by numerous experiments) as to the nature of dissipation under conditions of a random, elastically scattering potential. According to this concept, complete thermalization of the electron is the result not of momentum relaxation but of relaxation of the phase of the wave function due to inelastic scattering in regions with an equilibrium distribution (regions that are sinks and sources of charges), called reservoirs. The Landauer concept was used, in particular, in [7], in a calculation of the conductivity of an system. We shall follow the scheme of this calculation to determine the temperature contribution of coherent excitations in such a system. One of the important results of [7] was obtained by modeling a continuous random walk of an elastically scattered particle in a three-dimensional layer of the metal: the (diffusional) mean free path $\langle L \rangle$ depends linearly on the width of the normal layer. In terms of the Landauer concept, the layer width in our case should be understood to mean the distance between the boundary and the region of equilibrium distribution (reservoir), which is of the order of magnitude of the inelastic mean free path, $d \sim l_{inel}$. Since the probability (r for and excitations to pass from the boundary to the reservoir is inversely proportional to the layer width, it follows from the linear relation between $\langle L \rangle$ and d that $\langle L \rangle \sim \tau_r^{-1} \sim l_{inel}$. In [7] the probability of passage τ_r is introduced in the form $\tau_r = l_{el}/d$. However, it is also necessary to take into account the probability of realizing a diffusional trajectory by equating its length to the elastic mean free path, or, equivalently, equating the length $\langle L \rangle \sim \sqrt{\hbar D/k_B T}$ ($D \sim v_F l_{el}$ is the diffusion coefficient) to the real length of the trajectory. Since $\langle L \rangle = \sqrt{Dt}$ and $t = L/v_F$, we have $\langle L \rangle / L = l_{el}/\langle L \rangle$. In addition, we shall assume that the only temperature-dependent cause of inelastic scattering is inelastic electron-phonon collisions, with a corresponding mean free path $l_{inel} \sim l_{inel}^{e-ph} \sim T^{-3}$. This last quantity can be normalized (i.e., a normalizing coefficient can be found) according to some reliably determined value. This is most simply done from the temperature dependence of the resistance of the pure normal metal at some temperature $T^* \ll \Theta$ (at which the elastic effective electron-phonon length is numerically equal to l_{el}, so that

$$l_{inel}(T) = l_{inel}^* \left(\frac{T^*}{T}\right)^3, \qquad (2)$$

$$l_{inel}^* = l_{e-ph}^* \left(\frac{T^*}{\Theta}\right)^{n-3}, \quad l_{e-ph}^* = \alpha \left(\frac{\Theta}{T^*}\right)^n = l_{el},$$

where Θ is the Debye temperature and is the exponent in the temperature dependence of the resistance of the normal metal for $T \ll \Theta$.

As a result, the effective probability for coherent excitations to pass through the phase coherence region in elastic scattering can be written in the form

$$\tau_r = \frac{l_{el}}{l_{inel}} \frac{l_{el}}{\langle L \rangle} = \beta T^{3.5},$$
$$\beta = l_{el}^{3/2} \left(\frac{\hbar v_F}{k_B}\right)^{-1/2} [l_{inel}^* T^{*3}]^{-1}. \qquad (3)$$

In accordance with the Landauer concept, we find the relative contribution to the conductance in the phase coherence region by equating the fraction of coherent trajectories (those that return to the reservoir after reflections from the boundary, starting with the trajectory with and their contribution to the current and summing over all trajectories:

$$\frac{\delta G}{G_0} = \sum_{m=1}^{\infty} F(m) I(m), \qquad (4)$$

where $\delta G = G - G_0$, $G_0 \equiv G_{T=0}$,

$$F(m) = \tau_r^2 (1 - \tau_r)^{m-1} \quad (m \neq 0).$$

The probabilistic contribution to the current from a charge on trajectory with reflections is [7,11]

$$I(m) = 1 + |r_{eh}(m)|^2 - |r_{ee}(m)|^2,$$
$$|r_{eh}(m)|^2 + |r_{ee}(m)|^2 = 1,$$

where $|r_{ee}(m)|^2$ and $|r_{eh}(m)|^2$ are the probabilities for an electron incident on the boundary to leave the boundary, after reflections (Fig. 3), in the form of an electron wave or hole (Andreev) wave, respectively. The expression for $I(m)$ shows that for a large enough number of reflections, which increases the probability of Andreev reflection to such a degree that $|r_{eh}(m)|^2 \longrightarrow 1$, the contribution of the corresponding trajectory to the current increases by a factor of 2. If all of those trajectories reached the reservoir, the dissipation would be increased by the same factor. Formally this is a consequence of the same fundamental conclusion of the theory which was mentioned above: that in coherent Andreev reflection the efficiency of the elastic scattering of the electron momentum increases as a result of the interference of the and excitations [6, 8]. Actually, the fraction of the coherent trajectories that returns to the reservoir decreases rapidly with increasing distance to the reservoir from the boundary and with increasing number of reflections, which determines the length of the trajectory; thus we have the directly opposite result. In fact, assuming that for low electron energies ($eV/(\hbar v_F/l_{el}) \ll 1$) and a large contact area the main contribution to the change in con-

ductivity is from coherent trajectories with large numbers of reflections, so that $I(m) \approx 2$, and changing the sum in expression (4) to an integral, we find to a second approximation:

$$\frac{\delta G}{G_0} \approx 2\int_1^{m^*} F(m)dm \approx -\tau_r^2(m^{*2}\tau_r - 2m^*). \qquad (5)$$

The upper limit of integration m^* is the number of reflections corresponding to a certain critical length for a coherent trajectory L that reaches the reservoir. Longer trajectories (with larger give an exponentially small contribution to the change in conductivity, as do trajectories with a smaller number of reflections and shorter length, which do not reach the reservoir. Since varies in a direct relation with the number of reflections, while $\langle L\rangle$, as we have said, is inversely proportional to τ_r, for $W \gg l_{\text{inel}}$ (W is the transverse dimension of the contact) the upper limit of integration can be introduced as $m^* = \gamma\tau_r^{-1}$ with a certain coefficient γ that is to be determined experimentally. Substituting m^* into (5), we finally obtain

$$\frac{\delta G}{G_0} \approx -(\gamma^2 - 2\gamma)\tau_r = -AT^{3.5}, \quad A = \beta(\gamma^2 - 2\gamma). \quad (6)$$

This effect wherein the conductance increases appreciably with decreasing temperature is actually of the same nature as the growth of the conductivity as the number of elastic centers increases [7]. It consists in the fact that the number of trajectories leaving from the number of attainable reservoirs increases in the long-range phase coherence region, i.e., an ever greater number of trajectories appear on which the phase of the coherent wave functions does not relax; this decreases the dissipation. Thus, in accordance with (5), one expects that the temperature dependence of the relative effective resistance measured at probes separated by distances within the phase coherence region will be in the form of a function that decreases with decreasing temperature:

$$\frac{R}{R_N} = \frac{R_0}{R_N}(1 + AT^{3.5}), \qquad (7)$$

where $R_0 = G_0^{-1}$ and $R_N \equiv R_{\text{Al}}(T = T_c^{\text{In}})$ for the system under study.

In addition to this behavior, under certain conditions mentioned above there can also be an effect wherein the conductivity decreases according to Eq. (1). As we see in Fig. 3, besides the trajectories of types 1 and 3, which escape to the reservoir, for multiple Andreev reflections a doubled contribution to the dissipation which is independent of the position of the reservoir is given by momentum scattering on closed coherent trajectories of the type 2. At temperatures close to the temperature of the onset of the boundary, when $|r_{eh}(m)|^2 \ll 1$ ($\Delta \ll k_B T$, where Δ is the energy gap in the superconductor in contact with the normal metal) and is close to zero, the main contribution to the dissipation, as follows from

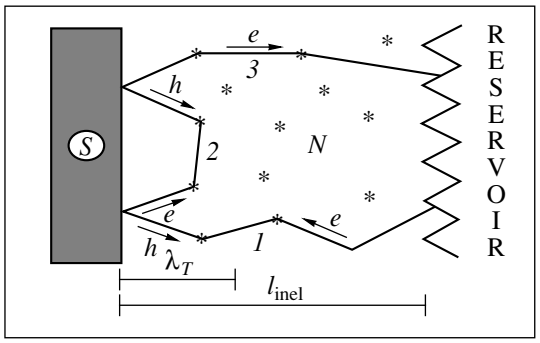


Fig. 3. An example of a trajectory with multiple Andreev reflections (from [7]): *1* and *3*—the part of the trajectory on which the phase of the wave functions relaxes in the reservoir; *2*—the part of the interference trajectory between reflections, on which the phase shift does not relax.

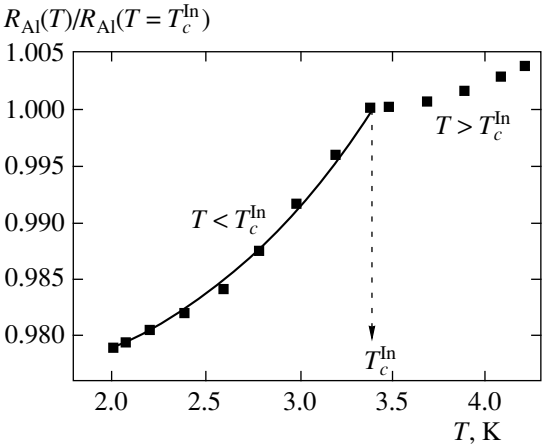


Fig. 4. Temperature dependence of the effective resistance (normalized to the value $T = T_c^{\text{In}}$ of aluminum for $l_{\text{inel}} > L_2$, $\bar{L} \gg l_{\text{el}}$, as compared with the theory (solid curve).

Eq. (1), is given by just such trajectories in a layer with a thickness of the order of the thermal coherence length λ_T. Thus in the case when the phase coherence length is comparable to the ballistic (elastic) mean free path, the temperature dependence of the effective resistance measured at probes placed a distance for the boundary of the order of this same distance can exhibit a maximum from which one can determine experimentally such an important theoretical parameter as the coherence length in a given metal.

3.2. Comparison with Experiment

3.2.1. $\bar{L} \gg l_{\text{el}}$, $l_{\text{inel}} > L_2$. Figure 4 demonstrates the application of expression (6) (the solid curve) to a description of the temperature dependence of the effective resistance of an elastically deformed region located a distance $\bar{L}$ from the boundary that is a factor of 10^3

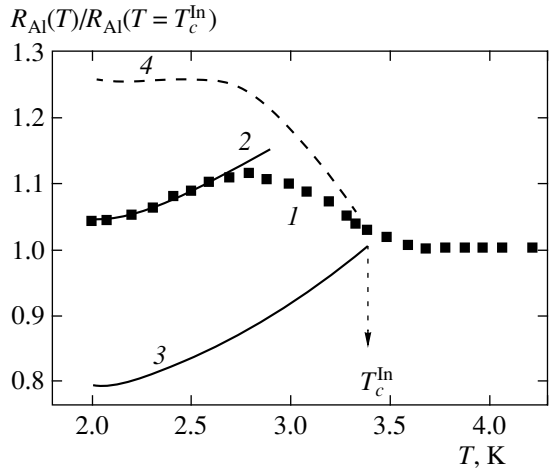


Fig. 5. Temperature dependence of the effective resistance (normalized to the value at $T = T_c^{In}$ of aluminum for $\bar{L} \sim l_{el}$, in comparison with the theory: (*1*) experiment; (*2*) theoretical (solid) curve, corresponding to expression (*7*); (*3*) the same curve normalized to $(R/R_N) = 1$; (*4*) the contribution to the growth of the resistance from trajectories of the type 2 from Fig. 3 in a layer of thickness $\hbar v_F/k_B T$ as a whole.

greater than the elastic mean free path (0.2 m), for a coefficient $A = 3.6 \times 10^{-4}$ [$K^{-3.5}$]. The good agreement of the theory and experiment makes it possible to estimate some of the parameters characterizing the quantum-interference effect wherein the conductivity increases under conditions of a strong interaction of the coherent excitations. The given value of implies that $\gamma \approx 45$. Measurements of the temperature dependence of the resistivity of aluminum at temperatures below 20 K give $n = 4$, $T^* \approx 10$ K, $l_{inel}^* = 3.5$ μm, $\beta \approx 10^2 l_{el}^{3/2}$ [$cm^{-1.5}$][$K^{-3.5}$]. This corresponds to the following estimates for the parameters: $\tau_r \sim 5 \times 10^{-4}$, $m^* \sim 5 \times 10^5$ for l_{inel} from 0.1 mm at 3 K to 0.44 mm at 2 K. If the number of reflections obtained is multiplied by the distance between nearest elastic centers, ΔL which for the case under consideration is of the order of $\sqrt{\sigma}\,(l_{el}/\sqrt{\sigma})^{1/3} \sim 10^{-2}$ (σ is the cross section of an elastic center, $\sqrt{\sigma}$ is of the order of the screening radius (2–5 Å), so that $m^* \Delta L \sim l_{inel}$, i.e., the main part of the length of the important coherent trajectories is indeed governed by the number of reflections, as we have assumed from the very beginning.

3.2.2. $\bar{L} \sim l_{el}$. We subtract the normal scattering contribution (curve *3* in Fig.2) from the temperature dependence of the resistance measured at the same probes after the elastic strain was removed and a large elastic mean free path had been restored (curve *1* in Fig. 2). The result of this subtraction, normalized to the value of the resistance at the temperature of formation of the boundary $T = T_c^{In}$, is shown by curve *1* (squares) in Fig. 5.

We apply Eq. (7) to the descending part of this curve. As we see in the figure, the theoretical curve *2*, with a coefficient $A = 3.87 \times 10^{-3}$ [$K^{-3.5}$] describes the experiment well. We decompose the experimental curve *1* into two competing components, consisting of the contributions (1) and (7). To do this we normalize Eq. (7) with the coefficient A determined above in such a way that $(R/R_N) = 1$ for $T > T_c^{In}$ (curve *3*) and subtract the resulting curve from curve *1* and add 1 to the difference. The result is shown by curve *4* in Fig. 5. It corresponds to the positive contribution (1) of coherent excitations to the conductivity of normal aluminum at distances from the boundary comparable to the elastic mean free path. The character of the curve is completely analogous to that which we were able to observe directly under similar conditions in an system with a different pair of metals: Cu–Sn [3,4]. It follows from the form of curve *4* that this contribution amounts to

$$\left[\frac{\delta R}{R_N}\right]_{\lambda_T} \approx \left[\frac{R}{R_N}\right]_{T < 2.5\ \mathrm{K}} - 1 \approx 20\%,$$

from which, in accordance with (1) for $\bar{L} \approx 125$ μm and $\bar{r} \sim 1$, it follows that $\lambda_T \approx 25$ μm at the theoretical value of the coherence length $(\hbar v_F/k_B T)_{A1} = 20$ μm for $T = 2.5$ K. The other parameters for the case considered are: $\gamma \approx 3.5$ and $\tau_r \approx 0.03$, from which we get $m^* \sim 10^2$ for the same value of the phase coherence length l_{inel} as in the "dirty" case discussed in Subsection 3.2.1.

4. CONCLUSION

For an aluminum-indium hybrid system with mechanical contact of the two metals, we have studied the temperature behavior of the transport in the normal aluminum in a temperature interval that includes the superconducting transition temperature of indium. The measurements were made at fixed probes near the contact in two regimes, in which the elastic mean free path of the electrons is much shorter than or is comparable to the distance from the probes to the contact. After the transition of the indium to the superconducting state and the formation of an boundary, two nonstandard types of behavior of the conductivity with decreasing temperature arose. In the first of the regimes indicated, which was brought about by increasing the concentration of elastic scattering centers by elastically deforming the sample, a decrease in the effective resistance was observed after formation of the boundary, while in the second regime the resistance was found to first increase and then decrease as the temperature was lowered. The change in conductivity is an order of magnitude or more greater than the usual change at helium temperatures in the absence of an boundary. It is shown that the observed phenomena completely correspond to the ideas about those features of the coherent transport which are due to the quantum-interference interaction of coherent excitations arising in Andreev reflections,

within the phase coherence region, which is limited by the scale of the inelastic mean free path. The observed features were used to estimate a number of microscopic parameters in aluminum—in particular, the phase and thermal coherence lengths.

REFERENCES

1. S. Washburn and R. A. Webb, Adv. Phys. **35**, 375 (1986).
2. A. G. Aronov and Yu. V. Sharvin, Rev. Mod. Phys. **59**, 755 (1987).
3. Yu. N. Chiang and O. G. Shevchenko, Zh. Éksp. Teor. Fiz. **113**, 1064 (1998) [JETP **86**, 582 (1998)].
4. Yu. N. Chiang and O. G. Shevchenko, Fiz. Nizk. Temp. **25**, 432 (1999) [Low Temp. Phys. **25**, 314 (1999)].
5. Yu. N. Tsyan, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 481 (2000) [JETP Lett. **71**, 334 (2000)].
6. J. Herath and D. Rainer, Physica C (Amsterdam) **161**, 209 (1989).
7. B. J. van Wees, P. de Vries, P. Magnic, and T. M. Klapwijk, Phys. Rev. Lett. **69**, 510 (1992).
8. A. Kadigrobov, R. Shekhter, and M. Jonson, Physica B (Amsterdam) **218**, 134 (1996).
9. Yu. N. Chiang, Prib. Tekh. Éksp., No. 1, 202 (1981).
10. R. Landauer, *Internal Memo* (IBM Thomas J. Watson Research Center, 1966).
11. G. E. Blonder, V. Tinkham, and T. M. Klapwijk, Phys. Rev. B **25**, 4515 (1982).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1242–1249.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1434–1442.

SOLIDS
Electronic Properties

Manifestations of the Pseudogap and Superconducting Gap of Cuprates in Photoemission

A. A. Ovchinnikov and M. Ya. Ovchinnikova
Semenov Institute of Chemical Physics, Russian Academy of Sciences, Moscow, 117977 Russia
Received June 5, 2000

Abstract—The degree to which the interpretation of the existence of a pseudogap and a superconducting gap in cuprates on the basis t–t'–U of the Hubbard–model corresponds to the data obtained from the photoemission spectra is discussed. The pseudogap in the model is interpreted as the work function of electrons from the insulating parts of the Brillouin zone boundary. On this basis one can explain the angle dependence of the gap measured in the photoemission spectra and its evolution on changes in doping and temperature. In particular, an explanation is found for the decline in the ratio of the angle derivative of the gap near the site, $v_\Delta = (1/2)d\Delta(\varphi)/d\varphi$, to the maximum value of the gap, Δ_{max}, with decreasing doping. That behavior and the different temperature dependence of the gap $\Delta(\varphi)$ for different angles are due to the presence of two contributions to Δ with different anisotropies—from the pseudogap and from the superconducting gap. The calculation of the spectral functions confirms the sharp Fermi boundary observed in the direction and the smeared edge of the distribution along the path $\Gamma(0, 0) – M(\pi, 0) – Y(\pi, \pi)$. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Many important results have now been obtained on the electronic structure of cuprates (see reviews [1–5]). These include the demonstration of the symmetry of the superconducting order, the discovery of anisotropy of the pseudogap of the normal state, the "small" Fermi surface of underdoped cuprates, the possible destruction of the Fermi surface in the region $k \sim (\pi, 0)$ [6, 7], and the direct measurement of the superconducting gap in tunneling spectroscopy [8–11]. Recently, a detailed study of the angle-resolved photoemission (ARPES) has revealed exotic temperature dependence of the gap and doping dependence of its anisotropy [15]. An apparent difference between these data and the characteristics obtained from measurements of the penetration depth was established. It was shown that the angle dependence of the gap $\Delta(\varphi)$ according to the ARPES data cannot be fitted by a simple function $\propto \cos(2\varphi)$ and that the form of $\Delta(\varphi)$ depends on the doping. In particular, the ratio $\xi = v_\Delta/\Delta$ of the derivative $v_\Delta = (1/2)d\Delta(\varphi)/d\varphi$near the site of the gap $\varphi = \pi/4$ to the maximum value of the interval Δ falls off appreciably with decreasing doping in underdoped samples. Finally, the standard form of the Fermi surface for overdoped bismuth ceramics was recently re-examined [16]. The ARPES data [16] were interpreted in the framework of a "large" Fermi surface with regions of the electronic type near the points $M(\pi, 0)$ instead of a continuation of a standard arc around the point $Y(\pi, \pi)$. Here the standard notation is used for the symmetric points in quasimomentum space $k = (k_x, k_y)$.

New experiments on the angle dependence of the gap could be of help for choosing among the various interpretations of the presence of a pseudogap [1]. According to one of the hypotheses that have been discussed, the pseudogap of the normal state is a precursor of superconducting pairing. However, the pairs formed are incoherent in a pseudogap normal state in the situation where phase fluctuations destroy the coherence of the pairs [17].

Another view of the origin of the pseudogap and the phase diagram in cuprates has arisen from studies of the Hubbard t–t'–U and t–t'–J models [18–22]. It attributes the appearance of the pseudogap to a specific structure of the split Hubbard bands, namely, to a change in the topology of the Fermi surface upon optimal doping. We shall call this mechanism of formation of the pseudogap the structural-band mechanism. Various approaches are used to describe the correlation mechanism of superconducting pairing. One such approach is the spin-polaron approximation in the t–t'–J model [18]. Another, the variational approach, takes into account the correlation of the type of valence bonding (the band analog of Anderson's RVB states) in the Hubbard t–t'–U model. Since in this method the structure of the quasiparticles is characterized in explicit form, it is the most transparent method. Common to each of these models is a Hubbard splitting of an initially unified band due to long-range antiferromagnetic spin correlations. Here the fine details of the upper edge of the lower Hubbard subband, which depend on t', determine the structure of the Fermi surface and govern the low-energy phenomena. Here is the hopping interaction of the next-nearest neighbors located along the diagonal. In particular, plays an important role in the formation of the pseudogap of the normal state, and the sign of determines its anisotropy. For systems with $t'/t > 0$, the coin-

cidence of the position of the Van Hove singularity in the density of states with the chemical potential corresponds precisely to the optimal doping δ_{opt}. It is at this value of the doping that the change of the topology of the Fermi surface occurs. At the transition from underdoped to overdoped systems the hole pockets around the points $k = (\pi/2, \pi/2)$ and the "small" Fermi surface give way to a large Fermi surface. For δ_{opt}, in the underdoped region, the Van Hove singularity in the spectrum of states lies below the chemical potential μ. As a result, certain parts of the boundary of the magnetic Brillouin zone near $k_M = (\pi, 0)$ turn out to be insulating regions. According to [20], the work function for the electrons that are removed in photoemission from the levels in the regions of this kind is just the pseudogap of the normal state observed in ARPES and other experiments.

Another criterion for checking the models is the shape of the Fermi boundary and the distribution function of the electrons in photoemission from states with different directions of the quasimomentum. Up till now the topological change predicted in the t–t'–U and t–t'–J models has been confirmed only by the observation of a "small" Fermi surface in the region $\delta < \delta_{opt}$. However, the predicted crossing of the segment $\Gamma(0, 0) - M(\pi, 0)$ near the point by the Fermi boundary for systems with $\delta > \delta_{opt}$ has not been confirmed. According to the models of [18–22], small Fermi-boundary regions of the electronic type arise near the point for $\delta > \delta_{opt}$. However, it is in this region of quasimomenta that the smearing of the distribution functions of the photoelectrons creates ambiguities in the interpretation of the ARPES data. In particular, a new interpretation of the ARPES data for BiBCCO at $\delta > \delta_{opt}$ apparently confirms the crossing of the segment $\Gamma - M$ by the Fermi boundary [16]. In view of the remaining uncertainty, it is of interest to study the properties of the one-electron spectral functions of the t–t'–U model and the influence of spin fluctuations on them even in the case of the simplest solutions.

The goal of the present study is to understand the degree to which the model based on a structural-band origin of the pseudogap can describe new experiments on the anisotropy of the gap as determined from the photoemission data. For this purpose in Section 2 we calculate the doping dependence and temperature dependence of the anisotropy of the total gap resulting from the combined contributions from the superconducting gap and the pseudogap. The calculation is done on the basis of the simplest variational approach to the description of the correlated state [20]. The initial state is one with correlations of the type of valence bonding and antiferromagnetic spin order, reflecting the long-range spin fluctuations. In Section 3 the manifestations of the Fermi boundary in the single-particle spectral functions n_k and $A(k, \omega \longrightarrow 0)$ and the shape of the Fermi surface in different directions of are discussed for the simplest antiferromagnetic and helical spin states of the mean field model. The high sensitivity of the position of the Fermi boundary in the region $k \sim (\pi, 0)$ to spin fluctuations is in qualitative correspondence with the observed destruction of the Fermi boundary in this region [6, 7], while a sharp Fermi boundary is preserved in the direction $k_x = k_y$.

2. DEPENDENCE OF THE ANISOTROPY OF THE GAP ON DOPING AND TEMPERATURE

Calculations were done for the t–t'–U–V Hubbard model by the method proposed in [20]. The formula of the model has the form

$$H = H(U, t) + \Delta H(V, t'),$$

$$\Delta H(V, t') = V \sum_{\langle nm \rangle} n_n n_m + t' \sum_{\langle\langle nm \rangle\rangle} \sum_{\sigma} (c^{\dagger}_{n\sigma} c_{m\sigma} + \text{H.c.}). \quad (1)$$

Here $H(U, t)$ is the Hamiltonian of the basic Hubbard model with a one-center interaction and a hopping integral for neighboring centers. The additional term ΔH includes the hopping interaction of next-nearest (diagonal) neighbors and an interaction of the Coulomb type between nearest sites.

There are deep reasons for including the interaction t'. For a one-band model of the CuO plane [23, 24] this parameter is determined by the competition of the contribution from direct hops between the orbitals of oxygen ($\sim t_{pp}$) and the contribution from the second-order process $\sim t_{pd}^2/(\epsilon_p - \epsilon_d)$. As a result, the parameter turns out to be sensitive to the parameters of the initial three-band model and can be different for different cuprates [25]. Another aspect is the strong influence of on the structure of the bands near the Fermi boundary and on the low-energy properties of the system [18–22].

The variational correlation state Ψ with correlations of the type of the valence bonds is constructed [20] with the aid of a unitary transformation of the uncorrelated state Φ:

$$\Psi = \hat{W}(\alpha)\Phi, \quad \hat{W}(\alpha) = \exp\left[\alpha \sum_{\langle nm \rangle} Z_{nm}\right],$$
$$Z_{nm} = \frac{1}{2} \sum_{\sigma} (c^{\dagger}_{m\sigma} c_{n\sigma} + \text{H.c.})(n_{n-\sigma} - n_{m-\sigma}). \quad (2)$$

For Φ we consider an uncorrelated state of the most general type, viz., a function of the BCS type with anomalous averages of symmetry and with a doubled magnetic cell for testing the possibility of antiferromagnetic spin ordering and superconductivity in the channel. The unitary operator $W(\alpha)$ responsible for the formation of the valence bonds depends on the varia-

tional parameter α, and the operator Z_{nm} in Eq. (2) pertains to the bond $\langle nm \rangle$ of nearest centers.

The meaning of this transformation and its efficacy in lowering the energy of the system can be explained for the example of a two-center system with molecular orbitals $q_\sigma(u_\sigma) = (a_\sigma \pm b_\sigma)/\sqrt{2}$. In this case the operator

$$W = \exp(\alpha Z_{ab}) \equiv \exp[\alpha(g_\uparrow^\dagger g_\downarrow^\dagger u_\downarrow u_\uparrow - \text{H.c.})]$$

transforms the uncorrelated two-hole state $F(a, b) = g_\uparrow^\dagger g_\downarrow^\dagger >$ H.c. to an exact singlet state Ψ of the dimer for the corresponding optimum value of the parameter α. In the case of a large ratio U/t this state describes a localized singlet of the valence bond. At medium values of it becomes possible to optimize the charge state of the bond or the degree of localization of holes at sites of the bond.

The unitary transformation (unlike a nonunitary transformation of the Gutzwiller type) makes it possible to derive an effective Hamiltonian $\tilde{H}$ that operates in the basis of uncorrelated states $\{\Phi\}$ and to obtain an exact value for the average energy:

$$\begin{aligned} \tilde{H}(\alpha) &= W^\dagger(\alpha) H W(\alpha), \\ \overline{H} &= \langle \Psi H \Psi \rangle = \langle \Phi \tilde{H} \Phi \rangle. \end{aligned} \tag{3}$$

Thus one can use a self-consistent procedure of minimizing $\overline{H}$ with respect to Φ and α simultaneously. Such a treatment of the Hamiltonian (3) in the mean field approximation determines the electronic bands $E_{k,\lambda}$ and the single-particle Fermi operators $\chi_{k,\lambda}^\dagger$, which are the eigenvalues and eigenoperators of the localized effective Hamiltonian $(\overline{H})_{\mathrm{L}}$ [20]. In contrast to the dimer structure with nonoverlapping valence bonds [26], in the case of a homogeneous state of the valence bonds (2) the effective Hamiltonian was found to the two lowest orders in α:

$$\hat{H}(\alpha) \approx H + \alpha[H, Z] + \frac{\alpha^2}{2}[[H, Z], Z]. \tag{4}$$

Thus the treatment is limited to Hubbard models with $U/t \leq 9$, which correspond to small values of the optimal parameter $\alpha \leq 0.22$. The detailed procedure of the solution is given in [20]. Practically all of the calculations were done using the effective Hamiltonian $\tilde{H} = \hat{H}(\alpha, U, t) + \Delta H(V, t')$, in which only the principal part $H(U, t)$ was subjected to the transformation $H(U, t)$, while for small interactions ΔH in (1) only the zeroth order in α is retained.

The attraction of holes in the channel is induced mainly by the correlated hopping terms,

$$\sim \alpha U[c_{n\sigma}^\dagger c_{m\sigma} n_{n,-\sigma} n_{m,-\sigma} + \text{H.c.}], \tag{5}$$

that arise in the effective Hamiltonian upon the formation of valence bonds. Unlike the empirical interaction of correlated hops, such as that introduced by Hirsch [27], the interaction constant in (5), which is proportional to α, is determined by a variational procedure. Further, unlike the t–J model, the application of the mean field approximation to the new effective Hamiltonian (4) is admissible, since here doubly occupied configurations of the centers are not strictly forbidden, as they are in the t–J models. Minimization of the energy with respect to α optimizes the charge configurations of the bonds and the degree of localization of charges on them [20]. The band energies $E_{k\lambda}$ (eigenvalues of the linearized effective Hamiltonian $\tilde{H}_{\mathrm{L}}$) correspond to the self-consistent solution Φ and to the optimum value of α.

The investigated homogeneous solutions exhibit two-dimensional antiferromagnetic spin order in a wide doping region, $\delta = |1 - n| \leq \delta_c \sim 0.3$, which is substantially greater than the region of bulk antiferromagnetic order ($\delta_c \sim 0.05$). This means that the finite antiferromagnetic spin correlation length is considerably larger than the lattice constant. Under such conditions the mean-field solutions and the picture of the antiferromagnetically split band can serve as a basis for describing the structure of the quasiparticles and low-energy phenomena. In the normal state the lower Hubbard subband has the following structure [20]:

$$\begin{aligned} E_k &= [\epsilon_0 + 4t' c_x c_y + \ldots] \\ &- \sqrt{[U d_0 + \ldots]^2 + [2t(c_x + c_y) + \ldots]^2}. \end{aligned} \tag{6}$$

Here only the leading contributions from the lower harmonics $c_{x(y)} = \cos k_{x(y)}$ are written in explicit form, and $d_0 = \langle (-1)^n S_{zn} \rangle$ is the value of the alternating spin. In a superconducting state of symmetry the quasiparticle energies in the self-consistent solution are very close to the form

$$\begin{aligned} E_{k\lambda} &= \pm\sqrt{(E_k - \mu)^2 + W_k^2}, \\ W_k &\sim F_k[k_{11}(c_x - c_y) w_1 + \ldots]. \end{aligned} \tag{7}$$

Here W_k is determined by the superconducting order parameter $w_l = (l_x^2 - l_y^2)\langle c_{n\uparrow}^\dagger c_{n+l,\uparrow}^\dagger \rangle$, $|l| = 1, \sqrt{5}, 3$. The form factor F_k is determined by the structure of the quasiparticles, which are single-particle states of the lower Hubbard band, and the constants $k_{i,j}$, the second derivatives of the mean energy with respect to and w_i, w_j. A calculation with all the contributions exactly taken into account shows that the first harmonic w_1 gives the main contribution to the superconducting gap function W_k and that its angle dependence is close to $W_k \sim (c_x - c_y)$.

The doping dependence of the superconducting transition temperature $T_c(\delta)$ for the Hubbard model with the parameters $U/t = 8$, $t'/t = 0.05$, and $V/t = 0.1$ is

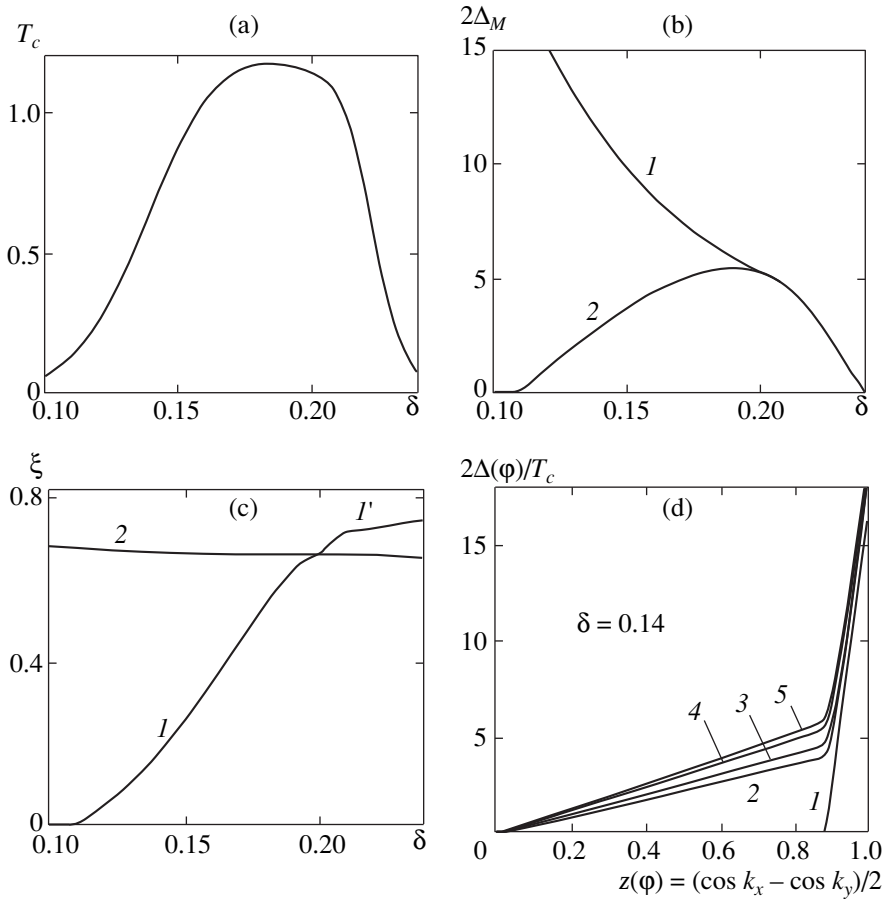


Fig. 1. (a) Doping dependence of the superconducting transition temperature $T_c(\delta)$ in units of $[0.01t]$. (b) Twice the value of the shift of the photoemission edge $2\Delta_M(\delta)$ as a function of doping for $k = (\pi, 0)$ (*1*) and the corresponding superconducting gap $2W_k(\pi, 0)$ (*2*) in the same units. (c) Ratio $\xi = v_\Delta/\Delta_M$ (*1*) and $\xi_W = v_\Delta/W_k(\pi, 0)$ (*2*) as a function of doping. (d) Angle dependence of the gap $\Delta(\varphi)$ at different temperatures for an underdoped system with $\delta = 0.14$, $T_c/T_c^{\max} = 0.55$ at $T/T_c = 0.4$ (*1*), 0.6 (*2*), 0.8 (*3*), 0.9 (*4*), 1.2 (5).

shown in Fig. 1a. We note that for the value of the maximum temperature $T_c^{\max}$ increases by a factor of 1.64 [20]. Curve *1* in Fig. 1b shows the doping dependence of $W_M(\delta) = W(k_M)$, the value of the superconducting gap W_k at the point $k = k_M = (\pi, 0)$. The calculation was done at a temperature $T/t = 0.002$. According to Eq. (7), however, the quasiparticle energy and the corresponding shift $\Delta = \Delta\omega$ of the edge of the photoelectron distribution function at $k = (\pi, 0)$ is determined not only W_k by but also by the corresponding band energy $|E_k - \mu|$ of the lower Hubbard band relative to the chemical potential. Unlike the unsplit band E_k, the levels of the lower Hubbard subband are periodic within the magnetic Brillouin zone. The functions E_k pass through a maximum at the boundary of this zone as is varied k along any path connecting the points $\Gamma(0, 0)$ and $Y(\pi, \pi)$. This makes for a different doping dependence of the ratio $\xi = v_\Delta/\Delta_M$ in the under- and overdoped regions (Fig. 1c). Here $v_\Delta = (1/2)d\Delta/d\varphi$ is the angle derivative of the gap at $\varphi = \pi/4$, and Δ_M is the value of the gap Δ at $k = (\pi, 0)$.

For the overdoped regime each $\Gamma - Y$ path crosses the "large" Fermi boundary: $E_k - \mu = 0$ at $k = k_F$, and at $E_k - \mu > 0$ all values k of on the boundary of the magnetic Brillouin zone. In this case the gap $\Delta(\varphi)$ in the excitation spectrum is determined solely by the *d*-wave superconducting gap W_k at the Fermi boundary $k = k_F$. Consequently, the anisotropy of the gap $\Delta(\varphi)$ coincides with $\Delta(\varphi) = W_k(\varphi) \sim z(\varphi) = 0.5(\cos k_x - \cos k_y)$ at $k = k_F$. The corresponding ratio $\xi = v_\Delta/\Delta_M$ of v_Δ to the gap measured at the point $k = k_M$ is given by $\xi = (1/2)(dW_k/d\varphi)/W_M$. It is close to the constant value expected for a purely-wave superconducting order without a pseudogap.

In underdoped systems with $t'/t > 0$, hole pockets form in the region $k = (\pi/2, \pi/2)$ with a "small" Fermi boundary around them. Here the segments of the magnetic Brillouin zone near $M(\pi, 0)$ become insulating

segments of the boundary, on which $E_k - \mu < 0$. The shift $\Delta(k_M)$ in the photoelectron distribution function in ARPES at $k_M = (\pi, 0)$ here is given by $\Delta(k_M) = \sqrt{\Delta_N^2 + W_M^2}$. Here W_M is the value W_k of at k_M, and the quantity $\Delta_N = |E_k - \mu|$ at k_M is the pseudogap of the normal state, which is equal to the work function for electrons being removed from levels $k \sim k_M$ of the normal state. For bands described by expression (6), the value of Δ_N increases with decreasing doping. For the same reason the total shift $\Delta(\varphi) = \sqrt{\Delta_N^2(\varphi) + W_k^2(\varphi)}$ of the photoemission edge from the insulating parts of the magnetic Brillouin zone differs substantially from the anisotropy of the superconducting gap $W_k(\varphi)$. For the underdoped regime ($\Delta_N \neq 0$) we have

$$\xi \sim 0.5[W'(\pi/4)]/\sqrt{\Delta_N^2 + W_M^2}.$$

where $W'(\varphi) = dW_k/d\varphi$ at $\varphi = \pi/4$. The value of ξ falls off sharply ($W_k \longrightarrow 0$) as the doping decreases, in complete agreement with the experimental behavior of $\xi(\delta)$ [15]. Unlike $\xi(\delta)$ the quantity $\xi_W(\delta) = v_\Delta/W_M$ calculated with allowance for the superconducting gap in the numerator remains nearly constant (curve *2* in Fig. 1c).

Figure 1d shows the anisotropic shift $\Delta(\varphi)$ of the photoemission edge for highly underdoped systems ($\delta = 0.14$, $T_c/T_c^{\max} = 0.55$). The function $\Delta(\varphi)$ was calculated as the vector k moved along a generalized Fermi boundary consisting of the nonshadow parts of the "small" Fermi boundary and the insulating segments of the boundary of the magnetic Brillouin zone (see [20]). A number of features of the calculated anisotropy are in agreement with the anisotropy measured in ARPES experiments.

(1) The deviation of $\Delta(\varphi)$ from a linear dependence, $\Delta(\varphi) \sim z(\varphi) = |\cos k_x - \cos k_y|$, is actually observed in Bi-Sr-Ca-Cu-O (BSCCO) in the underdoped regime [15], and the signs of the coefficient in front of the next harmonic in $\Delta(\varphi)$ are the same for the calculation and experiment.

(2) For underdoped systems the function Δ_k turns out to be more sensitive to changes in temperature at values of in the region k near the sites of the gap than for $k \sim (\pi, 0)$. The shift $\Delta(\varphi)$ for $\Delta(\varphi) \sim \pi/4$ vanishes as $T \longrightarrow T_c$, whereas the pseudogap of the normal state in the region $k \sim (\pi, 0)$ retains a nonzero value, $\Delta_M = \Delta_N = \mu - E(\pi, 0) > 0$, at $T > T_c$. This property also corresponds to experiment [12, 13], which showed a different temperature sensitivity of $\Delta(\varphi)$ for different φ. The calculation gives a sharp bend in the linear dependence $\Delta(z(\varphi))$, unlike the smoothed curve in the experiment. The reason may be that the dispersion with k_z respect to was neglected in the calculation, as was the finite resolution with k respect to in the experiment.

(3) The calculated ratio $\eta = 2\Delta_M/kT_c$ has larger values in the underdoping region. For example, for $\delta = 0.16$ ($T_c = 0.9T_c^{\max}$) it has the value $\eta = 9.5$. At the same time, at optimum doping and in the overdoped region the total gap is determined solely by the superconducting part, $\Delta = W(k_F)$, and the values $\eta \approx 5$ obtained for the ratio are close to the value $\eta = 4.5$ obtained in the empirical BCS models with *d*-wave pairing. The decrease of the ratio $2\Delta/kT_c$ with increasing doping is in agreement with experiment [1].

Independent estimates of the anisotropy of the gap near its sites are obtained in [15] from data for the superconducting hardness ρ_s at low temperatures on the basis of the relation between v_Δ and the penetration depth for $T \longrightarrow 0$. This relation $d\rho_s(T)/dT \sim d(\lambda^{-2})/dT \sim 1/v_\Delta$, gives a doping dependence of v_Δ that differs from the mutually agreeing functions $v_\Delta(\delta)$ obtained from the ARPES data and our calculations. The reasons for the differences remain unclear.

3. REFLECTION OF THE FERMI BOUNDARY IN THE PHOTOELECTRON SPECTRA AND THE INFLUENCE OF SPIN FLUCTUATIONS

To check the structural-band origin of the pseudogap which was proposed in [18–22] and to confirm the topological rearrangement of the Fermi surface upon optimum doping, it is of interest to study how the change in the topology of the Fermi surface and the appearance of insulating regions on the boundary of the magnetic Brillouin zone in the underdoped region are manifested in the density of states and in the features of the single-electron spectral functions. Such a study, even in the framework of the simplest mean-field solutions for the *t*–*t*'–*u* model, can be useful.

Another goal of such calculations is to illustrate the influence of spin fluctuations on the photoelectron spectra for different directions of the quasimomentum. The strong influence of the low-frequency collective spin modes is well known [28]. In the simplest version the spin waves are described by states with a helical spin structure. In this section we therefore carry out a comparative study of the density of states and the spectral reflections of the band structure for antiferromagnetic and helical states of the *t*–*t*'–*U*–*V* Hubbard model (1). This study is of a qualitative character. For this reason and for the sake of simplicity we restrict the calculations to the mean field approximation without allowance for correlations of the type of valence bonds.

The standard helical state Ψ_Q with helicity vector $\mathbf{Q}$ is a mean-field state with one-electron averages $\langle c_{n\uparrow}^\dagger c_{n\downarrow}\rangle = b_0 \exp[iQn]$. For $Q \longrightarrow \mathbf{Q}_{AF} = (\pi, \pi)$ it goes over to an antiferromagnetic state. Neglecting correlations of the type of the valence bonds has a substantial effect on the energy of the system and causes the critical doping level, at which the antiferromagnetic (or helical) mean-field solutions go over to paramagnetic solutions, to shift to higher values [20]. For the same reason, minimization of the energy with respect to at a

given doping δ overestimates the difference between the optimum vector $Q(\delta)$ and $Q_{AF} = (\pi, \pi)$. For this reason the calculations and the comparative discussion of the density of states and the spectral functions are given for the helical states with an arbitrary helicity vector without a minimization of the energy with respect to at each δ.

For mean-field states without allowance for the correlations of the type of valence bonds and without superconducting pairing, we determined the regions of under- and overdoping of the t–t'–U–V model as those regions for which the Van Hove singularity in the system of states lies below or above the chemical potential of the system. Then the "optimal" doping corresponds to the situation when the position of the Van Hove singularity is precisely coincident with the chemical potential μ. For the calculated model with parameter $U/t = 8$, $t'/t = 0.5$, and $V/t = 0.1$ such a determination gives a value $\delta_{opt} \sim 0.3$ for the optimum doping in the case of an antiferromagnetic state. This value is greater than $\delta_{opt} \sim 0.19$ for the same system in the antiferromagnetic state with valence bonds, for which the correlations of the type of valence bonds was taken into account.

Figure 2 shows the density of states of the lower Hubbard band for the antiferromagnetic and helical states for different directions of the helicity vector with Q, $|Q - Q_{AF}| = 0.3$. The density of states with a single Van Hove singularity for the antiferromagnetic state transforms into a density of states with split singularities for the helical states. The formation of the spin structure with a preferred direction lifts the degeneracy of the Van Hove singularity for $k = (\pm\pi, \pm\pi)$ in analogy with how the lattice distortions lift this degeneracy in the ordinary Jahn–Teller effect [3].

Below is a comparative analysis of the influence of the spin structure on the occupation function

$$n_k = \frac{1}{2}\sum_{\sigma} \langle c^{\dagger}_{k\sigma} c_{k\sigma} \rangle$$

in momentum space and the single-particle spectral functions $A(k, \omega)$ at low frequencies $\omega \longrightarrow 0$:

$$A(k, \omega = 0) = \frac{1}{ZN}\sum_{\sigma}\sum_{i,\,f} |\langle f|c_{k\sigma}|i\rangle|^2 \times \exp\left(-\frac{E_i}{kT}\right)\frac{\gamma}{\pi[(E_i - E_f - \omega)^2 + \gamma^2]}\Bigg|_{\omega = 0}. \quad (8)$$

Here E_i and E_f are the energies of the initial and final states, Z is the partition function of the system, and the standard δ function $\delta(E_i - E_f - \omega)$ has been replaced by a Lorentzian of width γ.

Figure 3 shows the functions n_k and $A(k, \omega = 0)$ calculated for k varying along the contour $Y(\pi, \pi) - \Gamma(0, 0) - M(\pi, 0) - Y$ for mean-field states of the t–t'–U–V model for two doping values—in the underdoped region $\delta < \delta_{opt}$ and in the overdoped region $\delta > \delta_{opt}$. Here the optimal doping $\delta_{opt} = 0.3$. The model parameters are $U/t = 8$, $t'/t = 0.5$, $V/t = 0.1$ and the width in (8) $\gamma = 0.003t$.

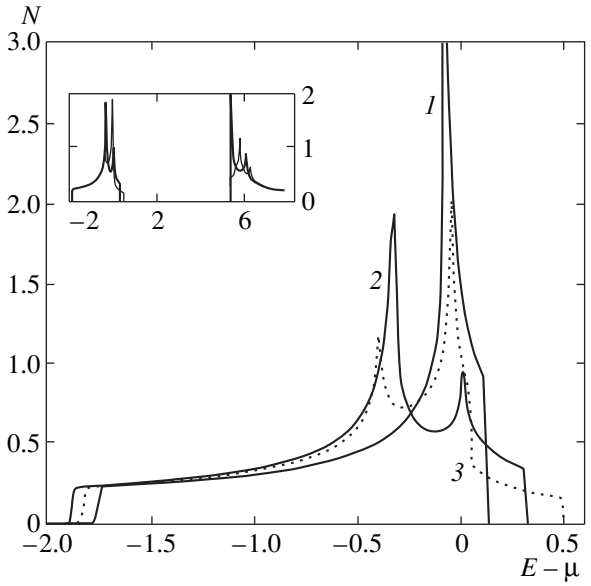


Fig. 2. Density of states of (*1*) the lower Hubbard band for an antiferromagnetic mean-field state and for helical states with (*2*) $Q_2 = (0.9\pi, \pi)$ and (*3*) $Q_3 = (0.93\pi, 0.93\pi)$.

Some of the properties of the spectral functions obtained are partially known.

Even in antiferromagnetic states, for which the band energy is periodic within the magnetic Brillouin zone, the spectral functions n_k and $A(k, \omega)$ do not possess such periodicity. Fine details aside, we see that the occupied $n_k > 0.5$ and empty $n_k < 0.5$ quasimomentum states correspond to regions around the Γ or Y points, respectively, in analogy with the case of noninteracting particles. This property is well known [29] and holds true for the spiral states as well.

On the segment $\Gamma - Y$ the sharpest falloff of occurs for and corresponds to the crossing of the nonshadow Fermi boundary (i.e., located in the main magnetic Brillouin zone) with the line $\Gamma - Y$. Putting in a helical spin order instead of the antiferromagnetic order does not alter the position of the Fermi boundary in this direction. The corresponding peak in $A(k, \omega = 0)$ is the most intense one for any doping and is not affected by spin fluctuations, since it does not depend on the helicity vector **Q**. This finding agrees with the fact that the experimentally observed Fermi boundary in the $\Gamma - Y$ direction is well defined for any doping [6]. The second peak of the function $A(k, \omega = 0)$ in this same direction, corresponding to a step in n_k, pertains to the shadow Fermi boundary. The intensity of this peak is much lower, and its position depends substantially on the helicity vector. This means that this feature will be smoothed by spin fluctuations.

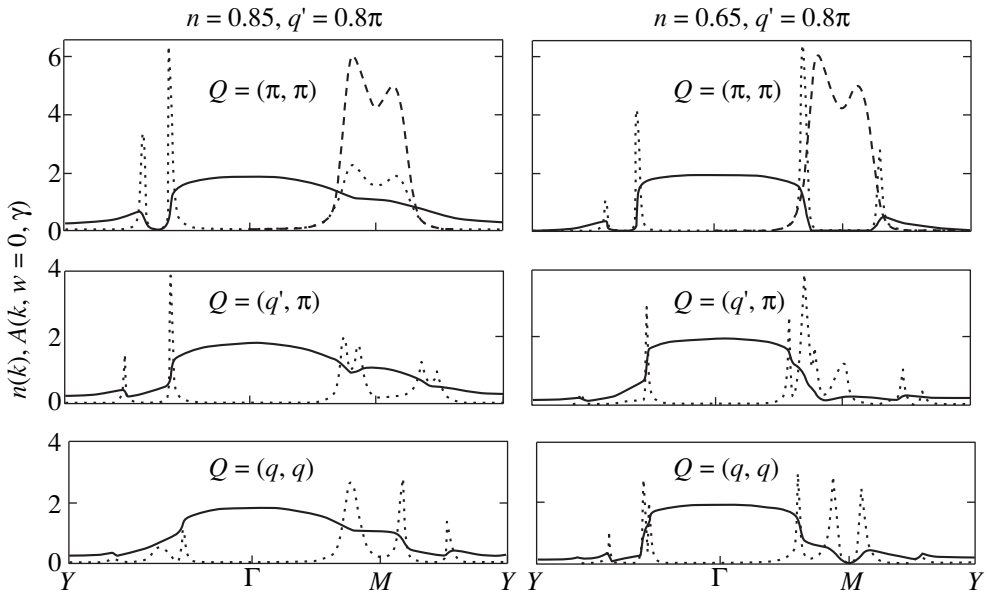


Fig. 3. Occupation functions n_k (solid curves) and $A(k, \omega = 0, \gamma)$ for $\gamma = 0.003t$ (dotted curve) for the antiferromagnetic and helical states of models that take into account the underdoping (left) and overdoping (right). The vector **k** varies along the path $T - \Gamma - M - Y$. For the helical states the vector **Q** was specified by the values $q' = 0.8\pi$ and $q = 0.86$. The dashed curves on the segments $\Gamma - M - Y$ (upper plots) correspond to the antiferromagnetic state at nearly "optimal" doping, $\delta = 0.3$.

Another situation takes place in the region of the point $M(\pi, 0)$. The occupation function n_k varies smoothly with k along $\Gamma - M - Y$, and the low-frequency excitation peaks in $A(k, \omega = 0)$ have a rather high energy in a wide region k of near the point M. The maximum intensity is reached at the optimum doping. The same situation, but with a lower intensity, takes place at any doping: both for underdoped systems, when in the region k in the vicinity of the point M becomes insulating ($E(\pi, 0) - \mu < 0$), and for overdoped systems, when the point lies between the main and shadow Fermi boundaries ($E(\pi, 0) - \mu > 0$). This smearing (or destruction) of the Fermi boundary in the region of the point has been observed in the ARPES data [6, 7]. A large total photoemission intensity from the region $k \sim (\pi, 0)$ has in fact been observed [16] in BSCCO, but it turned out to depend on the photon energy, a circumstance that cannot be described on the basis of two-dimensional models alone. The model must be generalized to take into account the interlayer interaction.

4. CONCLUSIONS

We have shown on the basis of the t–t'–U Hubbard model that the hypothesis of a structural-band origin of the pseudogap can explain a number of features in the behavior of the anisotropy of the gap $\Delta(\varphi)$ observed in photoemission experiments. The change in topology of the Fermi surface at optimum doping, which is due to the structure of the edge of the low Hubbard band, causes the flat parts of the band in underdoped systems to become insulating, and a pseudogap $\Delta_N(\varphi)$ of the normal state opens up. This pseudogap and the superconducting gap $\Delta_{SC}(\varphi)$ together determine the total shift $\Delta(\varphi)$ of the photoemission edge. The different dependence of each of the components Δ_N and Δ_{SC} on temperature, doping, and the direction k of make it possible to describe the following observed features of the anisotropy of the gap in the underdoped region.

(1) The presence of higher harmonics in $\Delta(\varphi)$, i.e., a deviation from a simple d-wave form $\propto [\cos k_z - \cos k_y]$.

(2) the appreciable falloff, with decreasing doping, of the ratio $\xi = v_\Delta/\Delta$ of the slope of the angle dependence of the gap $v_\Delta = (1/2)d\Delta(\varphi)d\varphi$ near the site of the gap $\varphi = \pi/4$ to the maximum value of the gap Δ.

(3) The stronger temperature dependence of the gap in the region of its site ($\varphi = \pi/4$) than in the region of the maximum of the gap.

(4) The increase in the ratio Δ/T_c with decreasing doping.

The spectral functions for the antiferromagnetic and helical mean-field states of the t–t'–U model qualitatively confirm the presence of a sharp Fermi boundary in the region of the site of the gap and a smeared Fermi boundary in the region $k \sim (\pi, 0)$, with a sharp increase in the total photoemission intensity from this region at optimum doping.

The main unsolved problem for this interpretation of the origin of the pseudogap is to match it up with the observed tunneling spectra. The difficulties of such a matching in the underdoped region have been discussed in [30]. In the t–t'–U model the tunneling gap Δ_{tunn} is due solely to the superconducting contribution, while the pseudogap controls only the asymmetry of the spectra outside the optimum doping region. On the contrary,

the observed tunneling gap in underdoped compounds coincides with the pseudogap [8–11].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 00-03-32981 and 00-15-97334). The authors are also grateful to V.Ya. Krivnov and N.M. Plakida for stimulating discussions of the problem.

REFERENCES

1. T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
2. Z. M. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
3. R. S. Markiewicz, J. Phys. Chem. Solids **58**, 1179 (1997).
4. N. M. Plakida, Philos. Mag. B **76**, 771 (1997).
5. Yu. A. Izyumov, Usp. Fiz. Nauk **42** (3), 215 (1999).
6. N. M. Norman, H. Ding, M. Randeria, *et al.*, cond-mat/9710163 (1997).
7. N. M. Norman, cond-mat/9904048.
8. T. Hasegawa, H. Ikuta, and K. Kitazawa, in *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Ginsberg (World Scientific, Singapore, 1992), Vol. III, p. 525.
9. Ch. Renner, B. Revaz, J.-Y. Genoud, *et al.*, Phys. Rev. Lett. **80**, 149 (1998).
10. A. Matsuda, S. Sugita, and T. Watanabe, Phys. Rev. B **60**, 1377 (1999).
11. Y. DeWilde, N. Miyakawa, P. Guptasarma, *et al.*, Phys. Rev. Lett. **80**, 153 (1998).
12. J. M. Harris, P. J. White, Z. X. Shen, *et al.*, Phys. Rev. Lett. **79**, 143 (1997).
13. J. M. Harris, Z.-X. Shen, P. J. White, *et al.*, Phys. Rev. B **54**, R15665 (1996).
14. P. J. White, Z.-X. Shen, C. Kim, *et al.*, Phys. Rev. B **54**, R15669 (1996).
15. J. Mesot, N. R. Norman, M. Randeria, *et al.*, Phys. Rev. Lett. **83**, 840 (1999).
16. Y.-D. Chuang, A. D. Gromko, D. S. Dessau, *et al.*, Phys. Rev. Lett. **83**, 3717 (1999).
17. V. J. Emery and S. A. Kivelson, Nature **374**, 4347 (1995).
18. N. M. Plakida, V. S. Oudovenko, R. Horsch, and A. J. Liechtenstein, Phys. Rev. B **55**, 11997 (1997).
19. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Phys. Lett. A **249**, 531 (1998).
20. A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 350 (1998) [JETP Lett. **67**, 369 (1998)]; Zh. Éksp. Teor. Fiz. **114**, 985 (1998) [JETP **87**, 534 (1998)]; Zh. Éksp. Teor. Fiz. **115**, 649 (1999) [JETP **88**, 356 (1999)].
21. R. O. Kuzian, R. Hayn, A. F. Barabanov, and L. A. Maksimov, Phys. Rev. B **58**, 6194 (1998).
22. F. Onufrieva, P. Pfeuty, and M. Kiselev, Phys. Rev. Lett. **82**, 2370 (1999).
23. J. H. Jefferson, H. Eskes, and L. F. Feiner, Phys. Rev. B **45**, 7959 (1992).
24. H. B. Schuttler and A. J. Fedro, Phys. Rev. B **45**, 7588 (1992).
25. D. Duffy and A. Moreo, Phys. Rev. B **52**, 15607 (1995).
26. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **110**, 342 (1996) [JETP **83**, 184 (1996)]; Zh. Éksp. Teor. Fiz. **112**, 1409 (1997) [JETP **85**, 767 (1997)].
27. J. E. Hirsch, Phys. Rev. Lett. **54**, 1317 (1985).
28. J. Schmalian, D. Pines, and B. Stojkovic, Phys. Rev. B **60**, 667 (1999).
29. Z.-X. Shen and J. R. Schriefer, Phys. Rev. Lett. **78**, 1771 (1997).
30. A. A. Ovchinnikov and M. Ya. Ovchinnikova, cond-mat/9912096 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1250–1260.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1443–1455.

SOLIDS
Electronic Properties

Piezoelectric Effect and the Long-Term Resistance Relaxations Induced by Uniaxial Compression in *p*-GaAs/$Al_xGa_{1-x}As$ Heterostructures

V. N. Kravchenko[a, *], N. Ya. Minina[a], A. M. Savin[a], and O. P. Hansen[b]

[a]*Moscow State University, Moscow, 119899 Russia*
**e-mail: krav@mig.phys.msu.su*
[b]*Oersted Laboratory, Niels Bohr Institute, DK-2100 Copenhagen, Denmark*

Received July 13, 2000

Abstract—Long-term resistance relaxations induced by uniaxial compression in (001)*p*-GaAs/$Al_{0.5}Ga_{0.5}As$ heterostructures are observed, and the main properties of these relaxations are investigated: the dependence of their character on the direction of the uniaxial compression, the change in concentration of current carriers during the relaxation processes, and the quenching of the relaxations by temperature, illumination, and high electric fields. It is found that the character of the relaxation process is different for compression directions [110] and $[1\bar{1}0]$: in the first case the concentration of two-dimensional holes in the quantum well decreases in the course of the relaxation, while in the second case it increases. A model is proposed in which the cause of the relaxations of the piezoresistance and the anisotropic character of these relaxations is assumed to be the piezoelectric field, the value of which, according to estimates, is $|E_z| = 1.152 \times 10^4$ V/cm per kbar. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The long-term relaxations of the resistance in semiconductors and heterostructures based on them are a manifestation of nonequilibrium processes in the system of charges and have attracted steady interest for several decades now. It was noted long ago [1, 2] that this phenomenon is due to the existence of energy barriers of various natures in the material. In massive semiconductors these barriers may be a consequence of the inhomogeneity of the material [1, 2], while in heterostructures the inhomogeneity at the heterointerface and the energy barriers associated with it are expressly created during growth of the system. The importance of studying nonequilibrium processes in the materials and structures that underlie modern solid-state electronics would be hard to overestimate.

The long-term relaxations of the resistance can be induced by various external influences, e.g., light, high electric fields, and changes in temperature. At low temperatures the characteristic relaxation times can reach very large values ($t > 10^5$ s), as a result of which metastable states are formed. The processes that have been studied most are the nonequilibrium processes arising as a result of irradiation by light with different wavelengths.

Recently we observed the long-term resistance relaxations induced by uniaxial compression in *p*-GaAs/$Al_{0.5}Ga_{0.5}As$ heterostructures at liquid-nitrogen and liquid-helium temperatures [3]. To the best of our knowledge, the long-term relaxations of the piezoresistance in heterostructures had not been reported before that, although similar phenomena occur in massive samples of gallium antimonide upon the application and removal of a uniaxial load [4].

In this paper, we present results of a detailed study of the properties of the relaxations of the electrical resistance and residual piezoresistance observed in *p*-GaAs/$Al_{0.5}Ga_{0.5}As$ heterostructures and propose a most likely mechanism by which the set of observed effects can be explained.

2. SAMPLES AND EXPERIMENTAL PROCEDURE

The relaxations of the piezoresistance were studied on samples prepared from two *p*-GaAs/$Al_{0.5}Ga_{0.5}As$ heterostructures with different levels of beryllium doping. The samples were grown my molecular beam epitaxy at the Niels Bohr Institute of Copenhagen University. The growth direction [001] was determined by the orientation of the GaAs substrates. The main growth parameters and characteristics of the heterostructures are given in Table 1.

The samples were punched out along the natural slip planes from a wafer on which a heterostructure had been grown; they were {110} in the form of parallelepipeds mm in size $3 \times 0.8 \times d$ mm is the thickness of the wafer in mm), with the long side directed along either [110] or $[1\bar{1}0]$. Hall bridges (mesas) were fabricated on the (001) plane of the structure by a photolithographic method. The contacts were formed by depositing Zn/Au

Table

	Structure	
	HCO218	HCO251
Manufacturer of the GaAs substrate	Sumitomo	Outokumpo
Substrate thickness, mm	0.625	0.5
Substrate temperature, °C	630	630
Rotation speed, rpm	20	20
Growth rate, μm/h	0.628	0.637
Undoped GaAs* (buffer), μm	1.0	1.0
Undoped $Al_{0.5}Ga_{0.5}As$ (spacer), nm	7.0	7.0
$Al_{0.5}Ga_{0.5}As$ doped with beryllium (active layer), nm	50	50
Be concentration in the $Al_{0.5}G_{0.5}As$, 10^{18} cm^{-3}	1	0.5
GaAs doped with beryllium, nm	5.0	5.0
Be concentration in the GaAs, 10^{18} cm^{-3}	2	1
Concentration of 2D holes, 10^{11} cm^{-2}	9.8	7.6
Mobility in the direction at 4.2 K, $m^2 V^{-1} s^{-1}$	4.7	4.5

* A GaAs/AlAs superlattice was deposited between the substrate and buffer layer.

with a subsequent "alloying in" for 3 min at a temperature of 380°C. The quality of the contacts was monitored from the linearity of the current-voltage characteristics, which was preserved at least to 50 μA, while the working current was 1–2 μA. Uniaxial compression to a pressure of 3 kbar along the long axis of the sample was effected by the technique described in Ref. [5]. During the measurements the samples were shielded from external radiation by a copper screen. Sixteen samples of the HCO218 heterostructure and four samples of the HCO251 heterostructure were investigated.

3. EXPERIMENTAL RESULTS

While the effect of uniaxial pressure on the optical transitions in $A^{III}B^{V}$ heterostructures has been studied for quite some time [6,7], the behavior of the transport properties of the charge carriers in them has practically escaped attention. This is apparently because the occurrence of long-term relaxations of the resistance upon the application and removal of an external uniaxial load was detected only comparatively recently [3]. The essence of the observed phenomenon is demonstrated in Fig. 1 for the heterostructures whose parameters are given in Table 1, for the cases when the uniaxial compression is directed along [110] and along $[1\bar{1}0]$. The effect consists in the fact that immediately after application of the load (intervals *1–2* in Fig. 1) the system of two-dimensional (2D) holes is found in a nonequilibrium state, and the corresponding electrical resistance slowly relaxes to its equilibrium value over a time that can reach 10^3–10^4 s at liquid-nitrogen temperatures and can exceed 10^5 s at 4.2 K. The equilibrium state (dashed lines) can be obtained by thermal cycling, i.e., by heating the sample to 180–200 K followed by a slow cooling to the temperature of the experiment (intervals *3–4*). If the relaxation process continues for a long enough time or the equilibrium state has been reached by the aforementioned thermal cycling under pressure, then after removal of the load (intervals *5–6*) the system is again found in a metastable state, in which the resistance is higher (or lower) than the initial equilibrium value ρ_0 at zero pressure $P = 0$. By analogy with the residual photoconductivity one can speak of a residual positive or negative piezoresistance. The initial value ρ_0 (denoted by the dotted line) can also be reached by thermal cycling (intervals *7–8*).

3.1. Anisotropic Character of the Relaxations

A distinctive feature of the relaxations of the piezoresistance is that its character changes when the compression direction is changed, as is clearly seen in Fig. 1. After the application of pressure along the [110] direction, the resistance in the loaded state relaxes to a higher value, and after the load is removed it exceeds the initial value of the resistance ρ_0, i.e., there is a positive residual piezoresistance (Figs. 1a, 1c). In the case when a compressive stress is applied in the perpendicular direction (along the $[1\bar{1}0]$ axis) the resistance relaxes to a lower value after application of the load, and it takes on a lower value than ρ_0 upon recovery to

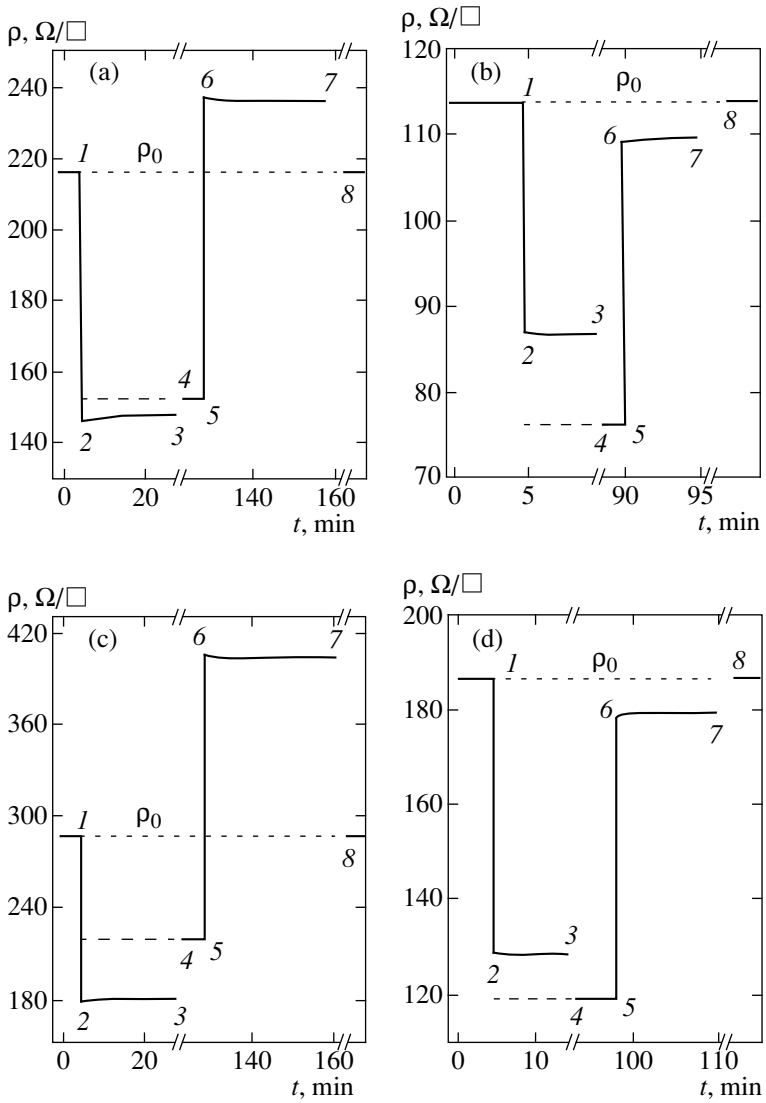


Fig. 1. Long-term resistance relaxations induced by uniaxial compression at $T = 4.2$ K for compression directions [110] (a, c) and $[1\bar{1}0]$ (b, d) in samples of the HCO218 (a, b) and HCO251 (c, d) heterostructures. A uniaxial pressure of 1 kbar is applied in the time intervals *1–2* and removed in the intervals *5–6*. In the intervals *3–4* and *7–8* the samples were heated to 200 K and slowly cooled back to 4.2 K. The dashed and dotted lines correspond to the equilibrium values of the resistance in the loaded and unloaded states, respectively.

the unloaded state, i.e., one can speak of a residual piezoconductivity (Figs. 1b, 1d). Here it should be noted that the difference of the values of ρ_0 in the two mutually perpendicular directions [110] and $[1\bar{1}0]$ for the same heterostructure is a consequence of the anisotropy of the mobility of 2D holes in the plane of the structure [8].

Figure 1 demonstrates the identical behavior of samples of the heterostructures HCO218 and HCO251 both in the nature of the response to the application of pressure and in relation to the different directions of the compressive stress. In what follows we shall give the results for the HCO218 samples, which were investigated in greater detail.

3.2. Dependence of the Amplitudes of the Relaxations on The Value of the Pressure

The amplitude of the relaxations here is understood to mean the difference $\Delta\rho_A$ of the resistances in the metastable and equilibrium (obtained after thermal cycling) states. A certain error in the determination of this difference arises because of the fact that even at 4.2 K in the first minutes after the application (removal) of the load there is a slight but noticeable relaxation of the electrical resistance, the value of which, however, is small compared to $\Delta\rho_A$ (Fig. 1). Figure 2 shows the dependence of the resistance on the load, $\rho(P)$, for two directions of uniaxial compression: [110] (Fig. 2a) and

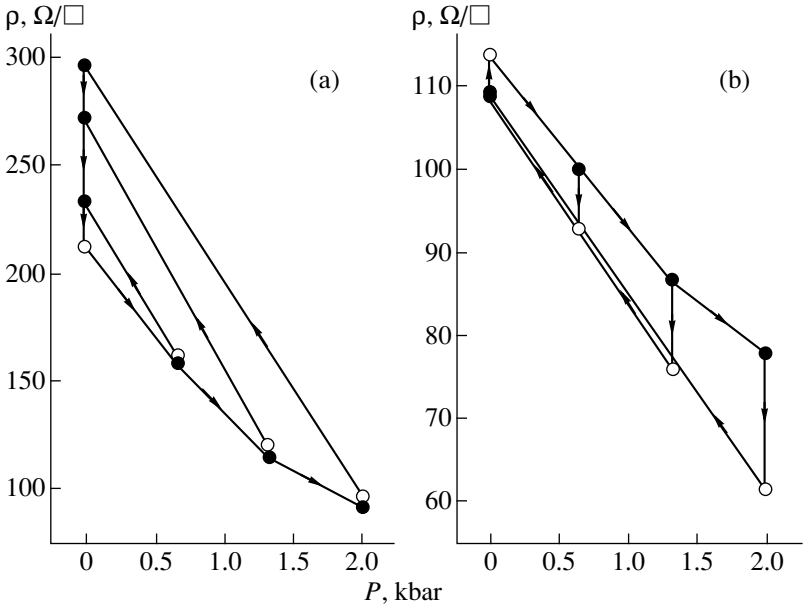


Fig. 2. Resistance of a two-dimensional hole gas versus pressure at 4.2 K for uniaxial compression directions (a) [110] and (b) [1$\bar{1}$0]. The filled symbols correspond to metastable states after application (removal) of the pressure, and the unfilled symbols to the equilibrium states obtained by heating the sample to 200 K and slowly cooling it to 4.2 K (thermal cycling). The arrows indicate the sequence in which the experimental points were obtained.

[1$\bar{1}$0] (Fig. 2b). The filled points correspond to the value of the resistance immediately after the application (removal) of the load, while the unfilled symbols are for the equilibrium state obtained after thermal cycling to 200 K.

There are two noteworthy characteristic features that follow from the curves plotted in Fig. 2.

(1) The amplitude of the relaxations on both application and removal of the load increases with increasing applied (removed) pressure.

(2) In the case of uniaxial compression along [110] the amplitude of the relaxations after removal of the load was noticeably larger than the amplitude of the relaxations after application of the load (Figs. 1a, 2a), whereas for the [1$\bar{1}$0] direction the picture changes to the opposite (Figs. 1b, 2b).

3.3. Time Dependence of the Resistance Relaxations

The time dependence of the piezoresistance during the relaxation process is of a nonexponential character. In the majority of cases this process can be described by a logarithmic law like that given in [9, 10]:

$$R^{-1}(t) = R^{-1}(t=0) \pm A\ln\frac{t}{\tau_0}, \qquad (1)$$

which is valid for the recombination of spatially separated carriers with a rectangular profile of the spatial distribution of charge traps. Here it is assumed that the measurement time $t \gg \tau_0$, where τ_0 is the characteristic lifetime of the charge carriers in the absence of their spatial separation from the trapping centers. It should be noted that in many cases [11, 12] just such a logarithmic time dependence gives a good description of the relaxation of the residual photoconductivity in *n*-GaAs/$Al_xGa_{1-x}As$.

Figure 3a shows the characteristic time dependence of 1/ρ in the linear and logarithmic time scales for one of the samples in the [110] orientation after the removal of a pressure $P = 0.85$ (curve *1*) and 1.7 kbar (curve *2*) at a temperature of 77 K. It should be noted that two linear segments can be discerned on the logarithmic plots. Analysis of the time dependence of the resistance relaxations at 4.2 K (Fig. 3b) after the application (curve *4*) and removal (curve *3*) of the load also reveals the presence of two processes, the faster of which obeys a logarithmic law, while the other is practically "frozen." Here the first linear parts of the curves in Fig. 3 pertain to a relaxation process which even at low temperature lasts only a few minutes. This relaxation process can be seen in Fig. 1 immediately following the application (removal) of the pressure, and its contribution to the change in resistance is insignificant. It is most likely due to the tunneling of charge from residual ionized impurities lying nearby in the buffer layer. We are mainly interested in the long-lived metastable state described by the second linear segments of the curves.

3.4. Memory Effect

If the loading-unloading cycle is repeated without intermediate thermal cycling but the relaxation process under load goes on for a rather long time (tens of minutes),

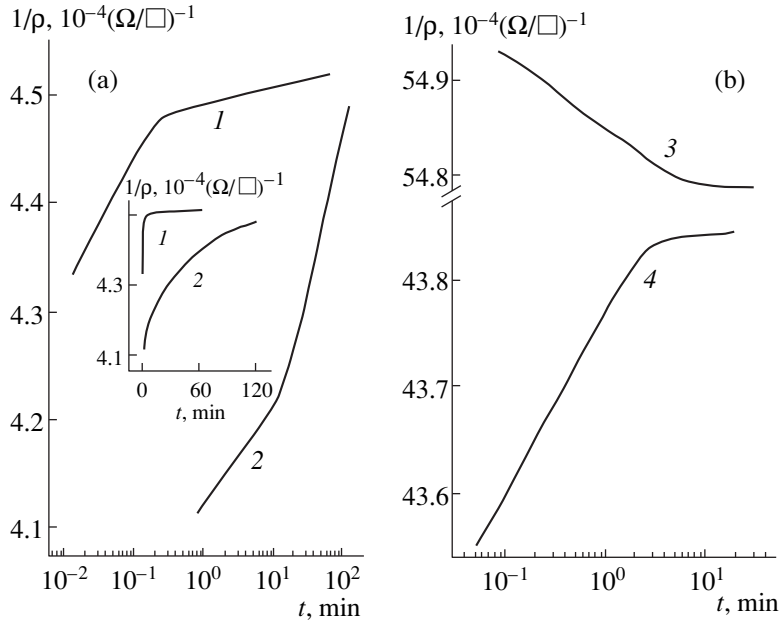


Fig. 3. Sheet conductance of a two-dimensional hole gas versus time for T = (a) 77 and (b) 4.2 K. Curves *1*, *2*, and *4* are recorded after the removal of pressures of 0.85, 1.7, and 0.35 kbar, respectively, and curve *3* after the application of a load of 0.35 kbar. The inset $\rho^{-1}(t)$ shows plotted on a linear time scale.

then, after the load is removed, a residual piezoresistance is also observed, but its value now depends not only on the value of the pressure applied prior to the removal but also on the duration of the relaxation process under load. At helium temperatures this process, as we have said, does not go to completion in the time intervals actually realizable in experiment. If the pressure is removed immediately after it is applied, then the system returns nearly to its initial state: no relaxation process occurs in the loaded state.

This is the basis for the memory effect, wherein the sample "remembers" previous loadings and relaxations. This effect is demonstrated in Fig. 4. For example, if the relaxation process after removal of the load (interval *4–5*) is interrupted by a brief cycle in which a uniaxial stress of the same magnitude is applied and removed (interval *5–6–7–8*), the resistance of the sample "remembers" its state prior to the last loading cycle, and the subsequent relaxation (*8–9*) is a continuation of the previous relaxation process (*4–5*). This does not occur if the relaxation in the loaded state continues for a long enough time (*10–11*). In that case after the load is removed the relaxation process *12–13* is not a continuation of the relaxation process *8–9*, i.e., the resistance of the sample also "remembers" the relaxation processes that have occurred in the previous loaded state. It is clear that in order to exactly reproduce the states of the sample before each new measurement cycle it must be brought to an equilibrium state by thermal cycling.

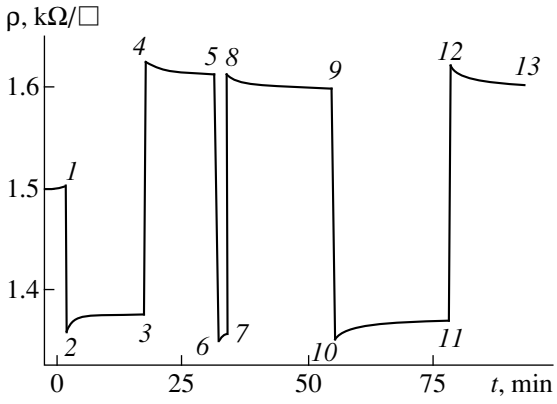


Fig. 4. Sheet resistance of a two-dimensional hole gas versus time at T = 77 K when the load is applied and removed. In the time intervals *1–2*, *5–6*, and *9–10* a uniaxial pressure P = 0.85 kbar was applied to the sample along the [110] direction, and in the intervals *3–4*, *7–8*, and *11–12* the load was removed.

A memory effect is also characteristic for the previously observed nonequilibrium processes induced by illumination [4], for example.

3.5. Quenching of the Relaxations of the Piezoresistance by Temperature, Illumination, and a High Electric Field

Quenching of the relaxations of the piezoresistance in the thermal cycling process has in fact already been demonstrated in Fig. 1. When the sample is heated from 4.2 to 200 K and then slowly cooled, the residual piezoresistance is "erased" (intervals *7–8* in Fig. 1) and the sample is brought back to the initial equilibrium state. By analogy, in this case one can regard the relaxation pro-

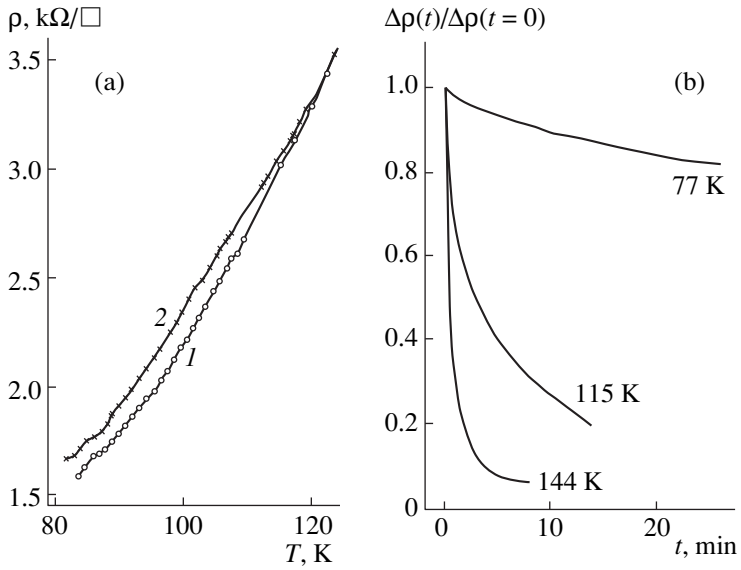


Fig. 5. (a) Temperature dependence of the sheet resistance in the initial equilibrium state (*1*) and after a loading and unloading cycle is carried out at 77 K with a load of $P = 1$ kbar (*2*); (b) relaxation of the residual piezoresistance $\Delta\rho$ after removal of a pressure $P = 0.85$ kbar at different temperatures.

cess under load as being quenched by thermal cycling, which brings the system of charge carriers to an equilibrium state. The temperature dependence of the resistance in the unloaded equilibrium state and after a loading-unloading cycle with $P = 1$ kbar is shown in Fig. 5a. The fact that the two states have already become indistinguishable at 120 K is explained by the increase in the rate of relaxations as the temperature is raised (Fig. 5b).

If the heating brings the system of charge carriers to a thermodynamic equilibrium state, as a result of which a temperature quenching of the relaxations of the piezoresistance occurs, then illumination or a strong electric field pulse can themselves cause nonequilibrium processes leading to charge redistribution in the system. This is apparently the basis for their quenching effect on the long-term relaxations induced by uniaxial compression and for the residual piezoresistance at low temperatures. For the illumination we used red light with a photon energy of 1.9 eV from a (Al)GaAs light-emitting diode. Figure 6a illustrates the quenching of the residual piezoresistance by illumination at 77 K: when the diode is turned on briefly (point *5*) the residual piezoresistance induced by a loading-unloading cycle with a pressure of 0.5 kbar is "dumped," and the electrical resistance rapidly relaxes to its equilibrium value. If the loading (unloading) of the sample is carried out under illumination, then the relaxation process and the residual piezoresistance are not observed at all. The quenching of the relaxations of the piezoresistance by illumination persists at helium temperatures, but in that case the effect takes place against the background of the residual photoconductivity.

For studying the effect of a high electric field on the relaxation of the piezoresistance we used a pulsed method to preclude overheating of the sample. Square voltage pulses 30–400 ns long and with a repetition frequency of 1–100 Hz were produced by a generator across a mercury relay. At a temperature of 77 K the effect of the high electric field pulses $E \geq 3000$ V/cm is manifested in a substantial acceleration of the relaxations of the piezoresistance, so that 2–3 min after application of the electric field pulses the resistance will essentially reach its equilibrium value (Fig. 6b, for a perturbing uniaxial pressure of 1.5 kbar). At helium temperatures a high enough electric field $E \geq 200$ V/cm will itself cause switching of the samples into a long-lived high-resistance state that is sensitive to illumination.

3.6. Behavior of the Total Concentration of 2D Holes During Relaxation of the Piezoresistance

The concentration of 2D holes was monitored from the Hall effect and the frequency of quantum oscillations of the magnetoresistance at a temperature of 1.7 K. It was found that in the course of the relaxations of the piezoresistance the total concentration of 2D holes in the quantum well varies in correlation with the change in the electrical resistance. Figure 7 shows the characteristic curves of the total concentration of holes as a function of the uniaxial compressive stress during the application and removal of pressure along the [110] (Fig. 7a) and $[1\bar{1}0]$ (Fig. 7b) axes; the cycles of application and removal of the pressure were carried out with an intermediate thermal cycling to accelerate the relaxations. As we see from Fig. 7a, in the case of compression along [110] the hole concentration in the well decreases in the course of the relaxations after the application of pressure, while after the removal of pressure the hole concentration increases in the relaxation

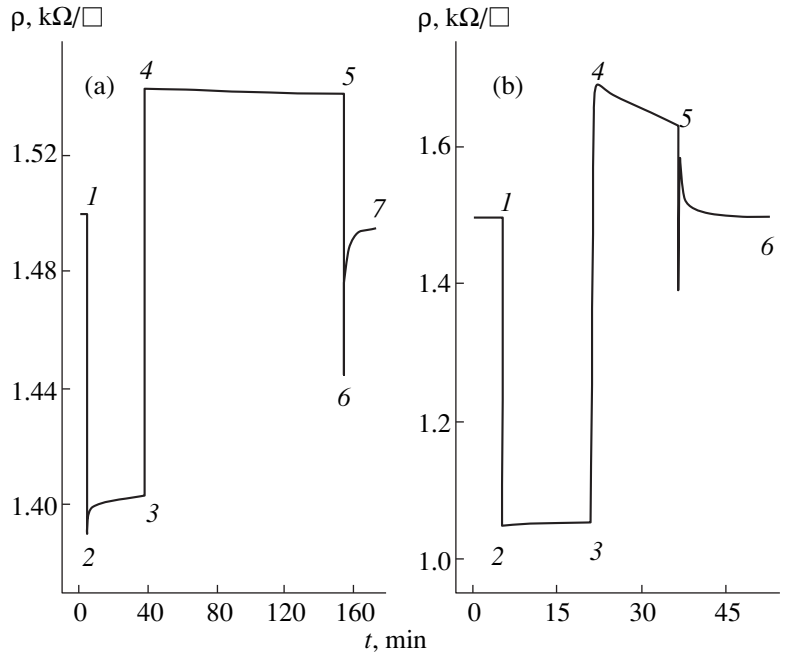


Fig. 6. Quenching by (a) light and (b) by a pulsed electric field of the residual piezoresistance induced by a cycle of application and removal (intervals *1–2–3–4*) of a uniaxial pressure at a temperature of 77 K. At time 5 the sample (a) is illuminated by a red LED for about 1 s or (b) is subjected to several pulses of a high electric field.

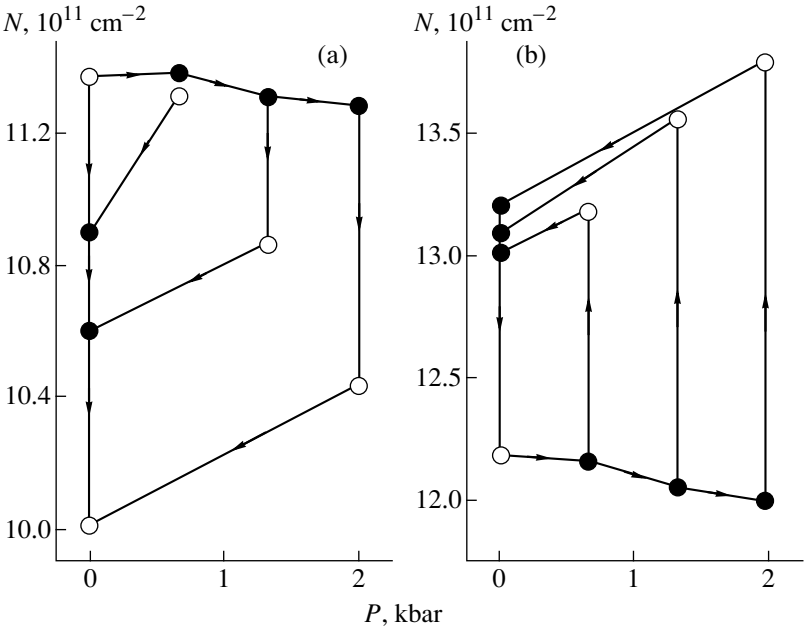


Fig. 7. Concentration of 2D holes as a function of pressure for compression directions [110] (a) and [1$\bar{1}$0] (b). The filled symbols correspond to metastable states immediately after the application (removal) of the load, and the unfilled symbols to equilibrium states obtained as a result of thermal cycling. The arrows indicate the sequence in which the experimental points were obtained.

process. For compression along [1$\bar{1}$0] (Fig. 7b) the opposite picture is observed: after loading of the sample the concentration of 2D holes relaxes to a higher value, while after removal of the load it relaxes to a lower value.

Despite the qualitative difference, the dependence of the concentration on the uniaxial compression along the [110] and [1$\bar{1}$0] directions has the following features in common:

(1) the strain dependence of the total hole concentration in the quantum well in the metastable state is the same for both compression directions—the hole concentration falls off slightly with increasing pressure (by 2–3% at $P = 2$ kbar);

(2) the removal of pressure from a sample in an equilibrium stressed state (obtained by thermal cycling) is always accompanied by an appreciable decrease in the concentration of 2D holes, independent of the direction of the uniaxial compression.

4. DISCUSSION

The observed properties of the relaxations of the piezoresistance, such as quenching of the metastable state by illumination, the application of a high electric field, and heating to a comparatively low temperature (around 200 K), indicate that the relaxations of the electrical resistance and the residual piezoresistance are not due to the introduction of some new defects in the crystal during the uniaxial compression. Therefore the observed effect should be considered from the standpoint of the electron-hole system and the internal properties of the structure itself.

Since a two-dimensional hole gas is formed in a quantum well on account of ionization of the Be acceptor impurity in the $Al_{0.5}Ga_{0.5}As$ active layer, in an analysis of the causes of the resistance relaxations one must consider the possibility that they can be explained by the tunneling of holes from the quantum well to Be impurity levels separated from the two-dimensional layer by an undoped spacer. However, according to the data of [13], the characteristic times for tunneling in this process at 4.2 K is $t \approx 10^{-8}$ s, whereas the observed relaxations of the piezoresistance are characterized by times $t > 10^5$ s.

It is known that nonequilibrium processes with such long relaxation times may be due to deep traps. It is possible that in our case, too, the resistance relaxations can be attributed to such widely known deep centers in GaAs and (Al)GaAs as DX and EL2 centers. However the problem is that they and all the other point defects of the crystal lattice, in view of the symmetry of the three-dimensional crystal, cannot account for the difference observed in this study in the character of the relaxations for compression in the directions [110] and $[1\bar{1}0]$ (Figs. 1, 7). The causes of the relaxations of the piezoresistance must be sought first among those mechanisms for which such a difference exists. Gallium arsenide, like other $A^{III}B^{V}$ compounds, exhibits piezoelectric properties on account of the absence of inversion symmetry of the lattice along the [111] crystallographic direction. Therefore, such a difference can be given by the piezoelectric effect, since the strain-induced electric field along the [001] direction changes sign when the compression direction is changed from [110] to $[1\bar{1}0]$.

The piezoelectric effect in such materials is determined by the off-diagonal components of the strain tensor ε_{xy} and it therefore absence for compression along [100]. For compression along the [110] and $[1\bar{1}0]$ directions of interest to us here, the strain components $\varepsilon_{xy} \neq 0$, taking the values $\varepsilon_{xy} = S_{44}P/4$ and $\varepsilon_{xy} = -S_{44}P/4$, respectively, where S_{44} is a component of the elastic compliance tensor. As a result, in the absence of compensating charges in the crystal, an electric field arises along the [001] axis, with a value given by the expression [14]

$$E_z = -\frac{2e_{14}\varepsilon_{xy}}{\varepsilon\varepsilon_0}, \qquad (2)$$

where e_{14} is the piezoelectric constant, ε is the static dielectric constant, and ε_0 is the permittivity of free space. Using the values $\varepsilon = 13.18$ and $e_{14} = -0.16 \times 10^{-4}$ C/cm^2 for GaAs [15], we find that at the strain corresponding to a uniaxial pressure of 1 kbar, a piezoelectric field with a value of $|E_z| = 1.152 \times 10^4$ V/cm should arise in GaAs. The electric field is positive (directed from the substrate toward the heterointerface) for compression along $[1\bar{1}0]$ and negative (directed from the heterointerface toward the substrate) for compression along [110].

In the $Al_{0.5}Ga_{0.5}As$ layer, since it is doped with beryllium and there is a two-dimensional hole gas at the GaAs/AlGaAs heterointerface, the strain-induced electric field should be compensated in a time of less than 1 s by a redistribution of charge in the active layer and quantum well [13]. In the GaAs buffer layer, on the other hand, the electric field that arises cannot be immediately compensated, since the compensating charge carriers must penetrate from impurity states in the GaAs buffer and substrate into the quantum well at the GaAs/$Al_{0.5}Ga_{0.5}As$ heterointerface through the entire GaAs buffer layer, which is 1 μm thick. The piezoelectric field distorts the band structure, leading to the changes shown schematically in Fig. 8.

An estimate using the model presented by the authors of Ref. [10] shows that in such a case the lifetime of the metastable state corresponding to different positions of the Fermi quasilevel near the heterointerface and in the substrate takes on very large values and is apparently limited at finite temperatures mainly by thermally activated processes involving residual impurities and defects in the material of the substrate and buffer layer. At helium temperatures the probability of such processes is extremely small and the metastable state is practically frozen, whereas heating to 200 K promotes the rapid formation of a charge that compensates the piezoelectric field, so that the material is brought to an equilibrium state.

For compensation of the piezoelectric field in the GaAs buffer layer at a pressure of 2 kbar it is necessary that a charge density of 2.69×10^{-8} C/cm^2 be formed at

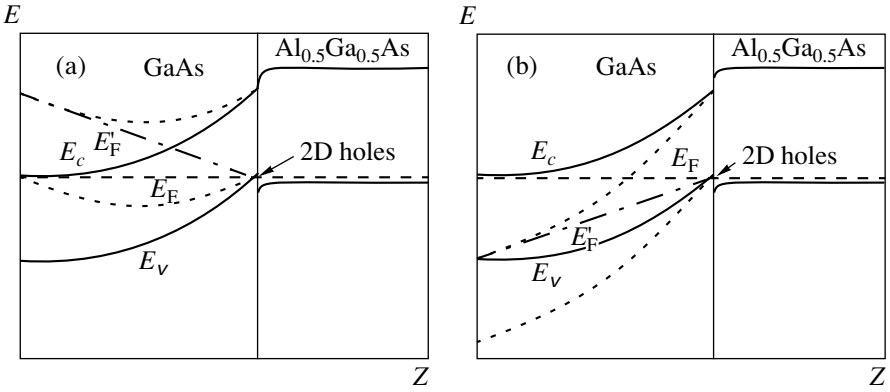


Fig. 8. Piezoeffect-induced change in the band structure in *p*-GaAs/$Al_{0.5}Ga_{0.5}As$ under uniaxial compression. The solid curves correspond to the unloaded equilibrium state, and the dotted curves to the equilibrium state immediately after application of a uniaxial pressure of 1 kbar along the (a) [110] and (b) $[1\bar{1}0]$ directions. The dashed curve corresponds to the position of the equilibrium Fermi E_F level, and the dot-and-dash curve to the nonequilibrium Fermi E'_F level immediately after application of the load.

its boundaries; this value corresponds to an additional concentration of holes or electrons (depending on the direction of the field due to the piezoelectric effect) at the GaAs/$Al_{0.5}Ga_{0.5}As$ heterointerface of $\Delta N = 1.68 \times 10^{11}$ cm^{-2}. This estimate of the density of compensating charge carriers is in rather good agreement with the observed difference between the values of the concentration of 2D holes in the metastable and equilibrium states (Fig. 7).

In accordance with the direction of the field E_z the compensating charge should preferentially increase the hole concentration in the well in the case when the pressure is applied along $[1\bar{1}0]$ and decrease it in compression along [110] (accordingly, in the equilibrium state under pressure the hole concentration should be higher or lower, respectively, than in the metastable state). This correlates completely with the experimental data (Fig. 7).

The total change in the carrier concentration in the quantum well under pressure is determined not only by the influence of the piezoelectric field but also by a set of other factors such as the change in the band structure and the value of the discontinuity of the bands at the heterointerface, the ionization energy of the Be acceptor impurity in the active layer, and the state of the defects and deep centers near the heterointerface. The present uncertainty in the values of some of the parameters and their strain dependence makes it difficult to interpret the total pressure dependence of the concentration of 2D holes. Using the difference in the time scales of the transient processes for different mechanisms responsible for changes in the concentration of free holes, we have attempted to separate and analyze the contributions due to several of them.

As we have said, all of the processes of redistribution of the holes between shallow acceptors in the active layer and the 2D layer at the heterointerface occur over rather short times. Therefore, we can assume that the change in concentration due to the changes of the band structure and the activation energy of Be shallow acceptor levels occur practically simultaneously with the application of the pressure.

On the other hand, such processes as the recharging of deep centers and the tunneling of carriers through the GaAs buffer layer occur very slowly at helium temperatures, and the corresponding characteristic times can exceed 10^5 s.

Starting from this fact, we can write the change in concentration under pressure, $\Delta N(P)$, in the form of a sum:

$$\Delta N(P) = \Delta N_1(P) + \Delta N^*(P),$$

where $\Delta N_1(P)$ is the contribution of "fast" processes due to changes in the band structure and to the redistribution of holes between the 2D layer and the Be shallow acceptors in the $Al_{0.5}Ga_{0.5}As$ active layer and $\Delta N^*(P)$ is the change in concentration due to the long-term processes: the recharging of deep centers and the tunneling of carriers through the GaAs buffer layer.

The first term $\Delta N_1(P)$ is determined from the dependence $N(P)$ obtained at 1.7 K without thermal cycling, when the long-term relaxation processes cannot appreciably influence the concentration of 2D holes at the heterointerface. Arguing in favor of the interpretation that it is these values that correspond to changes of the band structure and to the redistribution of holes between the 2D layer and shallow acceptors in the active layer is the fact that when the pressure is removed, the concentration of 2D holes in this case (i.e., without thermal cycling) returns to its initial value. The $\Delta N_1(P)$ curves obtained from Fig. 7 are presented in Fig. 9a.

The concentration change $\Delta N^*(P)$ responsible for the long-term relaxation of the resistance can be

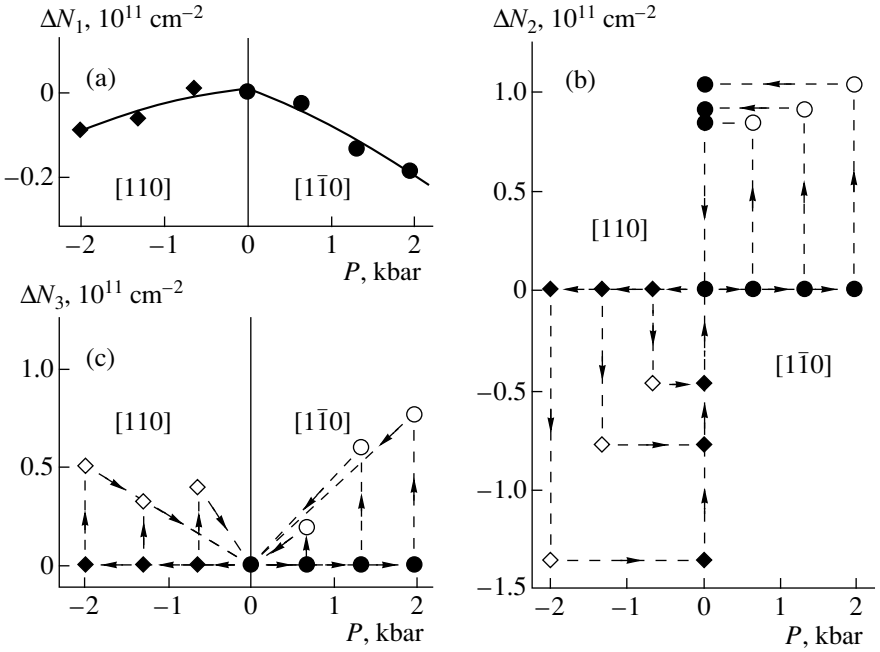


Fig. 9. Partial contributions of the change in concentration of 2D holes ΔN_1 (a), ΔN_2 (b), and ΔN_3 (c) under pressure (see text) to the overall dependence of $N(P)$. The data for the compression directions [110] and [1$\bar{1}$0] are presented on the same graph in such a way that negative values of the pressure correspond to compression along [110] and positive to compression along [1$\bar{1}$0]. The filled symbols correspond to metastable states immediately after the application (removal) of the pressure, and the unfilled symbols to the equilibrium states obtained as a result of thermal cycling. The arrows indicate the sequence in which the experimental points were obtained.

obtained on the basis of Fig. 7 by the simple subtraction

$$\Delta N^* = \Delta N(P) - \Delta N_1.$$

Here one notices a difference in the value of the change in concentration of 2D holes when the total relaxation (after thermal cycling) occurs under load and after the load is removed (Fig. 7), whereas the corresponding numbers of carriers entering and leaving the quantum well should be equal to each other if only the piezoelectric field is responsible for this effect.

In addition, the two compression directions [110] and [1$\bar{1}$0] have a common feature, which was mentioned in Subsection 3.6: in both cases when the load is removed from a sample in the equilibrium state under pressure, the concentration of 2D holes decreases noticeably (Fig. 7). In view of the identical behavior for the compression directions [110] and [1$\bar{1}$0], this phenomenon cannot be attributed to the piezoelectric effect. Apparently, the long-term relaxation processes also involve some mechanism for which these directions are equivalent.

That mechanism might be due to the presence of point defects, which have been observed in large numbers near heterointerfaces in *n*- and *p*-type $GaAs/Al_xGa_{1-x}As$ [16, 17] and are identified as deep centers. Their concentration reaches $\sim 10^{10}$ cm^{-2} [17]. If we suppose that electron traps with a pressure-dependent thermal activation barrier exist near the heterointerface, we can assume that these states, being occupied as a result of thermal cycling under load, will easily give up the trapped electrons when the pressure is removed, thus reducing the concentration of 2D holes in the quantum well. On this basis, the piezoeffect-induced change in the hole concentration, ΔN_2, is defined as the difference between the values of the concentration in the equilibrium state at and in the nonequilibrium state after the pressure is removed. Then on the basis of the curve (Fig. 7) it is proposed that the change in concentration ΔN^* in the two states can be separated:

$$\Delta N^* = \Delta N_2 + \Delta N_3,$$

where the main contribution ΔN_2 is due to the piezoeffect, and the considerably smaller part ΔN_3 that is symmetric with respect to the compression direction is due to deep traps. The result of this separation is shown in Figs. 9b and 9c. The data obtained on the concentration ΔN_2 of the holes that compensate the piezoelectric field are in good agreement with the estimates given above on the basis of formula (2). The change in the hole concentration due to deep centers also appears completely realistic in accordance with the known published data on the concentration of the latter near a $GaAs/Al_xGa_{1-x}As$ heterointerface [17].

The quenching of the metastable states by light and electric field pulses can be explained in the proposed model as being due to the formation of additional free electrons and holes due to the optical radiation and high pulsed electric field, which promotes the rapid formation of the compensating charge.

As we have said, the logarithmic time dependence of the resistance in the course of the relaxations of the residual photoconductivity is predicted for the case of recombination of spatially separated charge, when the carriers of one type are captured by traps and those of the other type remain free and participate in the conduction [10]. The charge separation in this model presupposes the presence of barriers, which are formed in heterostructures and which arise as a result of fluctuations of the composition or of the doping impurity. The logarithmic time dependence of the resistance here is due to the fact that the damping of the photoconductivity is slowed as the recombination front moves away from the layer of free charge carriers. Apparently, such a model can be used to explain the time dependence of the resistance relaxation due to the compensation of the piezoelectric field, since in that case we are dealing with the motion of the charge carriers through the buffer layer, with the participation of the residual impurities in the process. Since the probability of ionization of these impurities increases with increasing temperature, the rate of relaxations must also grow accordingly, which explains the quenching of the relaxations by the temperature.

It should be noted that besides the piezoeffect, there exists another mechanism in terms of which one might interpret both the main properties of the relaxations of the piezoresistance and also the different signs of the change in concentration of the 2D holes in the course of the relaxations for the compression directions [110] and $[1\bar{1}0]$. Such a difference can be given by misfit dislocations at the (001) GaAs/(Al)GaAs heterointerface, the chemical nature of which differs for these two directions: exclusively α dislocations (with As atoms in the dislocation core) are arranged along [110], and only β dislocations (with Ga atoms in the core) are arranged along $[1\bar{1}0]$. A model in which the change in concentration of 2D holes due to uniaxial compression is attributed to the recharging of the cores of misfit edge dislocations is described in detail in [18].

This mechanism is undoubtedly important for heterojunctions with a large lattice misfit of the epitaxial layers and should be taken into account in the fabrication of structures of such materials as GaAs/(Ga)InAs, GaAs/GaAs(P), and InSb/GaAs. In our case, however, the lattice misfit of GaAs and $Al_{0.5}Ga_{0.5}As$ is small and, according to [19, 20], should not lead to any misfit dislocations at the heterointerface. In spite of the fact that the question of whether misfit dislocations can form in such structures is far from an unambiguous answer, and there are examples that contradict this assertion (e.g., see [21]), we have been unable to detect the characteristic pattern of misfit dislocations in a scanning cathodoluminescence study of the structures investigated here.

ACKNOWLEDGMENTS

We are deeply grateful to V.I. Petrov for assistance in the scanning cathodoluminescence studies, to V.I. Panov and the staff of his laboratory for collaboration in the monitoring of the orientation of the samples on a scanning atomic force microscope, and to S.D. Beneslavskiĭ and E.V. Bogdanov for a helpful discussion.

This study was supported by the Russian Foundation for Basic Research (project nos. 00-15-96784 and 97-02-17685).

REFERENCES

1. M. K. Sheĭnkman and A. Ya. Shik, Fiz. Tekh. Poluprovodn. (Leningrad) **10**, 209 (1976) [Sov. Phys. Semicond. **10**, 128 (1976)].
2. A. Ya. Shik and A. Ya. Vul', Fiz. Tekh. Poluprovodn. (Leningrad) **8**, 1675 (1974) [Sov. Phys. Semicond. **8**, 1085 (1974)].
3. V. N. Kravchenko, N. Ya. Minina, Ya. S. Olsen, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 417 (1995) [JETP Lett. **61**, 424 (1995)].
4. A. Ya. Vul' and A. Ya. Shik, Fiz. Tekh. Poluprovodn. (Leningrad) **8**, 1952 (1974) [Sov. Phys. Semicond. **8**, 1264 (1974)].
5. N. B. Brandt, V. S. Egorov, M. Yu. Lavrenyuk, *et al.*, Zh. Éksp. Teor. Fiz. **89**, 2257 (1985) [Sov. Phys. JETP **62**, 1303 (1985)].
6. C. Mailhiot and D. L. Smith, Phys. Rev. B **36**, 2942 (1987).
7. G. Platero and M. Altarelli, Phys. Rev. B **36**, 6591 (1987).
8. A. M. Savin, C. B. Sorensen, O. P. Hansen, *et al.*, Semicond. Sci. Technol. **14**, 632 (1999).
9. H. J. Queisser, Phys. Rev. Lett. **54**, 234 (1985).
10. H. J. Queisser and D. E. Theodorou, Phys. Rev. B **33**, 4027 (1986).
11. E. F. Schubert, A. Fisher, and K. Ploog, Phys. Rev. B **31**, 7937 (1985).
12. M. J. Chou and D. C. Tsui, Appl. Phys. Lett. **47**, 609 (1985).
13. A. Dargys, N. Zurauskiene, and K. Bertulis, J. Phys.: Condens. Matter **9**, L557 (1997).
14. C. Mailhiot and D. L. Smith, Phys. Rev. B **35**, 1242 (1987).
15. Sadao Adachi, J. Appl. Phys. **58**, R1 (1985).
16. O. P. Hansen, J. Szatkowski, E. Placzek-Popko, *et al.*, Cryst. Res. Technol. **31**, 313 (1996).
17. P. Krispin, R. Hey, and H. Kostial, J. Appl. Phys. **77**, 5773 (1995).
18. V. N. Kravchenko, Candidate's Dissertation in Mathematical Physics (Mosk. Gos. Univ., Moscow, 1999).
19. J. W. Matthews and A. E. Blakeslee, J. Cryst. Growth **27**, 118 (1974).
20. B. C. Cooman, C. B. Carter, Kam Toi Chan, *et al.*, Acta Metall. **37**, 2779 (1989).
21. D. Cherns, D. Loretto, N. Chand, *et al.*, Philos. Mag. A **63**, 1335 (1991).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1261–1267.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1456–1462.

SOLIDS
Electronic Properties

A Diagram Technique for Perturbation Theory Calculations of the Effective Conductivity of Two-Dimensional Systems¶

I. M. Khalatnikov[a–c, *] and A. Yu. Kamenshchik[a, b]

[a]*Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117334 Russia*
[b]*Landau Network-Centro Volta, 22100, Como, Italy*
[c]*School of Physics and Astronomy, Raymond and Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, 69978 Israel*
**e-mail: landau@ICIL64.CILEA.IT*

Received July 27, 2000

Abstract—The perturbation theory for calculating the effective conductivity of the plane consisting of pieces of different conductivities is constructed, and a convenient diagram technique is elaborated for this perturbation theory. It is shown that for the chessboard, perturbative calculations give results that are in agreement with the well-known formula $\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}$. The components of the effective conductivity tensor for the anisotropic three-color chessboard are calculated. It is shown that the isotropic (symmetric) part of the effective conductivity calculated up to the sixth order of perturbation theory satisfies the Bruggeman effective medium equation for symmetric three-color structures with equally partitioned components. We also consider an isotropic three-color chessboard with nonequal weights of colors. In this case, the perturbation theory in the fourth order contradicts the results following from the Bruggeman equation for nonequal weights. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of calculating the effective conductivity of composite materials is of interest from theoretical and phenomenological points of view and has been attracting the attention of theorists since the nineteenth century [1]. The most developed is the theory of the effective conductivity of the plane, where one can use the duality between conductivity and resistance, which is typical only of the two-dimensional Ohm's law. This duality consists in the fact that rotating the plane by $\pi/2$ simultaneously with interchanging the current density vector j and the electric field E does not change Ohm's law. The most interesting two-dimensional result is the one by Dykhne [2], who considered the plane covered by regions with two different conductivities σ_1 and σ_2. If the distribution is stochastic and the statistical weights of the two conductivities are equal, one can show that

$$\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}. \tag{1.1}$$

The same result also holds for some regular structures of two-conductivities on the plane (for convenience, we call them two-color in what follows), in particular, for the chessboard, as was shown by Keller in his earlier paper [3].

The main goal of this paper is to construct the perturbation theory for calculating the effective conductivity on the plane for an arbitrary distribution of the conductivity and to represent the resulting formulas in a convenient graphical form. As applications of this theory, we rederive the Keller–Dykhne formula for the two-color chessboard. Using our diagram technique, we calculate the components of the effective conductivity tensor for the three-color chessboard representing an anisotropic structure with equal weights. It is shown that the isotropic part of the effective conductivity calculated up to the sixth order of perturbation theory satisfies the Bruggeman effective medium equation for symmetric three-color structures with equally partitioned components [4].

We also consider the isotropic three-color chessboard with nonequally partitioned components and show that the corresponding Bruggeman equation fails in this case.

We note that numerical simulation of the conductivity of regular, isotropic equal-weighted three-color structures on the plane was performed and analyzed in a recent paper [5]. Three-component dielectric media were also considered in [6].

The structure of the paper is as follows. In Section 2, we construct the perturbation theory for the effective conductivity and give a convenient diagram representation for this technique. Section 3 is devoted to the application of this technique to the deduction of the formula for the effective conductivity of the two-color chessboard. In Section 4, we use the perturbative technique to calculate components of the effective conductivity tensor for the three-color two-dimensional chessboard

¶This article was submitted by the authors in English.

and show that its isotropic part satisfies the Bruggeman equation. We then consider the isotropic three-color chessboard with nonequal weights and show that in this case, fourth-order calculations contradict the predictions following from the corresponding Bruggeman equation.

2. PERTURBATION THEORY FOR THE EFFECTIVE CONDUCTIVITY

We begin with constructing the perturbation theory for the effective conductivity. The idea of using perturbative methods for calculating the effective conductivity has been considered by Herring [7], Dykhne [8], and Bergman [9]. We note that Herring obtained a formula for the effective conductivity tensor of a locally anisotropic medium in the second order [7], while we develop the technique applicable to arbitrary orders of perturbation theory for locally isotropic media.

Ohm's law for a locally isotropic medium has the form

$$\mathbf{j} = \sigma\nabla\varphi, \tag{2.1}$$

where $\mathbf{j}$ is the current density; σ is the conductivity, which in general depends on the coordinates x and y; and φ is the electric potential. The charge conservation rule for a stationary distribution of currents is

$$\mathrm{div}\mathbf{j} = \nabla\mathbf{j} = 0. \tag{2.2}$$

Inserting Eq. (2.1) in Eq. (2.2), we obtain

$$\Delta\varphi + \nabla\ln\sigma\cdot\nabla\varphi = 0. \tag{2.3}$$

We now assume that

$$\sigma(x, y) = 1 + \alpha(x, y), \quad \langle\alpha\rangle = 0. \tag{2.4}$$

Here, the average value of $\sigma(x, y)$ is chosen equal to 1 for convenience, and $\alpha(x, y)$ is a small function whose average over the plane, denoted by $\langle\rangle$, is equal to zero.

We represent the potential φ as

$$\varphi = \mathbf{E}\cdot\mathbf{r} + \psi(x, y), \tag{2.5}$$

where the vector field $\mathbf{E}$ must be chosen such that the relation

$$\langle\nabla\psi\rangle = 0 \tag{2.6}$$

is satisfied. For isotropic conductivity distributions, we can introduce the effective conductivity by the relation

$$\mathbf{J} = \sigma_{\mathrm{eff}}\mathbf{E}, \tag{2.7}$$

where $\mathbf{J} = \langle\mathbf{j}\rangle$. For anisotropic distributions of $\alpha(x, y)$, relation (2.7) becomes a tensor one,

$$J_i = \sigma_{\mathrm{eff}\,ij}E_j, \quad i, j = 1, 2. \tag{2.8}$$

Inserting φ from (2.5) in Eq. (2.3), we obtain the Laplace equation

$$\Delta\psi + \nabla\ln(1 + \alpha)\cdot(\mathbf{E} + \nabla\psi) = 0. \tag{2.9}$$

It is convenient to introduce Fourier expansions for the small functions α and ψ:

$$\alpha = \sum_{\mathbf{k}}\alpha_{\mathbf{k}}\exp(i\mathbf{k}\cdot\mathbf{r}), \tag{2.10}$$

$$\psi = \sum_{\mathbf{k}}\psi_{\mathbf{k}}\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{2.11}$$

With these expansions in Eq. (2.9), we obtain

$$\mathbf{k}^2\psi_{\mathbf{k}} = \left[\nabla\left(\alpha - \frac{\alpha^2}{2} + \frac{\alpha^3}{3} - \frac{\alpha^4}{4} + \ldots\right)\cdot(\mathbf{E} + \nabla\psi)\right]_{\mathbf{k}}. \tag{2.12}$$

Solving Eq. (2.12) by iterations, one can find a perturbative expansion for the function $\psi_{\mathbf{k}}$. Using the formulas found for $\psi_{\mathbf{k}}$in Eqs. (2.7) or (2.8), one can then find the components of the effective conductivity tensor.

We now temporarily forget about condition (2.6) providing the correct definition of the macroscopic vector $\mathbf{E}$. One can write the following formulas for the perturbative corrections to the potential $\psi_{\mathbf{k}}$ and for the isotropic part of the effective conductivity tensor:

$$\sigma_{\mathrm{eff}} \equiv \frac{1}{2}(\sigma_{xx} + \sigma_{yy}), \tag{2.13}$$

$$\psi_{\mathbf{k}1} = \frac{i\alpha_{\mathbf{k}}(\mathbf{k}\cdot\mathbf{E})}{\mathbf{k}^2}, \tag{2.14}$$

$$\sigma_{\mathrm{eff}2} = -\sum_{\mathbf{k}}\frac{(\mathbf{k}\cdot\mathbf{E})^2}{\mathbf{E}^2\mathbf{k}^2}\alpha_{\mathbf{k}}\alpha_{-\mathbf{k}}, \tag{2.15}$$

$$\psi_{\mathbf{k}2} = -i\sum_{\mathbf{p}}\frac{(\mathbf{p}\cdot\mathbf{E})(\mathbf{p}\cdot\mathbf{k})}{\mathbf{k}^2\mathbf{p}^2}\alpha_{\mathbf{p}}\alpha_{\mathbf{p}-\mathbf{k}}, \tag{2.16}$$

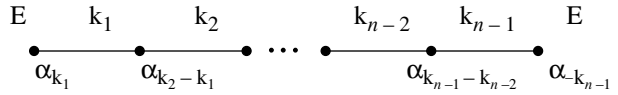
$$\sigma_{\mathrm{eff}3} = \sum_{\mathbf{p},\,\mathbf{k}}\frac{(\mathbf{k}\cdot\mathbf{E})(\mathbf{p}\cdot\mathbf{E})(\mathbf{p}\cdot\mathbf{k})}{\mathbf{k}^2\mathbf{p}^2\mathbf{E}^2}\alpha_{\mathbf{p}}\alpha_{\mathbf{k}-\mathbf{p}}\alpha_{-\mathbf{k}}, \tag{2.17}$$

and so on.

By induction, one can show that

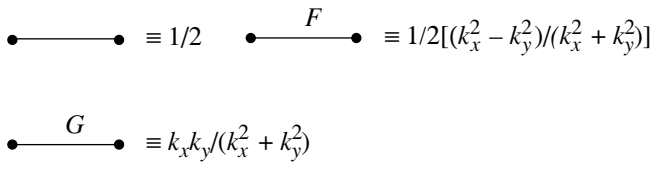
$$\sigma_{\mathrm{eff}n} = (-1)^{n+1}\sum_{\mathbf{k}_1,\ldots,\mathbf{k}_{n-1}}\frac{(\mathbf{k}_1\cdot\mathbf{E})(\mathbf{k}_1\cdot\mathbf{k}_2)\ldots(\mathbf{k}_{n-2}\cdot\mathbf{k}_{n-1})(\mathbf{k}_{n-1}\cdot\mathbf{E})}{\mathbf{E}^2\cdot\mathbf{k}_1^2\ldots\mathbf{k}_{n-1}^2}\alpha_{\mathbf{k}_1}\alpha_{\mathbf{k}_2-\mathbf{k}_1}\ldots\alpha_{\mathbf{k}_{n-1}-\mathbf{k}_{n-2}}\alpha_{-\mathbf{k}_{n-1}}. \tag{2.18}$$

Thus, all the corrections to the effective conductivity have a "chain" form and can be graphically represented as



However, to satisfy condition (2.6), it is necessary to subtract all the disconnected *n*th-order diagrams consisting of two chains from this diagram and then add the diagrams consisting of three chains. The construction of this alternating sum corresponds to the elimination of the terms containing the corrections proportional to possible homogeneous vector harmonics $\nabla\psi$.

We now analyze the structure of expression (2.18) and the corresponding diagram in some detail. This diagram contains "vertices" corresponding to the coefficients of the Fourier expansion (2.10) and "propagators" corresponding to two-dimensional wave vectors **k**. The propagators appear in three forms:



It is remarkable that Eq. (2.18) (after the subfraction of the alternating sum, as explained above) contains information about all the elements of the effective conductivity tensor. The diagrams corresponding to the isotropic part of this tensor defined in (2.13) contain an even number of the *F* propagators and an even number of the *G* propagators. The diagrams corresponding to the anisotropic part of the diagonal elements of this tensor,

$$\sigma_{\text{anysotr}} \equiv \frac{1}{2}(\sigma_{xx} - \sigma_{yy}), \tag{2.19}$$

contain an odd number of the *F* propagators and an even number of *G*. Finally, the diagrams corresponding to off-diagonal elements of the effective conductivity tensor σ_{xy} contain an odd number of the *G* propagators.

3. PERTURBATIVE CALCULATIONS FOR THE TWO-COLOR CHESSBOARD

To show how this perturbation theory and the diagram technique work, we consider the chessboard case. We assume that

$$\begin{aligned} \sigma_1 &= 1 + \delta, \\ \sigma_2 &= 1 - \delta. \end{aligned} \tag{3.1}$$

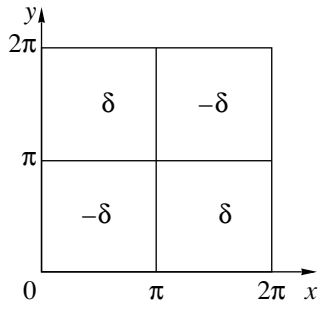


Fig. 1.

Then,

$$\sigma_{\text{eff}} = \sqrt{1-\delta^2} = 1 - \frac{\delta^2}{2} - \frac{\delta^4}{8} - \frac{\delta^6}{16} - \frac{5\delta^8}{128} - \frac{7\delta^{10}}{256} + \ldots. \tag{3.2}$$

To reproduce this formula by means of perturbation theory, we must calculate the Fourier coefficients for the function $\alpha(x, y)$. Let $\alpha(x, y)$ be a periodical function with the periods 2π in each coordinate such that

$$\begin{aligned} \alpha(x, y) &= -\delta \text{ if } 0 < x < \pi, \quad 0 < y < \pi, \\ \alpha(x, y) &= +\delta \text{ if } \pi < x < 2\pi, \quad 0 < y < \pi, \\ \alpha(x, y) &= +\delta \text{ if } 0 < x < \pi, \quad \pi < y < 2\pi, \\ \alpha(x, y) &= -\delta \text{ if } \pi < x < 2\pi, \quad \pi < y < 2\pi, \end{aligned} \tag{3.3}$$

see Fig. 1.

Then,

$$\alpha_{2k, m} = \alpha_{n, 2l} = 0, \tag{3.4}$$

$$\alpha_{2k+1, 2l+1} = \frac{16\delta}{(2\pi)^2(2k+1)(2l+1)}. \tag{3.5}$$

It is easy to see that all the diagrams with an odd number of vertices give a zero contribution. Indeed, the sum of wave numbers of all the vertices in a diagram must vanish (see Eq. (2.18)) and only odd values of wave numbers can occur (see Eqs. (3.4)–(3.5)). However, the sum of an odd number of odd numbers cannot vanish. Thus, we must consider only even orders of the perturbation theory.

Next, if a diagram contains an odd number of propagators *G*, the contribution of this diagram is equal to zero because, for any set of momenta $\mathbf{k} = (k_x, k_y)$, there is the set $\mathbf{k}' = (-k_x, k_y)$ giving the contribution of the opposite sign. Similarly, the diagrams containing an odd number of propagators *F* give vanishing contributions because the contribution of any diagram with a certain set of momenta $\mathbf{k} = (k_x, k_y)$ is cancelled by that with the momenta $\mathbf{k} = (k_y, k_x)$. Thus, σ_{anisotr} and σ_{xy} are equal to zero, and the effective conductivity of the chessboard is isotropic.

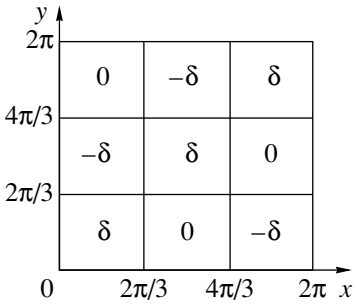


Fig. 2.

Now, we are in a position to calculate the diagrams corresponding to the effective conductivity of the chessboard. In the second order of the perturbation theory, we only have the diagram

that corresponds to the expression

$$\sigma_{\text{eff}2} = -\frac{1}{2}\sum_{\mathbf{k}}\alpha_{\mathbf{k}}\alpha_{-\mathbf{k}} = -\frac{256\delta^2}{2(2\pi)^4}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\frac{1}{(2k+1)^2(2l+1)^2}. \quad (3.6)$$

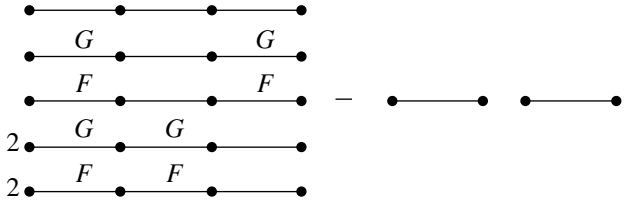
Using the well-known formula [10]

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \zeta_R(2) = \frac{\pi^2}{6}, \quad (3.7)$$

we obtain

$$\sigma_{\text{eff}2} = -\frac{\delta^2}{2}. \quad (3.8)$$

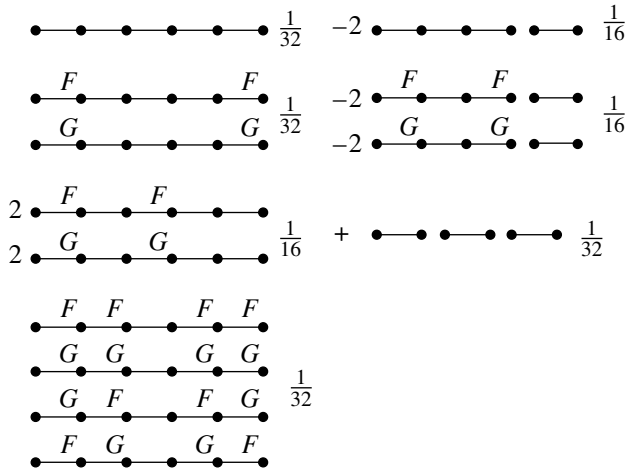
The set of the fourth-order diagrams is given by



Here, the first connected diagram gives the contribution $\delta^4/8$. The combination of the second and the third diagrams again gives $\delta^4/8$, the disconnected diagram gives the contribution $\delta^4/8$, while the last two connected diagrams do not give a contribution. Thus, the result is

$$\sigma_{\text{eff}4} = -\frac{\delta^4}{8}. \quad (3.9)$$

Next, we present only the diagrams giving nonvanishing contributions in the sixth order of the perturbation theory with the corresponding results:



Summing these, we have

$$\sigma_{\text{eff}6} = -\frac{\delta^6}{16}. \quad (3.10)$$

Quite similarly, we can also obtain

$$\sigma_{\text{eff}8} = -\frac{5\delta^8}{128} \quad (3.11)$$

and

$$\sigma_{\text{eff}10} = -\frac{7\delta^{10}}{256}. \quad (3.12)$$

The results of perturbative calculations in Eqs. (3.8)–(3.12) coincide with those obtained by expanding the general formula (3.2). This confirms the reliability of the perturbation theory developed here and encourages us to apply it to more complicated situations.

4. PERTURBATIVE CALCULATIONS FOR THE THREE-COLOR CHESSBOARDS

We now consider three-color structures with a square lattice symmetry, which we call the three-color chessboards. To do it, we again turn to formula (1.1). It is easy to notice that Eq. (1.1) can be rewritten as

$$\frac{\sigma_{\text{eff}} - \sigma_1}{\sigma_{\text{eff}} + \sigma_1} + \frac{\sigma_{\text{eff}} - \sigma_2}{\sigma_{\text{eff}} + \sigma_2} = 0. \quad (4.1)$$

Equation (4.1) was written by Bruggeman [4] as an approximative one (in the dipole approximation), but it appears to be exact for a large class of two-color coverings [2]. One can try to generalize this equation for the case of three conductivities (a three-color system) with equal weights as

$$\frac{\sigma_{\text{eff}} - \sigma_1}{\sigma_{\text{eff}} + \sigma_1} + \frac{\sigma_{\text{eff}} - \sigma_2}{\sigma_{\text{eff}} + \sigma_2} + \frac{\sigma_{\text{eff}} - \sigma_3}{\sigma_{\text{eff}} + \sigma_3} = 0, \quad (4.2)$$

which is equivalent to

$$\sigma_{\text{eff}}^3 + \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)\sigma_{\text{eff}}^2 - \frac{1}{3}(\sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_2\sigma_3)\sigma_{\text{eff}} - \sigma_1\sigma_2\sigma_3 = 0. \tag{4.3}$$

There are different tessellations of the plane allowing three-color (i.e., three-conductivity) coverings (see, e.g., [11]). Here, we consider the simplest tessellation: it is again a chessboard but is covered by three colors with the same statistical weights. We again introduce the function $\alpha(x, y)$ that is periodic with the periods 2π and is defined by the rules

$$\begin{aligned}
&\alpha(x, y) = \delta \text{ if } 0 < x < \frac{2\pi}{3}, \quad 0 < y < \frac{2\pi}{3},\\
&\alpha(x, y) = 0 \text{ if } \frac{2\pi}{3} < x < \frac{4\pi}{3}, \quad 0 < y < \frac{2\pi}{3},\\
&\alpha(x, y) = -\delta \text{ if } \frac{4\pi}{3} < x < (2\pi), \quad 0 < y < \frac{2\pi}{3},\\
&\alpha(x, y) = -\delta \text{ if } 0 < x < \frac{2\pi}{3}, \quad \frac{2\pi}{3} < y < \frac{4\pi}{3},\\
&\alpha(x, y) = \delta \text{ if } \frac{2\pi}{3} < x < \frac{4\pi}{3}, \quad \frac{2\pi}{3} < y < \frac{4\pi}{3},\\
&\alpha(x, y) = 0 \text{ if } \frac{4\pi}{3} < x < 2\pi, \quad \frac{2\pi}{3} < y < \frac{4\pi}{3},\\
&\alpha(x, y) = 0 \text{ if } 0 < x < \frac{2\pi}{3}, \quad \frac{4\pi}{3} < y < 2\pi,\\
&\alpha(x, y) = -\delta \text{ if } \frac{2\pi}{3} < x < \frac{4\pi}{3}, \quad \frac{4\pi}{3} < y < 2\pi,\\
&\alpha(x, y) = \delta \text{ if } \frac{4\pi}{3} < x < 2\pi, \quad \frac{4\pi}{3} < y < 2\pi,
\end{aligned} \tag{4.4}$$

see Fig. 2.

Obviously, this function corresponds to the periodic distribution of three conductivities

$$\begin{aligned}
\sigma_1 &= 1 + \delta,\\
\sigma_2 &= 1 - \delta,\\
\sigma_3 &= 1,
\end{aligned} \tag{4.5}$$

and the corresponding Fourier coefficients are given by

$$\alpha_{3k, m} = \alpha_{n, 3l} = 0. \tag{4.6}$$

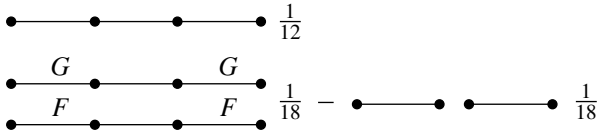
$$\alpha_{3k+1, 3l+1} = \alpha_{3k+2, 3l+2} = 0, \tag{4.7}$$

$$\alpha_{3k+2, 3l+1} = \alpha^*_{3l+1, 3k+2} = \frac{9\sqrt{3}ie^{i\pi/3}}{(2\pi)^2(3k+2)(3l+1)}. \tag{4.8}$$

Now, we are in a position to calculate perturbative contributions to the isotropic part of the effective conductivity for the three-color chessboard. In the second order of the perturbation theory, we again have only one diagram, which contributes

$$\sigma_{\text{eff}2} = -\frac{\delta^2}{3}, \tag{4.9}$$

where Eq. (3.7) was used again. In the fourth order, we have the following set of diagrams with nonvanishing contributions:



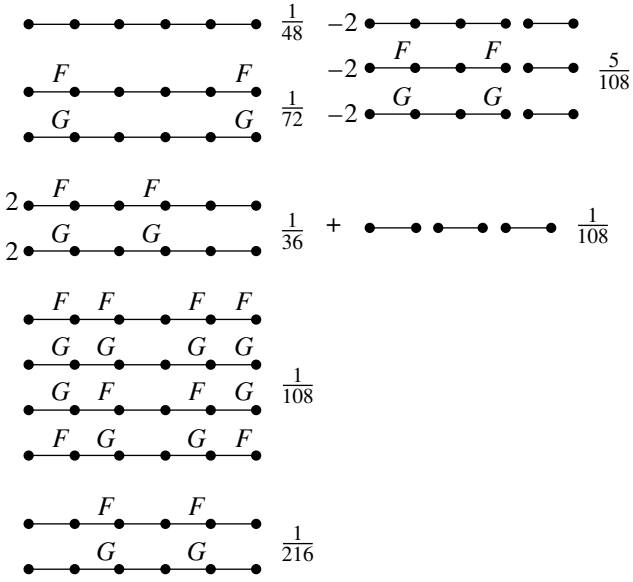
These result in

$$\sigma_{\text{eff}4} = -\frac{\delta^4}{12}. \tag{4.10}$$

In calculating the sums corresponding to the above diagrams, we have used the formula [10]

$$S \equiv \sum_{k=-\infty}^{\infty} \frac{1}{3k+1} = -\sum_{k=-\infty}^{\infty} \frac{1}{3k+2} = -\sum_{k=0}^{\infty} \frac{1}{(3k+1)(3k+2)} = \frac{\pi}{3\sqrt{3}}. \tag{4.11}$$

We now present the set of all the sixth-order diagrams giving nonvanishing contributions:



Summing these contributions, we have

$$\sigma_{\text{eff}6} = -\frac{17\delta^6}{432}. \tag{4.12}$$

Thus, we have calculated the symmetric part of the effective conductivity tensor

$$\sigma_{\text{eff}} \equiv \sigma_{\text{sym}} = \sigma_{xx} = \sigma_{yy}.$$

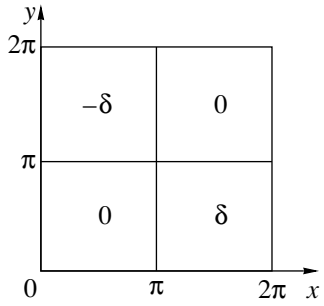


Fig. 3.

We can now substitute the expression for the isotropic (symmetric) part of the effective conductivity found up to the sixth order of the perturbation theory,

$$\sigma_{\text{eff}} = 1 - \frac{\delta^2}{3} - \frac{\delta^4}{12} - \frac{17\delta^6}{432} - \ldots, \tag{4.13}$$

into the three-color Bruggeman equation (4.2). We then see that this equation is satisfied. This gives us some grounds to think that this equation describes precisely the effective conductivity for the chaotic three-color distributions of conductivity on the plane just like it does for the two-color distributions, where Bruggeman formula (4.1) coincides with the exact Dykhne formula.

It is necessary to stress that, for the chessboard with a conductivity described by the function $\alpha(x, y)$ in Eq. (4.4), there is an anisotropic term already in the second order of the perturbation theory corresponding to the diagram

This term is equal to

$$\sigma_{xy} = -\frac{486\delta^2}{(2\pi)^4} \times \sum_{k,\,l=-\infty}^{\infty} \frac{1}{(3k+2)(3l+1)((3k+2)^2+(3l+1)^2)} \approx 0.06\delta^2. \tag{4.14}$$

We stress that for isotropic structures, the conductivity is purely local in the second order and has a universal form that does not depend on its particular structure. In anisotropic structures, we encounter nonlocal propagators already in the second order and the results are not universal.

We now consider another three-color chessboard with nonequal weights, which however is isotropic in contrast to the previous one. The conductivity is given by the formulas

$$\begin{aligned} &\alpha(x, y) = 0 \text{ if } 0 < x < \pi, \quad 0 < y < \pi, \\ &\alpha(x, y) = +\delta \text{ if } \pi < x < 2\pi, \quad 0 < y < \pi, \\ &\alpha(x, y) = -\delta \text{ if } 0 < x < 2\pi, \quad \pi < y < 2\pi, \\ &\alpha(x, y) = 0 \text{ if } \pi < x < 2\pi, \quad \pi < y < 2\pi, \end{aligned} \tag{4.15}$$

see Fig. 3.

Although this chessboard is a three-color one, it possesses all the symmetries of the two-color chessboard and is isotropic.

The nonzero Fourier coefficients are given by

$$\alpha_{2k+1,\,0} = \frac{1}{i\pi(2k+1)}, \tag{4.16}$$

$$\alpha_{0,\,2l+1} = \frac{1}{i\pi(2l+1)}. \tag{4.17}$$

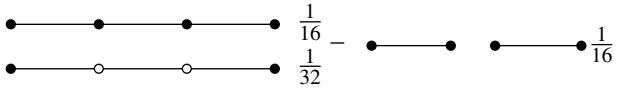
It is convenient to denote these Fourier coefficients by different vertices: the coefficient in (4.16) by a black circle and the one in (4.17) by a white circle.

The contribution of the second order of perturbation theory to the effective conductivity is given by the diagram

and is equal to

$$\sigma_{\text{eff}2} = -\frac{\delta^2}{4}. \tag{4.18}$$

The fourth-order diagrams have the form



Thus, the fourth-order contribution is

$$\sigma_{\text{eff}4} = -\frac{\delta^4}{32}. \tag{4.19}$$

On the other hand, the Bruggeman equation for this structure can be written as [4]

$$2\frac{\sigma_{\text{eff}} - \sigma_1}{\sigma_{\text{eff}} + \sigma_1} + \frac{\sigma_{\text{eff}} - \sigma_2}{\sigma_{\text{eff}} + \sigma_2} + \frac{\sigma_{\text{eff}} - \sigma_3}{\sigma_{\text{eff}} + \sigma_3} = 0, \tag{4.20}$$

or equivalently,

$$\sigma_{\text{eff}}^3 + \frac{\sigma_2 + \sigma_3}{2}\sigma_{\text{eff}}^2 - \frac{\sigma_1(\sigma_2 + \sigma_3)}{2}\sigma_{\text{eff}} - \sigma_1\sigma_2\sigma_3 = 0. \tag{4.21}$$

In our case, where

$$\sigma_1 = 1, \quad \sigma_2 = 1 + \delta, \quad \sigma_3 = 1 - \delta,$$

we easily obtain

$$\sigma_{\text{eff}} = 1 - \frac{\delta^2}{4} - \frac{\delta^4}{16} - \ldots. \tag{4.22}$$

It is obvious that in the fourth order of the perturbation theory, there is a contradiction between the results of perturbative calculations (4.19) and those following from the Bruggeman equation (4.22).

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (grant no. 99-02-18409). The work of A.Yu. Kamenshchik was also supported by CARIPLO Scientific Foundation. We are grateful to A.M. Dykhne, D.J. Bergman, and L.G. Fel who have introduced us to the problem of the effective conductivity of composite systems. We thank J.B. Keller for interest in our work. We are also grateful to A.S. Ioselevich, I.V. Kolokolov, V.G. Marikhin, A.V. Marshakov, A.Yu. Morozov, A.A. Rosly, and A.V. Zabrodin for useful remarks.

REFERENCES

1. Lord Rayleigh, Philos. Mag., Ser. 5 **34**, 481 (1892).
2. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
3. J. B. Keller, J. Math. Phys. **5**, 548 (1964).
4. D. A. G. Bruggeman, Ann. Phys. (Leipzig) **24**, 636 (1935).
5. L. G. Fel, V. Sh. Machavariani, and D. J. Bergman, cond-mat/0007022.
6. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **114**, 1121 (1998) [JETP **87**, 612 (1998)].
7. C. Herring, J. Appl. Phys. **31**, 1939 (1960).
8. A. M. Dykhne, private communication.
9. D. J. Bergman, private communication.
10. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1965; Nauka, Moscow, 1979).
11. P. Engel, *Geometric Crystallography: An Axiomatic Introduction to Crystallography* (Kluwer, Dordrecht, 1986).

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1268–1278.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1463–1475.

NONLINEAR PHYSICS

Multiple Fragmentation of Wave Packets in a Nonlinear Medium with Normal Dispersion of the Group Velocity

A. G. Litvak*, V. A. Mironov, and E. M. Sher

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia
**e-mail: litvak@appl.sci-nnov.ru*

Received May 22, 2000

Abstract—The features of the dynamics of the self-effects of the wave field are considered for the case when the dynamics is described by the nonlinear Schrodinger equation (NSE) with a hyperbolic spatial operator. An analytical investigation of the characteristic regimes of the self-effects is carried out; these regimes are due to the spatial competition of the self-focusing compression of the wave packet in one direction and its defocusing in another. The initial distributions of the wave field are analyzed with the goal of using them in a numerical modeling for illustration of the features of the self-effects. It is shown that three stages can be discerned in the evolution of the localized distributions: self-focusing filamentation of the wave field in the transverse direction, and compression and subsequent fragmentation in the defocusing (longitudinal) direction. The strongest nonuniformities are excited in the self-similar collapse of the wave field to hyperbolas (to the characteristic curves of the hyperbolic operator of the NSE). The stabilization of the development of the fragmentation instability is discussed separately. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In addition to the well-known nonlinear Schrodinger equation (NSE), the dynamics of the self-effects of wave packets is widely studied by means of the nonlinear Schrodinger equation (NSE) with a hyperbolic spatial operator [1]:

$$i\frac{\partial u}{\partial t} + \Delta_\perp u - \frac{\partial^2 u}{\partial z^2} + |u|^2 u = 0. \quad (1)$$

In the two-dimensional case (the Laplacian operator with respect to the transverse coordinates has the form $\Delta_\perp = \partial^2/\partial x^2$) it governs the dynamics of the self-effects of a wave on the surface of a liquid [1–4] and the spatial evolution of wave beams in a magnetized plasma in the parameter region corresponding to saddle points on the refractive index surface [5–7]. Recently, however, most of the interest has focused on the study of the self-effects of ultrashort laser pulses in media with normal dispersion of the group velocity $\partial v_{gr}/\partial\omega < 0$ (in this case, t corresponds to the coordinate of the centroid of the wave packet) [8–15].

A distinctive feature of Eq. (1) is that the distortion of the phase fronts due to the nonlinearity leads to focusing of the waves in the transverse direction and to dispersive spreading (defocusing) in the longitudinal direction. The competition between these two processes can result, in particular, in stabilization of the transverse collapse of an axially symmetric wave beam which would occur in the case $\partial/\partial z = 0$ [8]. The possibility of the onset of singularities in the evolution of a localized three-dimensional distribution is still in dispute [8, 14, 15].

The peculiarity of the self-effects of the waves in the framework of Eq. (1) is manifested even in the stage of modulational instability of a plane wave. The growth rate of this instability is given by the expression

$$\gamma^2 = (k_\perp^2 - k_z^2)[2A_0^2 - (k_\perp^2 - k_z^2)], \quad (2)$$

where A_0 is the amplitude of the plane wave, $k_\perp$ and k_z are the transverse and longitudinal wave numbers of the disturbance. The value of γ reaches a maximum not on a sphere, as in the NSE case, but on the hyperbolic surface

$$k_\perp^2 - k_z^2 = A_0^2, \quad (3)$$

i.e., in the framework of Eq. (1) the instability can occur on any arbitrarily small scales ($L \approx 1/k$), provided that the characteristic longitudinal ($L \approx 1/k_z$) and transverse ($L \approx 1/k_\perp$) scales of the nonuniformity satisfy relation (3). The most "large-scale" of the instabilities is the self-focusing instability ($k_\parallel = 0$, $k_\perp = A_0$). The evolution of wave packets which are rather "narrow" in the transverse direction, for which the stratification due to the development of this instability does not occur, has been considered in Refs. [5–10, 13, 15].

Numerical studies have shown that the dynamics of the self-effects is determined by the dimensionality of the problem [5, 8]. In a two-dimensional (one-dimensional in the transverse coordinate $r_\perp = x$) case a nonuniform compression of the wave field is accompanied by its fragmentation in the longitudinal direction [5]. In the central part of the localized distribution of the wave packet, where the intensity of the field is maximum, the compression is faster and the wave field becomes

separated in z direction into two isolated parts. This sequence of processes (compression, fragmentation into halves, etc.) is repeated several times against the background of the spreading of the wave field as a whole. The tendency to spread obviously follows from the equation for the second moments of the field distribution

$$a^2 = \int (x^2 + y^2)|u|^2 d\mathbf{r}, \quad b^2 = \int z^2 |u|^2 d\mathbf{r},$$

which can be obtained from (1):

$$\frac{d^2}{dt^2}(a^2 + b^2) = 8\int |\nabla u|^2 d\mathbf{r}. \tag{4}$$

Moreover, the rate of expansion of the localization region of the field is described by the same expression [the right-hand side of Eq. (4)] as in a linear medium. The nonlinearity governs the deformation of this region as the system evolves:

$$\frac{d^2}{dt^2}(a^2 - b^2) = 8H, \tag{5}$$

where H is the Hamiltonian of the system,

$$H = \int \left(|\nabla_\perp u|^2 - |u_z|^2 - \frac{|u|^4}{2} \right) d\mathbf{r}. \tag{6}$$

In the case of a smooth initial axially symmetric distribution which is highly elongated in the longitudinal direction z, fragmentation occurs if the power per unit length along z exceeds the critical power for self-focusing of an axially symmetric beam [this corresponds to $H < 0$ in Eq. (5)] [8, 13, 14].

These common properties of the self-effects of wave packets apparently depend on the specifics of the form of the spatial operator in (1). However, even from expression (2) for the growth rate of the modulation instability of a plane wave it is clear that these specifics can play an important role.

In the present study we examine several new self-effect regimes which derive from the hyperbolicity of the spatial operator of the initial Eq. (1) and whose realization depends on the form of the initial distribution $u(r_\perp, z)$. It is shown that some of them are initiated by an initial stratification of the wave field in the transverse direction, while others have the opposite property of weakening the development of the longitudinal and transverse instabilities on account of a competition between them. In a numerical study we have considered a two-dimensional version ($\Delta_\perp = \partial^2/\partial x^2$) of the initial equation. In contrast to earlier studies, here our attention is focused on investigating the long-term evolution of the system. The analytical results have a wider range of applicability, since, as will be easily seen, they are comparatively easy to generalize to the three-dimensional case.

2. ANALYTICAL STUDY OF THE FEATURES OF THE SELF-EFFECTS

Let us consider some general features of the self-effects of a wave system described by the nonlinear Schrodinger equation with a hyperbolic spatial operator (1), after first analyzing the stationary and self-modeling structures described by this equation and examining its hydrodynamic analog.

2.1. Stationary Self-Effects

The existence of a stationary spatially bounded distribution of the wave field is contradictory to the integral equations (4) and (5). They attest to the fundamentally dynamic character of the evolution of the field. Let us consider the structural features of nonlocalized stationary solutions in the case of a hyperbolic NSE for the purpose of further investigation of the dynamics of such initial field distributions.

We write the solution of the initial two-dimensional equation (1) in the form $u(x, z, t) = A\exp(i\varphi)$, where $A(x, z, t)$ and $\varphi(x, z, t)$ are real functions. As a result, we obtain

$$\frac{\partial \varphi}{\partial t} = \left(\frac{\partial \varphi}{\partial x}\right)^2 - \left(\frac{\partial \varphi}{\partial z}\right)^2 + A^2 + \frac{A_{xx} - A_{zz}}{A}, \tag{7}$$

$$\frac{1}{2}\frac{\partial A^2}{\partial t} = \frac{\partial}{\partial x}\varphi_x A^2 - \frac{\partial}{\partial z}\varphi_z A^2. \tag{8}$$

In contrast to the corresponding equations for the standard NSE, in (7) and (8) the derivatives with respect to z appear with a minus sign, i.e., the energy fluxes enter the continuity equation (8), for example, with different signs. One can construct a stationary solution

$$A^2(x, z) = \text{const}, \tag{9}$$

in which these fluxes compensate each other. Then the following relations will clearly hold:

$$\varphi_x A^2 = \frac{\partial \Phi}{\partial z}, \quad \varphi_z A^2 = \frac{\partial \Phi}{\partial x}, \tag{10}$$

where Φ is an arbitrary function. If Φ is chosen in accordance with the equation

$$\left(\frac{\partial \Phi}{\partial x}\right)^2 - \left(\frac{\partial \Phi}{\partial z}\right)^2 = 0, \tag{11}$$

then it is easy to see from (10) that φ satisfies the same equation (11):

$$\varphi_x^2 - \varphi_z^2 = 0. \tag{12}$$

At the same time, we find from Eq. (7) under condition (9) that

$$\varphi = \left(A^2 + \frac{A_{xx} - A_{zz}}{A}\right)t. \tag{13}$$

Substituting (13) into (12), we obtain the following equation for determining the stationary distribution of the field amplitude:

$$\left(A^2 + \frac{A_{xx} - A_{zz}}{A}\right)_x^2 - \left(A^2 + \frac{A_{xx} - A_{zz}}{A}\right)_z^2 = 0. \quad (14)$$

This equation, like the initial equation (1), contains wave operators and therefore does not have localized solutions [16]. To form a spatially bounded stationary distribution one can use a superposition of several nonlocalized wave structures. In what follows (see Section 3.3) we use extremely simple solutions of the form

$$|u|^2 = A^2 = B \pm C(x \pm z), \quad (15)$$

where B and C are arbitrary constants.

The initial localized distribution of the wave field constructed from solutions of the type (15) has the following distinctive feature. In accordance with the method used to construct the solution (9) for each individual component of a stationary structure of the type (15), the expression $A_{xx} - A_{zz}$ is equal to zero. In the present case of the hyperbolic NSE this term describes "diffraction" of the wave field. Thus the stationary distribution (9) is formed from a set of "nondiffracting" structures. The "diffraction" inherent to the wave field is obviously due to the presence of a transition region from one nonlocalized distribution (15) to another.

2.2. Self-Similar Structures

The axially symmetric solution of the NSE in the present case of a hyperbolic two-dimensional spatial operator corresponds to a self-similar field distribution that is a function of $\zeta = x^2 - z^2$. The quantity $\zeta = \text{const}$ is an invariant that is conserved in coordinate transformations of the "Lorentz" type (such a coordinate transformation leaves Eq. (1) unchanged). As a result, we arrive at the following equation for the self-similar field distribution:

$$i\frac{\partial u}{\partial t} + 4\zeta\frac{\partial^2 u}{\partial \zeta^2} + 4\frac{\partial u}{\partial \zeta} + |u|^2 u = 0. \quad (16)$$

The introduction of the new variable $\eta = \sqrt{\zeta}$ reduces (16) to an equation that is well known in the theory of self-focusing and describes axially symmetric wave beams [17]:

$$i\frac{\partial u}{\partial t} + \frac{1}{\eta}\frac{\partial}{\partial \eta}\eta\frac{\partial u}{\partial \eta} + |u|^2 u = 0. \quad (17)$$

In the sector $x^2 - z^2 = \zeta > 0$ (we call this the focusing sector) the distribution with $H < 0$ suffers a collapse. However, unlike the corresponding process in the NSE, here compression occurs not toward the axis of the system ($x = 0$, $z = 0$), but toward the hyperbola $x^2 - z^2 = \zeta_0 > 0$. It is clear that if the initial distribution is symmetric with x and z respect to and there will be two such hyperbolas, and, hence, one expects that simultaneously with the collapse there will be a stratification of the wave field in the transverse direction, as occurs in the self-focusing instability in the NSE. It is also important that the self-similar distribution localized with respect to $\eta\zeta$ is nonlocal with respect to the coordinate orthogonal to η.

In the sector $x^2 - z^2 = \zeta < 0$ the analogous transformation $\eta = \sqrt{-\zeta}$ reduces (13) to the equation

$$-i\frac{\partial u}{\partial t} + \frac{1}{\eta}\frac{\partial}{\partial \eta}\eta\frac{\partial u}{\partial \eta} - |u|^2 u = 0, \quad (18)$$

which describes self-defocusing of the wave field.

It should be specially noted that Eq. (17), which governs the evolution of the self-similar distributions, has the following feature. In the two-dimensional case under discussion the initial equation (1) describes a process of "one-dimensional" self-focusing weakened by defocusing. Equation (17), on the contrary, has the form of the "two-dimensional" NSE, which attests to the possibility of "strong" self-focusing (collapse) of a nonlocalized (with respect to the coordinate orthogonal to η) wave field in the system under study.

For an analysis of the process in which the collapse is weakened as a result of defocusing, it is convenient to transform to an orthogonal coordinate system $\zeta = x^2 - z^2$, $\mu = xz$. As a result, we arrive at the following equation, which is equivalent to the initial equation (1):

$$i\frac{\partial u}{\partial t} + 4\left(\zeta\frac{\partial^2 u}{\partial \zeta^2} + \frac{\partial u}{\partial \zeta} + 2\mu\frac{\partial^2 u}{\partial \zeta \partial \mu} - \frac{\partial^2 u}{\partial \mu^2}\right) + |u|^2 u = 0. \quad (19)$$

From this equation we see that for an initial distribution that depends weakly on μ, the derivatives with respect to μ in Eq. (19) can be neglected, and we naturally arrive at Eq. (17), which describes a collapse. However, as the collapse occurs the distribution becomes "one-dimensional" near $\zeta \approx \zeta_0$, and the role of the defocusing term $\zeta\partial^2 u/\partial\mu^2$ becomes appreciable. A detailed numerical investigation of the features of the self-effects in this case is presented in Section 3.2.

2.3. Solenoidal Distributions

As in the case of the NSE, the system of Eqs. (7), (8) for the amplitudes and phases of the field can be written in the form of hydrodynamic equations. Setting $\rho = A^2$ and introducing a velocity $\mathbf{v}$ vector with components $v_x = -\varphi_x$, $v_z = \varphi_z$, we obtain the equations of ideal hydrodynamics:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) = 0, \quad (20)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\mathbf{v} = \operatorname{div}\hat{P} \quad (21)$$

with a specific form of the pressure tensor,

$$\hat{P} = \begin{pmatrix} -\frac{\rho}{2} - \frac{(\sqrt{\rho})_{xx} - (\sqrt{\rho})_{zz}}{2} & 0 \\ 0 & \frac{\rho}{2} + \frac{(\sqrt{\rho})_{xx} - (\sqrt{\rho})_{zz}}{2} \end{pmatrix}, \quad (22)$$

which reflects all of the features of the evolution of the solutions of the NSE with a hyperbolic spatial operator. In particular, it turns out that the solenoidal motions of the liquid are strongly coupled to the potential motions. For example, the azimuthal modes of the NSE in the present case correspond to complex wave fields specified at $t = 0$ in the form

$$\Psi \propto \exp(i\alpha xz). \quad (23)$$

A hyperbolic phase of just this type ($\varphi = \alpha xz$) determines the circulation of the flux of the wave field around the origin of the coordinate system. This is because the fluxes along x and z appear in the continuity equation (7) with different signs. During the evolution of the system the "solenoidal" process, which is determined by the initial distribution (23), is damped out. One can see this by examining the linear part of the eikonal equation (5):

$$\frac{\partial \varphi}{\partial t} = \left(\frac{\partial \varphi}{\partial x}\right)^2 - \left(\frac{\partial \varphi}{\partial z}\right)^2. \quad (24)$$

Substituting the initial (solenoidal) distribution of the phase front $\varphi = \varphi_0 = \alpha xz$ into the right-hand side of Eq. (24), we see that it produces $\varphi_t = \alpha(z^2 - x^2)$ a phase distribution $\varphi \propto z^2 - x^2$, which determines the potential motion of the liquid. The potential distribution of the phase, as is easily seen from (24), generates itself but with the opposite sign, i.e., the complex conjugate wave field.

Thus, the initial distributions of the form (23), containing a "solenoidal" phase, can be of interest, since the averaging of the spatial disturbances on account of rotation can actually lead to stabilization of instabilities. The evolution of the wave field in the radial direction, in the presence of the circulation of its flux with respect to angle, occurs under conditions of a sort of averaged and therefore weakened nonlinearity. The results of a numerical study which demonstrate the development of the fragmentation instability are presented in Section 3.1.

For enhancement of instability, on the other hand, one can choose, for example, an initial condition of the form

$$u \propto \exp(i\alpha x^2 z^2). \quad (25)$$

Such a phase of the wave field, together with the parabolic one ordinarily used [14], guides the energy flux into the focusing sector and weakens the spreading of the distribution in the defocusing region. The features of the evolution of a distribution of the type (25) are discussed in Section 3.2.

3. NUMERICAL STUDY OF THE SELF-EFFECTS

Let us consider several characteristic regimes of the dynamics of the self-effects of a wave field. A numerical simulation was done by the spectral method, using a fast Fourier transform (see [18]) in a square with a side equal to 30.

The initial study was done using 128×128 Fourier harmonics. However, the generation of strong nonuniformities in the self-effects of the wave field reduced the reliability of the numerical calculations considerably. Therefore, in the final versions discussed below we used the maximum possible number of harmonics in our situation (750×750). The accuracy of the calculations was monitored by checking the integrals of the equation. Let us first consider the evolution of the wave field with an initial transverse form of Gaussian distribution, and then the dynamics of other localized distributions that more realistically reflect the specific features of the initial NSE (1) with a hyperbolic spatial operator.

3.1. Wave Packet with Gaussian Field Distribution

Let us consider a wave field with an initial distribution of the form

$$u = \frac{u_0}{a}\exp\left(-\frac{x^2 + z^2}{2a^2}\right). \quad (26)$$

The evolution of a packet with such a form is given in Fig. 1 for the parameter values $u_0 = 25$ and $u = 3$. The initial field amplitude was three times larger than the value used in [5], making it possible to observe the transverse filamentation of the beam.

In the first stage (Fig. 1a) the process occurred as described in [5]. The nonuniform compression of the wave field in the transverse direction as a result of self-focusing and the expansion in the longitudinal direction (defocusing) give rise to a dumbbell-shaped distribution. The redistribution of the wave field is accompanied by the formation of a maximum of the field at the ends of the "dumbbell" (Fig. 2).

Further, in the stage of expansion of the wave field as a whole one should distinguish two characteristic processes. First, it is seen that the regions of maximum field, as they expand along the z-axis (one to the right, the other to the left), excite a wave field with the characteristics of the hyperbolic operator ($\zeta = x + z$, $\eta = x - z$) (Fig. 1b). It has the form of a shock wave. Initially, at a high intensity of the source of the shock wave, the development of the fragmentation instability is observed at the leading edge (Fig. 1c). As the wave-

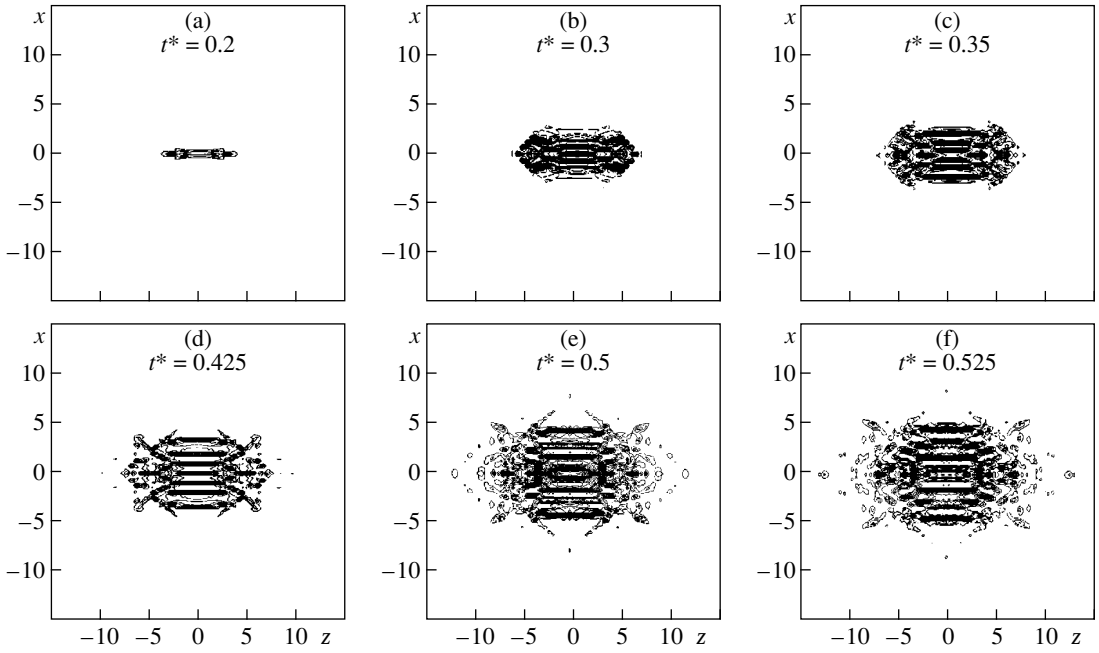


Fig. 1. Level lines $|u(x, z, t^*)|$ of in the evolution of a field with a Gaussian distribution.

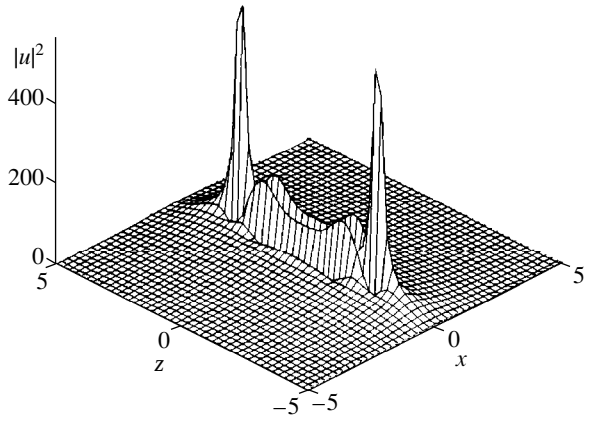


Fig. 2. Distribution of the wave field $|u(x, z, t^* = 0.2)|^2$.

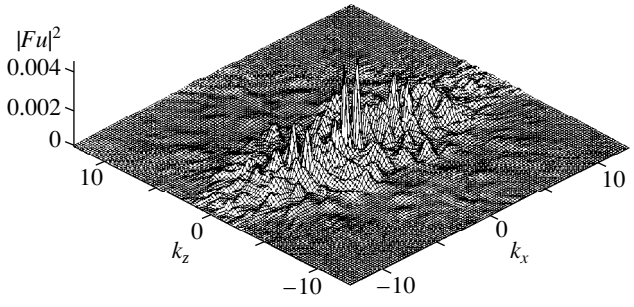


Fig. 3. Spatial spectrum of the wave field $|Fu(k_x, k_z, t^* = 0.45)|^2$.

field region expands and the maximum value of the field decreases, a weakening of the nonuniformities occurs behind the front of the shock wave (Figs. 1e, 1f).

In the central region one sees the development of a self-focusing filamentation of the wave field. Initially there is a structure with pronounced layering in the transverse direction (Fig. 1b), and then the fragmentation instability leads to the excitation of longitudinal nonuniformities as well. As the wave field expands, the number of transverse filaments grows. The maximum number of them corresponds to the number of filaments excited over the characteristic transverse dimension of the wave field as a result of the development of a self-focusing instability (for $k_z = 0$). In the longitudinal direction the layers extend along hyperbolas (Fig. 1c). We note further that for $t > 0.4$ there occurs a strong broadening of the spectrum. Up until that time the spectrum has a strong maximum at the center, which subsequently bifurcates (Fig. 3). We note that the regime of initial collapse turns out to be dominant not only in the case of Gaussian wave packets extending in the defocusing direction but also for fields which are more highly localized in that direction. A predominant nonuniformity of the wave field with respect to x leads to enhanced defocusing. As a result of the weakening of the "maximum" value of the field the self-effect process takes place in a manner analogous to the previous case, only more slowly than before, when the characteristic transverse scales of the wave field were the same ($a = b = 3$).

Let us conclude this section by considering the possibility of stabilizing the instability development. For

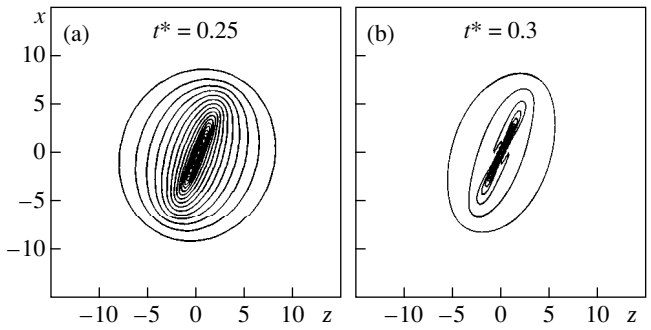


Fig. 4. Level lines of $|u(x, z, t^*)|$ in the evolution of a solenoidal distribution (27), $u_0 = 25$, $u = 3$.

this we use a complex generalization of the initial distribution (26) in the form

$$u = \frac{u_0}{a}\exp\left(-\frac{x^2 + z^2}{2a^2} + inxz\right) \tag{27}$$

with the same values of u_0 and a. A weakening of the effective (angle-averaged) nonlinearity was successfully achieved for $n = 3$ ($n = 1$ turned out to be insufficient). It is seen in Fig. 4 that the compression of the wave field occurs not toward the z-axis, as for $n = 0$, but toward some straight line in the xz plane. The distribution remains smooth up to $t = 0.25$, which is twice as long as in the case $n = 0$ (the self-focusing instability (see Fig. 1) turns out to be suppressed). The process culminates in multiple fragmentation. This fragmentation process differs from the previous versions of the formation of rather asymmetric nonuniformities.

3.2. Evolution of a Horseshoe-Shaped Field Distribution

Let us consider the evolution of a field distribution of the type

$$u = \frac{u_0}{a}\exp\left[-\frac{(x^2 - z^2)^2}{2a^4} - \frac{x^2 z^2}{2b^4}\right]. \tag{28}$$

The parameters a and b are chosen almost equal in order to study the role of a weak initial hyperbolicity of the distribution in the dynamics of a wave packet. Figure 5 shows the evolution of the wave field for $a = 3$, $b = 3.5$, and $u_0 = 25$. We note that with these parameters the initial distribution differs little from a product of super-Gaussian functions,

$$u \propto \exp\left(-\frac{x^4 + z^4}{2a^4}\right),$$

the dynamics of which we consider below for comparison.

The peculiarity of this self-effect regime is demonstrated in Figs. 5–7. A stratification of the wave field in the transverse direction develops immediately in the first stage. As a result, the field is concentrated in the focusing sector (Fig. 5b). Then the field increases rather sharply and becomes localized near the hyperbolas $\zeta = x^2 - z^2 = \pm 3$. This stage (Fig. 5c) was mentioned above

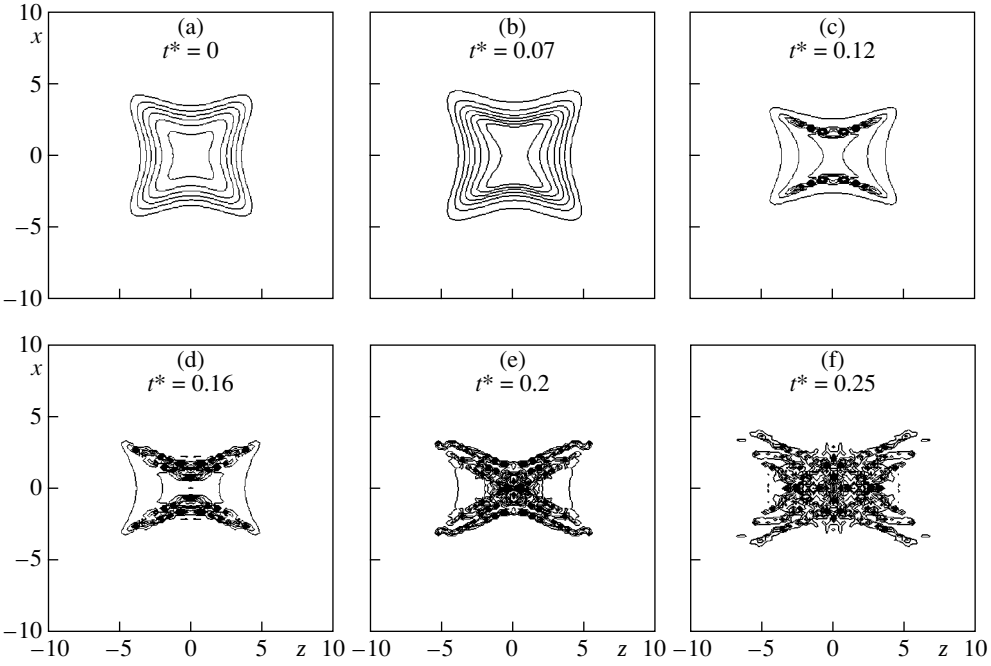


Fig. 5. Level lines of $|u(x, z, t^*)|$ in the evolution of a distribution with a slight hyperbolicity.

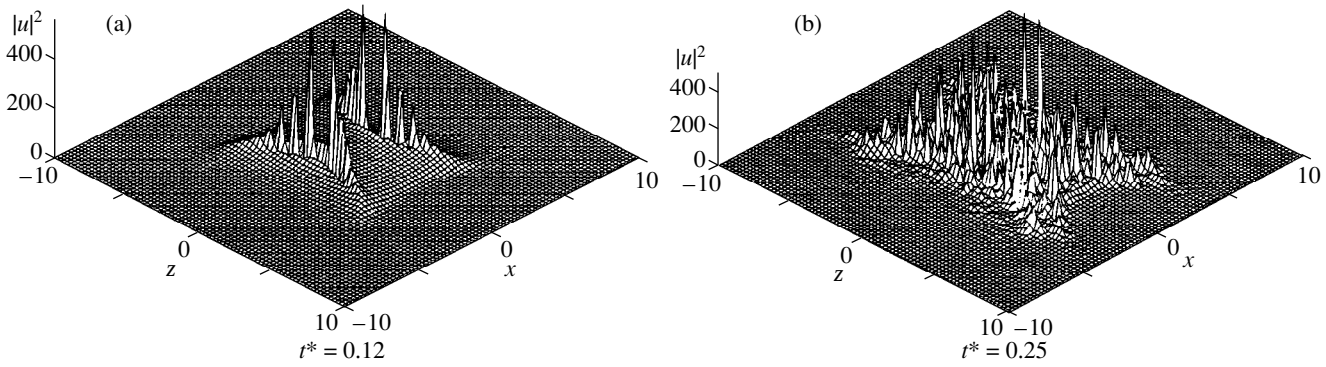


Fig. 6. Field distribution corresponding to the level lines of Figs. 5c and 5f.

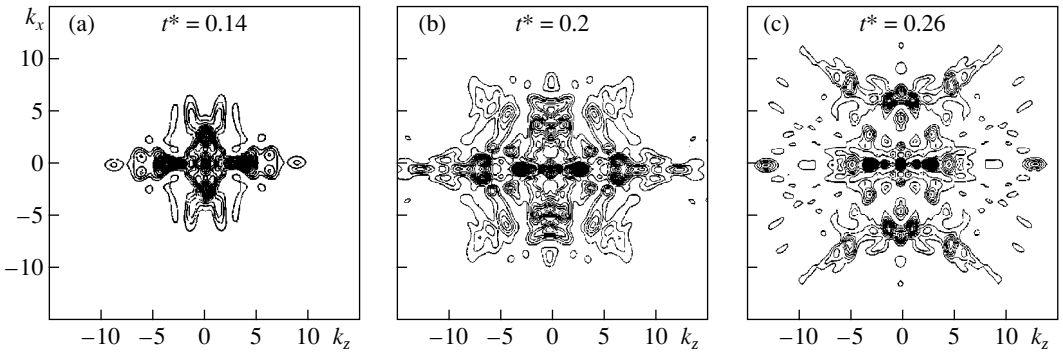


Fig. 7. Level lines of the spatial spectrum $|Fu(k_x, k_z, t^*)|$.

in the discussion of the self-similar structures. The subsequent increase in the field is accompanied by multiple fragmentation of the wave field along the hyperbolas $\zeta = \pm 3$ and an ultimate compression of the distribution in the focusing direction. The characteristic time for this process to take place is an order of magnitude shorter than that of the previous stage. It culminates in the formation of the structure shown in Fig. 6a.

Then one can see the following. The longitudinal (along the hyperbolas) nonuniformities are somewhat smoothed out, and the wave field localized near the hyperbola is stratified in the transverse direction (Fig. 5d). Thus two spatially bounded regions appear near the hyperbolas, with a consequent development of multiple fragmentation. This sequence of processes is repeated once again. As a result, the central region is filled with stronger nonuniformities than in the case of the evolution of a wave field with an initial Gaussian distribution. The number of transverse nonuniformities here is the same as in the previous version (Fig. 5e). However, the time of formation of the structure, with its considerably stronger nonuniformities, is about a factor of two shorter.

In the last stage (Figs. 5f, 6b) a rather uniform expansion of the wave field occurs as the nonuniformities fill the defocusing sector.

The main feature of the dynamics of the system in the case under consideration is an anomalous broadening of the spectrum of the wave field. A noticeable broadening begins after the self-similar collapse of the distribution toward the hyperbolas and the subsequent multiple fragmentation in the longitudinal direction.

This process can be explained as follows. As a result of the one-dimensional "collapse toward the hyperbolas" and the formation of a quasi-soliton field distribution in the transverse direction, the wave field exhibits a nonlinear frequency shift determined by the field A amplitude $u \propto \exp(iA^2t/2)$. It is clear that in the development of the fragmentation instability the phase is spatially modulated along the hyperbolas. This is manifested in a noticeable broadening of the spectrum. We note that the spectrum of the field amplitude is much narrower. The further modification of the spectrum is due to stratification of the hyperbolas.

The level lines of the modulus of the spectrum are shown in Fig. 7. The initial spectrum of the wave field ($t = 0$) is bounded by the region of the central peak (Fig. 7a). As the fragmentation instability develops and the hyperbolas become stratified, the level lines of the spectrum take on a rather exotic form (see Fig. 7).

It should be noted first that the distribution of the spatial spectrum is much smoother than the field distribution in coordinate space (cf. Figs. 5 and 7). A charac-

teristic feature of the spectrum is due to the presence of noticeable peaks (Fig. 7) at $k_x \approx \pm k_z = \pm 5$. This is a reflection of the rather periodic structure of the nonuniformities of the wave field which are excited as a result of the stratification and fragmentation instabilities. At long times the stratification of the spectrum and its localization in the central part and near the hyperbolas are more and more clearly visible (see Fig. 7c). The part of the spectrum on the hyperbola corresponding to nonuniformities whose growth rate is maximum (2) in a field with amplitude $u_A = 5$, i.e., in the mean field.

As u_0 falls off, the effect of a slight hyperbolicity in the initial distribution is manifested for quite a long time, all the way to $u_0 = 3$. Here there is a natural decrease in the number of "hyperbolas" that are generated in the course of the self-focusing filamentation of the wave beam. For $u_0 < 3$ the dominant process in the initial stage becomes the spreading of the wave field in the defocusing direction.

Analogous results are obtained for a somewhat different specification of an initial distribution with a slight hyperbolicity of the level lines:

$$u = \frac{u_0}{a}\exp\left[-\frac{(x^2-z^2)^2}{2a^4}-\frac{x^2+z^2}{2a^2}\right]. \qquad (29)$$

The development of self-focusing of the wave field near the hyperbolas and the subsequent multiple fragmentation in the longitudinal direction can be stabilized, for example, in the case of a complex generalization of the initial condition of the form (28):

$$u = \frac{u_0}{a}\exp\left[-\frac{(x^2-z^2)^2}{2a^4}-\frac{x^2z^2}{2b^4}\right]-i\frac{x^2z^2}{2}. \qquad (30)$$

The presence of the phase $-ix^2z^2/2$ in the initial distribution (30) leads to a solenoidal motion with a flux of the wave field into the defocusing sector. As a result, the formation of the field distribution occurs as in the development of the fragmentation instability of a wave packet with an initial Gaussian form (26) (Fig. 1a). The further evolution of the wave field takes place in an analogous way.

In the case of an initial distribution that is the complex conjugate of expression (30), the flux of the wave field is directed into the focusing sector. As a result, two rather strong peaks of the field initially form in the focusing sector, and then, as they spread in the defocusing direction, a multiple fragmentation develops. The subsequent evolution occurs as in the case of an initial distribution with a slight hyperbolicity (28).

Concluding this section, we note the results of a numerical study of the evolution of a super-Gaussian distribution (with $a = 3$, $b = a/\sqrt[4]{2}$, $u_0 = 25$, and in Eq. (28)). Analysis shows that the processes develop in an analogous way (see Fig. 5) but more slowly. The absence of a stage of self-similar collapse toward "hyperbolas" leads to field amplitudes that are insufficient for the onset of multiple fragmentation. The processes of stratification of the wave field and the fragmentation into halves occur as in the case of Gaussian beams (Fig. 1).

3.3. Evolution of Nonsmooth Distributions

Using a nonlocalized nondiffracting structure of the form (25), $u = u_0[(2a \pm x \pm z)/a^3]^{1/2}$, one can construct various spatially bounded initial distributions. We have investigated the dynamics of the self-effects for the three distributions whose square moduli are shown in Fig. 8. The first of them (Fig. 8a) is a quadrangular prism. The field distribution has the form

$$u = u_0 \times \begin{cases} \sqrt{(2a+x+z)/a^3} & \text{for } x, z<0;\ x+z\geq -2a, \\ \sqrt{(2a-x+z)/a^3} & \text{for } x>0,\ z<0;\ z-x\geq -2a, \\ \sqrt{(2a+x-z)/a^3} & \text{for } x<0,\ z>0;\ x-z\geq -2a, \\ \sqrt{(2a-x-z)/a^3} & \text{for } x, z>0;\ x+z\leq 2a. \end{cases} \qquad (31)$$

The second distribution (Fig. 8b) is specified with respect to z in a smaller interval $|z| \le a$ and is bounded in the defocusing direction by the vertical planes $z = \pm a$. Finally, the third distribution (Fig. 8c) is defined in a square $|x| \le a$, $|z| \le a$ and is additionally bounded by vertical planes at $x = \pm a$. We note that in the numerical investigation the initial distribution shown in Fig. 8 is approximated by a finite number of Fourier harmonics and is therefore smoothed accordingly. By sectioning the distribution of the prism type (Fig. 8) by vertical planes, one can form a "two-scale" distribution (Fig. 8c) analogous, for example, to the super-Gaussian distribution (28) with $b = a/\sqrt[4]{2}$.

A detailed numerical investigation of the dynamics of the self-effects of the distributions presented in Fig. 8 shows that the evolution of the wave field is largely analogous to that in the cases considered above, although the growth of the filamentation and fragmentation instabilities is substantial slower and the nonuniformities and wave-field generation that arise are smoother despite the nonsmooth nature of the initial distribution. For example, for initial conditions corresponding to Fig. 8a, the sequence of processes is the same as in the decay of a Gaussian distribution (26). First there is a focusing of the wave field in the transverse direction and its localization near the axis of the system, then the fragmentation instability develops, and only then comes the self-focusing instability accompanied by transverse stratification. The abrupt falloff of the initial field distribution in the defocusing direction due to the vertical planes (Fig. 8b) leads to a decrease in the spreading of the field in that direction.

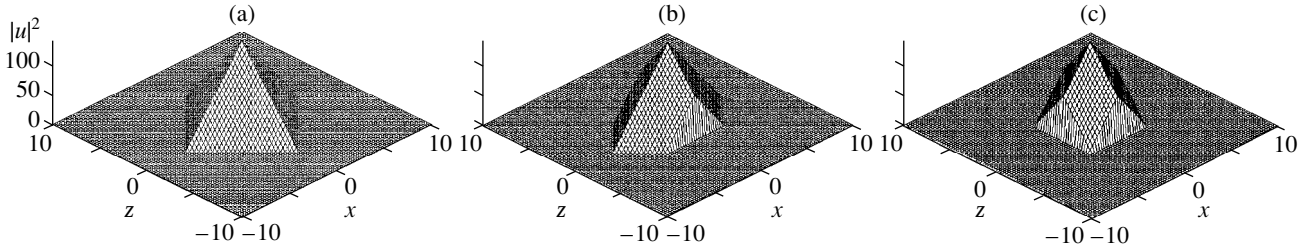


Fig. 8. "Nonsmooth" initial distributions.

The introduction of an additional bounding of the region of initial localization of the field in the focusing direction (Fig. 8c) makes the picture of the self-effects similar to the evolution of a beam of the super-Gaussian form (28) with $b = a/\sqrt[4]{2}$.

4. STATISTICAL DESCRIPTION OF THE LONG-TERM EVOLUTION

Numerical studies of the dynamics of the self-effects reveal the excitation of a broad spectrum of nonuniformities and thereby justify the use of a statistical approach for analysis of the subsequent behavior of the wave field. It is of particular interest to consider the long-term evolution of the system under study.

In a stochastic approach the wave field is described by a set of correlation functions. In the case of a uniform weak turbulence, i.e., in the spatially uniform case, the problem can be described by an equation for the spectrum of the radiation intensity [19].

In the spatially nonuniform wave field considered here, for solving the stated problem it is convenient to use the correlation function

$$\Gamma = \langle u(\mathbf{r}_1, z_1, t)u^*(\mathbf{r}_2, z_2, t)\rangle, \tag{32}$$

where the angle brackets denote a statistical averaging. The equation describing the evolution of Γ_{12} is obtained from the initial equation (1) for a complex amplitude by the necessary multiplications and ensemble averaging, in the same manner as the equations for the density matrix are found in quantum mechanics. The main difficulty in the nonlinear case under consideration is due to the necessity of decoupling the higher-order correlations. However, for a cubic nonlinearity on the assumption that the statistics of the field in a nonlinear medium differ only slightly from the initial Gaussian statistics, the equation for the correlation function turns out to be closed (see [17, 20] for details). In the variables

$$\begin{aligned} \mathbf{R} &= \frac{\mathbf{r}_1 + \mathbf{r}_2}{2}, \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \\ z &= \frac{z_1 + z_2}{2}, \quad q = z_1 - z_2 \end{aligned} \tag{33}$$

it has the form

$$\left(i\frac{\partial}{\partial t} + \nabla_{\mathbf{R}}\nabla_{\mathbf{r}} - \nabla_z\nabla_q\right)\Gamma_{12} = 2(\Gamma_{11} - \Gamma_{22})\Gamma_{12}. \tag{34}$$

We assume that for $t = 0$ the problem under study has translational symmetry with respect to t the correlation function Γ_{12} is determined by the expression

$$\begin{aligned} \Gamma_{12} = \frac{W}{ab}\exp\Bigg\{-\Bigg[&\frac{\mathbf{R}^2}{a^2} + \frac{\mathbf{r}^2}{\rho^2} \\ &+ i\alpha(\mathbf{R}\cdot\mathbf{r}) + \frac{z^2}{b^2} + \frac{q^2}{h^2} + i\beta zq\Bigg]\Bigg\}, \end{aligned} \tag{35}$$

where a and b are the widths of the wave field in the transverse and longitudinal directions, respectively, ρ and h are the effective correlation lengths of the field in the corresponding directions (they are related to the correlation lengths of ρ_0 and h_0 by the expressions $\rho^{-2} = \rho_0^{-2} + a^{-2}$, $h^{-2} = h_0^{-2} + b^{-2}$, and is the power of the wave field. In an aberrationless description of the processes we shall seek the solution of the equation in the form (35), where the parameters are functions of t. For the paraxial part of the beam ($R \ll a$) the Gaussian intensity distribution $\Gamma_{11}(\mathbf{R} + \mathbf{r}/2, z + q/2, t)$ and $\Gamma_{22}(\mathbf{R} - \mathbf{r}/2, z - q/2, t)$ are replaced by series expansions. Then to a first approximation for $\Gamma_{11} - \Gamma_{22}$ we find

$$\Gamma_{11} - \Gamma_{22} = \frac{W}{ab}\left(\frac{\mathbf{R}\cdot\mathbf{r}}{a^2} + \frac{zq}{b^2}\right). \tag{36}$$

Further, as usual, we substitute expression (35) into Eq. (34) with allowance for relation (36). Equating to zero the coefficients of different powers of $\mathbf{R}^n \cdot \mathbf{r}^m$ and $z^k q^p$ (n, m, k, and p are integers), we obtain equations for the parameters of the distribution (35).

We should mention first that the number of spatial nonuniformities in the nonlinear medium in the longitudinal and transverse directions, as in the analogous equation (34) corresponding to the standard NSE [8],

remains constant:

$$N_z = \frac{\beta(t)}{h(t)} = N_z(t=0),$$
$$N_r = \frac{\alpha(t)}{\rho(t)} = N_r(t=0). \quad (37)$$

When these invariants are taken into account, the equations for the width of the intensity distribution in the focusing and defocusing directions have the form

$$a'' = \frac{N_r}{a^3} - \frac{P}{a^2 b}, \quad b'' = \frac{N_z}{b^3} + \frac{P}{b^2 a}, \quad (38)$$

where $P = W/4$.

These equations obviously also describe the self-effects of a coherent field ($N_r = 1$, $N_z = 1$) in the paraxial approximation. A detailed study of the dynamics of the system (38) for $N_r = 1$, $N_z = 1$ was carried out in [5]. It showed that for $P/b_0 > 1$ ($b_0 = b(t = 0)$) the system after a certain number of oscillations in the transverse direction goes over to a regime of smooth spreading. In the cases considered in the present study the qualitative picture remains the same. However, the analytical results of Ref. [5] cannot be used to the fullest extent. The development of the instabilities and the formation of nonuniformities in the initial stage determine the values of N_r and N_z. The further evolution of the system is governed by Eqs. (35) and (36).

Numerical calculations show that the number of transverse nonuniformities N_r is determined by the self-focusing instability (Eq. (2) for $k_z = 0$, $N_r = 2\sqrt{P}/\pi^2$. As a result of the generation of nonuniformities and the expansion of the wave field in the longitudinal direction, the right-hand side of (36) changes sign and, consequently, the self-focusing compression gives way to a spreading in the transverse direction. No such definite conclusion can be reached for the value of N_z. It is evident and is confirmed by the numerical calculations that $N_z > N_r$, (the longitudinal fragmentation develops even in the absence of the self-focusing stratification [5]). Thus relations (35) and (36) show that in the long-term evolution of the system, the number of nonuniformities remains constant (38); their excitation leads to enhancement of the spreading of the wave field $a \sim \sqrt{N_r}\, z$, $b \sim \sqrt{N_z}\, z$.

5. CONCLUSION

We have considered a number of new scenarios for the dynamics of the self-effects of the wave fields in a system described by a nonlinear Schrodinger equation (1) with a hyperbolic spatial operator. For these scenarios there are characteristically three successive stages: self-focusing filamentation, compression, and fragmentation of transverse nonuniformities in the defocusing direction. The strongest nonuniformities are excited in the development of the fragmentation instability along hyperbolas during the self-similar collapse toward them (see Section 3.2).

The other scenarios, with initial distributions of the type in (27) and (31), are characterized by a slower development of the processes and the excitation of smoother nonuniformities. A statistical approach to the study of the long-term evolution of the system showed that the "number" of longitudinal and transverse nonuniformities is conserved during the expansion of the wave field.

Our results are applicable, e.g., to wave fields on the surface of a liquid. The investigated structural features in the behavior of wave packets should manifest themselves under suitable conditions. For example, the fragmentation of wave structures localized near hyperbolas (see Fig. 5) was observed in [21] in a study of the excitation of surface waves by nonuniform flows of a liquid.

For problems of the propagation of short laser pulses in media with normal dispersion of the group velocity a generalization of the analytical results to a three-dimensional geometry is necessary. For example, the self-similar variable is $\zeta = r^2 - z^2$, and the investigated equation (1) admits a solution that collapses near the surface of "hyperboloids." The other analytical relations are generalized in a corresponding way. The main observable effect should be anomalous broadening of the spectrum of the laser pulse along the propagation path.

It should be noted that when the growth rate of the stratification and fragmentation instabilities is increased, especially in the self-similar regime of collapse to the "hyperbolas," it becomes necessary to go beyond the framework of the equations with "quadratic dispersion." Clearly, taking the higher derivatives with respect to z into account will have a stabilizing effect. For example, it can be shown that when a third-derivative term $\propto i\partial^3\Psi/\partial z^3$ is added to the initial equation (1), the average of the maximum instability growth rate for a plane wave will remain as before ($\gamma \approx |u_0|^2$). However, the instability becomes convective, and the disturbances are carried out of the region with a velocity proportional to k_z^2 ($k_\perp \sim k_z$). From this, one can estimate that the characteristic scale of the nonuniformity $L \sim 1/k_z$ is stabilized at the level $L \sim 1/|u_0|$. Among the other mechanisms stabilizing the fragmentation instability, the most notable is multiphoton absorption [22].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 98-02-17205 and 99-02-16399).

REFERENCES

1. A. G. Litvak and V. I. Talanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **10**, 539 (1967).
2. V. E. Zakharov, Prikl. Mekh. Teor. Fiz. **9**, 86 (1968).

3. H. Yuen and B. Lake, *Nonlinear Dynamics of Deep-Water Gravity Waves* (Academic, New York, 1982; Mir, Moscow, 1987).
4. M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
5. A. G. Litvak, T. A. Petrova, A. M. Sergeev, and A. D. Yunakovskiĭ, Fiz. Plazmy **9**, 495 (1983) [Sov. J. Plasma Phys. **9**, 287 (1983)].
6. A. G. Litvak, A. M. Sergeev, and N. A. Shakhova, Pis'ma Zh. Tekh. Fiz. **5**, 862 (1979) [Sov. Tech. Phys. Lett. **5**, 33 (1979)].
7. J. R. Myra and C. S. Liu, Phys. Fluids **23**, 2258 (1980).
8. N. A. Zharova, A. G. Litvak, T. A. Petrova, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 12 (1986) [JETP Lett. **44**, 13 (1986)]; N. A. Zharova, A. G. Litvak, T. A. Petrova, *et al.*, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **29**, 1137 (1986).
9. J. E. Rothenberg, Opt. Lett. **17**, 583 (1992).
10. P. Chernev and V. Petrov, Opt. Lett. **17**, 172 (1992); Opt. Commun. **87**, 28 (1992).
11. G. G. Luther, A. C. Newell, and J. V. Moloney, Physica A (Amsterdam) **49**, 4085 (1994); G. G. Luther, A. C. Newell, J. V. Moloney, *et al.*, Opt. Lett. **19**, 789 (1994); **19**, 862 (1994).
12. X. D. Cao, G. P. Agrawal, and C. J. McKinstrie, Phys. Rev. A **49**, 4085 (1994).
13. G. Fibich, V. M. Malkin, and G. C. Papanicolou, Phys. Rev. A **52**, 4218 (1995).
14. L. Berge and J. J. Rasmussen, Phys. Plasmas **3**, 824 (1996); L. Berge, J. J. Rasmussen, E. A. Kuznetsov, *et al.*, J. Opt. Soc. Am. B **13**, 1879 (1996).
15. J. E. Ranka, R. W. Schirmer, and A. L. Gaeta, Phys. Rev. Lett. **77**, 3783 (1996).
16. L. Berge, E. A. Kuznetsov, and J. J. Rasmussen, Phys. Rev. E **53**, R1340 (1996).
17. S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981).
18. Ya. L. Bogomolov, E. N. Pelinovskiĭ, and A. D. Yunakovskiĭ, Preprint No. 275, IPF AN SSSR (Institute of Applied Physics, Academy of Sciences of USSR, Gorki, 1990).
19. V. E. Zakharov, in *Basic Plasma Physics*, Ed. by A. A. Galeev and R. Sudan (Énergoatomizdat, Moscow, 1984; North-Holland, Amsterdam, 1984), Vol. 2.
20. V. A. Aleshkevich, G. D. Kozheridze, and A. N. Matveev, Usp. Fiz. Nauk **161**, 81 (1991).
21. V. V. Bakhanov and V. I. Talanov, in *Collection of Scientific Works of Institute of Applied Physics, Russian Academy of Sciences "Near-Surface Ocean Layer. Physical Processes and Remote Probing"* (Nizhni Novgorod, 1999).
22. A. Brodeur and S. L. Chin, Phys. Rev. Lett. **80**, 4406 (1998).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 6, 2000, pp. 1279–1285.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 6, 2000, pp. 1476–1483.

MISCELLANEOUS

Minimal Surfaces and Fluctuations of Membranes with Nontrivial Topology

E. I. Kats[a, b, *] and M. I. Monastyrskiĭ[c]

[a] *Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117940 Russia*
[b] *Laue-Langevin Institute, F-38042 Grenoble, France*
[c] *Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia*
**e-mail: kats@cpd.landau.ac.ru*

Received August 1, 2000

Abstract—A new geometric approach to the description of phase transitions and fluctuations in membranes with nontrivial topology is proposed. The method is based on the possibility of representing real membranes and vesicles, defined in the space R^3, as minimal surfaces embedded in S^3. A change in the genus of the physical membrane corresponds to the formation of holes in the minimal surface. In the framework of mean field theory a model is constructed for a phase transition that can be characterized as the crystallization of holes in S^3. In real membranes this corresponds to a phase transition from a cubic phase to a sponge. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Membranes are thin, flexible layers of so-called amphiphilic molecules (i.e., having a polar head with a high affinity to a given solvent (usually water) and a hydrocarbon tail having an anomalously low affinity for the solvent). This structure makes membranes extremely unusual objects, much different from superficially similar systems such as the interface between liquids or solid shells. At interfaces between liquids an exchange of particles between the two bulk phases occurs, and this corresponds to a finite (as a rule, of the order of 10^2 erg/cm^2) surface tension. For a membrane found in equilibrium with a solution of the amphiphilic molecules of which it is made, the number of molecules of the membrane is fixed, which by definition means that the surface tension is zero.

Thus the energy of a membrane is determined not by surface tension but only by the elastic energy of its bending. For liquid membranes, which are the subject of the present paper, the characteristic bending modulus $\kappa \approx 10^{-14}$ erg (i.e., of the order of the temperature). This circumstance leads to fundamental differences between membranes and solid shells. The energy of the latter (like that of membranes) is determined mainly by the bending modulus. However, for solid shells this modulus is of the order of 10^{11} erg (i.e., 25 orders of magnitude (!) larger than for liquid membranes). Therefore, for a description of membranes one must use the formalism of statistical physics, whereas solid shells are described by the equations of mechanics.

In solutions, membranes can form closed surfaces called vesicles. The typical size of a vesicle is 1–10 μm. Vesicles and membranes have been the subject of many publications (see, e.g., monographs [1–3] and reviews [4, 5]). However, there are a number of questions on this topic that have not yet had much light shed on them in the literature. We are talking about vesicles and membranes with nontrivial topology (genus $g > 1$). There have been several publications devoted to investigating vesicles with the topology of a torus ($g = 1$) and Lawson surfaces ($g = 2$) [6–8]; some qualitative observations were made, and their shape fluctuations were discussed.

Usually the three simplest types of membrane shapes are considered: spherical, cylindrical, and planar (lamellar). There also exist more complex aggregates with nontrivial topology. The following circumstance will be important for further discussion: Spherical aggregates (vesicles) are naturally bounded in size, whereas cylindrical or lamellar vesicles can in principle have an infinite extent in one or two directions, respectively.

These supermolecular aggregates in turn can self-organize on large scales, forming phases with some type of orientational and/or translational order. For example, spherical vesicles can form a three-dimensional cubic lattice, or infinite cylinders can be packed into a two-dimensional hexagonal lattice. We note that in these examples we are talking about crystalline order occurring on large scales (in typical cases of the order of 10^{-4} cm), whereas on smaller scales the system is a liquid. Various kinds of structures made up of infinite layers are also possible. The simplest example is a one-dimensional lattice of nearly planar layers (such a structure, which is an analog of smectic liquid crystals, is called lamellar and is usually denoted as L_α. There are also phases having cubic spatial symmetry, the structural unit of which is an infinite periodic surface

formed by a bilayer and having a complex topology. This type of structure is denoted as Q_α. Schematically it can be described as a three-dimensional cubic structure of "handles" (the number of handles N_h is equal to the topological genus of the surface, and for the surface under consideration we have $N_h \longrightarrow \infty$ or $g \longrightarrow \infty$ in the thermodynamic limit). From this we already see the fundamental difference of the Q_α phase from the cubic structure of spherical vesicles considered above, which has a trivial topology.

In the case of melting of the periodic distribution of handles an isotropic "sponge" phase forms, conventionally denoted as L_3. The structural unit of the phase L_3 is an aperiodic infinite minimal surface (i.e., in other words, a surface with zero mean curvature at every point) and a finite density of handles or genus of the surface g, although the spatial distribution of g is characterized by only short-range order.

In this paper we present a new method of describing and classifying the possible types of structures formed by membranes. In Section 2 we describe the mathematical aspects of the classification method. In Section 3 that method is used to study the phase transitions involving a change of the topology and structure of the membranes. Section 4 is devoted to an analysis of the number of so-called conformal degrees of freedom (zero modes of the Helfrich energy [9] that describes the deformation of the membranes and vesicles), which exist only in membranes having a nontrivial topology ($g > 1$). In principle the thermodynamic and geometric characteristics of membranes can be studied by standard physical methods (e.g., the scattering of light or X-rays), which are determined by the spectrum of the Laplacian operator on the given surface (the latter can be measured, in principle, from the distribution of the concentration of marker atoms or temperature on the surface). We briefly discuss the connection between the physical and topological characteristics of membranes. In Section 5 we summarize the main results of this study.

2. RELATION OF WILLMORE SURFACES IN S^3 TO MINIMAL SURFACES IN R^3

In this section we discuss the mathematical facts that enable one to describe phase transitions in membranes.

We recall the basic definitions.

Let M^2 be a surface (closed, with an edge or noncompact) embedded in a three-dimensional Euclidean space R^3.

The surface M^2 is called a Willmore surface if it is an extremum of the Helfrich–Willmore functional [9, 10]:

$$F = \int_{M^2} H^2 dA, \qquad (1)$$

where H^2 is the square of the mean curvature and dA is an element of area. According to the Gauss–Bonnet theorem, the functional (1) is equivalent to the functional

$$\int_{M^2} (k_1^2 + k_2^2) dA,$$

where k_1 and k_2 are the principal curvatures of the surface, $H = (k_1 + k_2)/2$. Therefore the minima of the functional (1) describe the flattest surface of the given topology.

We shall make important use of an observation due to Weiner [11], which consists in the following.

Statement 1. Let $\tilde{M}^2$ be a minimal surface in S^3, and let γ be a stereographic projection $S^3 \longrightarrow R^3$. Then

$$\gamma(\tilde{M}^2) = M^2 \quad \text{and} \quad F(M^2) = \sigma(\tilde{M}^2), \qquad (2)$$

where $\sigma(\tilde{M}^2)$ is the area of the minimal surface.

This assertion is true both for closed surfaces and for surface with a boundary. The Willmore surfaces M^2 do not exhaust all the projections of minimal surfaces in S^3. For example, there exists an infinite family of Willmore tori (this family was constructed by Pinkal [12]) that do not correspond to minimal tori in S^3. However, these tori, although they are extrema of the functional (1), do not give local minima. This result is apparently valid for surfaces of arbitrary genus. The Willmore surfaces of arbitrary genus, obtained from the minimal surfaces in S^3, are constructed in [13, 14].

It is important to note that the Willmore surfaces embedded in R^3 satisfy the relation

$$\int_{M^2} H^2 dA \geq 4\pi n,$$

where n is the number of self-crossing points (of different types in projection [15]).

From this and the estimate of the area of the minimal surfaces embedded in S^3 it follows that all Willmer surface (for $g \geq 2$) have at least crossing points.[1]

3. DESCRIPTION OF PHASE TRANSITIONS INVOLVING A CHANGE OF MEMBRANE TOPOLOGY

In the previous section we showed that the Willmore surfaces that give a minimum of the Helfrich energy in a given topological class (i.e., under the condition that the genus of the surface g = const) is related to the min-

[1] Equivalence between the given class of Willmore surfaces and the minimal surfaces, which follows from Statement 1, leads to a far-ranging analogy with string theory. The minimal surfaces $M^2 \subset S^3$ are "world" sheets of a string with the Nambu–Goto action, defined in the sphere S^3. The surfaces M^2 play the role of instantons in a field theory of the Yang–Mills type, defined in the space S^3.

imal surfaces $M^2 \subset S^3$, which give a minimum of the area functional $\sigma(M^2)$. Therefore, instead of calculating the partition function for the Helfrich Hamiltonian (1), we can calculate the partition function determined by the Gibbs factor

$$\exp\left[-\frac{1}{T}\int \sigma(M^2)dA\right]$$

for the minimal surfaces in S^3. This approach is especially convenient for describing structures with a complex topology, e.g., cubic and sponge phases formed by membranes. Both of these types of structures are phases with a finite density of g. In the language of minimal surfaces S^3 in such structures correspond, according to (2), to minimal surfaces with a finite density of punched-out points (holes).

Before going on to a more detailed discussion of this correspondence, we need to say a few words about the physical structure of the cubic and sponge phases. The ordinary crystals studied in solid-state physics are infinitely repeating motifs of a single building block (the unit cell). The unit cell usually consists of a small number of atoms rigidly fixed in their equilibrium positions by the interatomic interaction potential.

Cubic phases of membranes are constructed differently. Cubic crystals of complex liquids form infinite threefold-periodic liquid surfaces. The unit cell of such a structure (from 100 to 1500 Å in size) contains a large number of molecules, which can freely diffuse along the membrane. The cubic structure is described in this case by a periodic distribution of the mass density $\rho(\mathbf{r})$ and related characteristics (including $g(\mathbf{r})$). The (sponge) phase L_3 is characterized by only short-range order in the distribution $\rho(\mathbf{r})$ or $g(\mathbf{r})$. In the language of the dual system (holes on the minimal surface in S^3), we can talk about a periodicity (in the case of the cubic phases L_3) or an amorphous (in the case of the sponge phase) distribution of the hole density $n(\mathbf{r})$.

If the energy E_h associated with the formation of a hole in the minimal surface is finite, then the average density of holes can be estimated as follows:

$$n \approx \frac{1}{\xi^2}\exp\left(-\frac{E_h}{T}\right), \tag{3}$$

where ξ is the average size of the unit cell formed by holes on the minimal surface [and is related to the average size of the physical unit cell of the cubic structure by the transformation (2)]. Formula (3) has the meaning of a dimensional estimate of the area per hole multiplied by the probability of formation of a hole. To find ξ we can make use of the following observation. We write the Helfrich-Willmore functional (1) out completely, i.e., with the Gaussian curvature taken into account K:

$$F = \frac{\kappa}{2}\int H^2 dA + \bar{\kappa}\int K dA. \tag{4}$$

Here we have introduced two elastic moduli κ and $\bar{\kappa}$, which determine the deformation of the membrane with respect to changes in the mean and Gaussian curvature, respectively. The formation of a sponge or cubic phase involves the spontaneous rise of the topological genus of the surface (i.e., the spontaneous creation of handles or, what is the same, passages between layers). In turn the spontaneous creation of handles occurs for $\bar{\kappa} > 0$, since, according to the Gauss–Bonnet theorem, for a closed surface the second term in Eq. (4) is equal to $4\pi\bar{\kappa}(1 - N_h)$, where N_h is the number of handles. This process of spontaneous creation of handles will inevitably lead to the appearance of an ever greater number of ever smaller handles. The process stops at a certain equilibrium scale ξ (which we are to find) such that the positive definite terms of higher order in the curvature become equal to the negative contribution from the Gaussian curvature. For finding ξ we must take into account in expansion (4) the terms proportional to H^{-4} and K^{-2}. Minimizing the energy thus obtained yields the equilibrium size ξ:

$$\xi \approx \sqrt{\frac{\kappa}{\bar{\kappa}}}a,$$

where we have used the natural estimate κa^2 (a is the molecular size) for the fourth-order elastic moduli.

It is important to note that the minimal surface in S^3 possesses surface tension (the energy determined by the functional (2) is proportional to the area, the coefficient of proportionality having the meaning of a surface tension).[2] Therefore the interaction between holes in such a manifold will necessarily be short-ranged (unlike physical membranes, where, because of the absence of surface tension in the Helfrich energy, this interaction is long-ranged) [16].

A natural estimate for the interaction energy is the temperature. If R is the characteristic interaction radius, then if the distance between holes is smaller than R, the interaction energy will be $U \approx T$, while for large distances the interaction can be neglected. What we have said can be formalized in the form the following expression for the free energy of a system of holes on a minimal surface in S^3:

$$F_h = Tn\ln\frac{n}{en_0} - Tn\ln(1 - n\xi^2). \tag{5}$$

Here $e \equiv \exp(1)$, and n_0 is the characteristic hole density in the dilute solution. The first term in (5) specifies the free energy of an ideal gas of holes, while the second

[2] Of course, the effective surface tension σ_{eff} of the minimal surfaces in S^3 depends on the physical parameters of the real membranes in R^3 in an extremely complicated way. To find the relation between them is a complicated problem whose solution is beyond the scope of this paper. We shall regard σ_{eff} as a phenomenological parameter that permits a description of phase transitions in physical membranes of nontrivial topology.

gives the increase in the free energy due to overlap of the deformed regions created by neighboring holes.

The free energy (5) can be used to describe the transition from a dilute hole phase to a dense liquid phase. We recall that the formation of a hole on a minimal surface in S^3 corresponds to an increase in the genus of the physical membrane (of the Willmore surface in R^3). Therefore the dilute hole phase can be used as a model for the lamellar phase L_α for membranes with infrequent "passages" between layers. The hole liquid phase can model the sponge phase L_3. For describing the ordered distributions of holes on the minimal surface (cubic phases formed by physical membranes) it is necessary to include in (5) terms which depend not only on the average hole density but also on the hole distribution on the minimal surface.

In the mean field approximation one can introduce the following single-particle distribution function for the displacements of the holes **u** relative to their equilibrium positions:

$$\rho_1(\mathbf{u}) = \frac{1}{Z}\exp\left[-\frac{1}{2T}\sum_q \varepsilon(q)|u(\mathbf{q})|^2\right], \tag{6}$$

where $\varepsilon(q)$ is the energy of deformation of the hole distribution, and

$$Z = \int Du(\mathbf{q})\exp\left[-\frac{1}{2T}\sum_q \varepsilon(q)|u(\mathbf{q})|^2\right]$$

is a normalization factor.

The energy $\varepsilon(q)$ can be expanded in a natural way as

$$\varepsilon(q) = \tau_{\text{eff}} + \sigma_{\text{eff}}\, q^2, \tag{7}$$

where the contribution τ_{eff} is determined from the energy of condensation of the dense phase, and the energy of deformation is related to the surface tension σ_{eff}.

From Eq. (7) we can find the mean-square displacement

$$\left\langle |u^2| \right\rangle = \frac{T}{\sqrt{\tau_{\text{eff}}}}\frac{1}{\sigma_{\text{eff}}}$$

and apply the classical Lindemann criterion for the melting of a crystal [17]

$$\frac{T}{\sqrt{\tau_{\text{eff}}}}\frac{1}{\sigma_{\text{eff}}} \approx 0.1\xi^2, \tag{8}$$

which determines the melting temperature of a crystal of holes on the minimal surface or (in accordance with the aforementioned equivalence of the minimal surfaces and physical membranes) the temperature of the cubic-sponge phase transition:

$$T_c \approx 0.1\xi^2\sigma_{\text{eff}}\sqrt{\tau_{\text{eff}}}\,. \tag{9}$$

We also note that this type of phase transition (the ordering of holes on the minimal surface or of topological handles on the Willmore surface) should not be accompanied by the a significant change in the average density and can therefore be a weak first-order transition (the so-called weak-crystallization phase transitions; see, e.g., [18]). In this case one expects that there will be appreciable fluctuational pre-transition effects. It is entirely possible that these effects are responsible for the flow-induced anomalous birefringence observed in sponge phases [19]. Indeed, in the isotropic sponge phase the birefringence $\delta\epsilon_{ik}$ can be induced by a gradient of the hydrodynamic velocity $\gamma_{ik} \equiv \nabla_i v_k$. In the simplest approximation we have the following linear relation:

$$\delta\epsilon_{ik} = \alpha\frac{\nabla_i v_k}{1 + i\omega\tau},$$

where α is a coefficient of proportionality, ω is the frequency, $\tau \propto \alpha/B$ is the relaxation time, and B is the characteristic elastic modulus of the cubic phase. Therefore, anomalous birefringence is expected at a weak first-order transition (when the modulus B is small and the coefficient α is large).

Let us recall once again that the characteristics of the minimal surface in S^3 the effective surface tension σ_{eff} and the effective hole condensation energy τ_{eff} are related to the physical characteristics of the real membrane in R^3 in an extremely complicated way. To establish the relation we would need a complete theory of phase transitions involving a change of the shape and topology of the membranes, and such a theory does not exist at present. The main content of this Section is the statement that a phenomenological theory can be easily formulated for minimal surfaces in S^3, but with unknown coefficients σ_{eff} and τ_{eff}, which are phenomenological parameters of the theory.

The restrictions on these coefficients follow directly from the fact that the formation of stable holes in $\tilde{M}^2 \subset S^3$ is possible. It is easy to understand, for example, that with a surface tension σ_{eff} = const (independent of the sizes and distribution of the holes) the formation of holes is altogether impossible. Indeed, the energy of formation of a single hole includes the energy cost of the line tension $\gamma_{\text{eff}}L$, where γ_{eff} is the phenomenological coefficient of linear tension (like the coefficients σ_{eff} and τ_{eff} introduced above, it depends on the physical parameters of the real membranes R^3 in a complicated way), and L is the perimeter of the hole. There is also an energy benefit coming from the energy of surface tension, i.e., $\sigma_{\text{eff}}A$ (A is the area of the hole). Thus for a circular hole of radius we have

$$E_h = 2\pi\gamma_{\text{eff}}r - \pi\sigma_{\text{eff}}r^2. \tag{10}$$

In the case of constant γ_{eff} and σ_{eff} the energy (10) has only one stable minimum, $r = 0$, which means the

absence of a hole. Furthermore, the energy (10) has a maximum at $r_c = \gamma_{\text{eff}}/\sigma_{\text{eff}}$, i.e., an energy barrier

$$\Delta E \approx \frac{\gamma_{\text{eff}}^2}{\sigma_{\text{eff}}}$$

such that for $r < r_c$ the hole should collapse and for $r > r_c$ it should grow to infinity. However, all of this discussion presupposes constant parameters σ_{eff} and γ_{eff}, i.e., the absence of a back effect of the holes on the properties of the surface (or, equivalently, an effect of the handles back on the properties of the physical membrane). Of course that is generally not the case, and the effective surface and linear tensions are far from being constant but depend in a complicated way both on the physical parameters of the real membrane in R^3 and on the external conditions and, in particular, on the sizes and distribution of the holes on S^3.

For example, in the presence of many holes one must take into account the entropy of their distribution and also the interaction between them. In addition, a membrane can have a certain orientational order. This order is also present in the minimal surface in S^3. The formation of a hole in the minimal surface can lead to deformation of the orientational field present at the surface and thereby bring about the aforementioned back effect of holes on σ_{eff} and γ_{eff} (the details of this "instanton" mechanism of stabilization of the holes is described in [20]). We note also that for any feedback mechanism (i.e., influence of holes on the properties of the surface) σ_{eff} and γ_{eff} depend on the material parameters of the different media: σ_{eff} is a characteristic of the solvent, and γ_{eff} is mainly determined by the properties of the amphiphilic molecules making up the membrane.

4. CONFORMAL MODES IN VESICLES AND MEMBRANES

In this section we discuss the properties of the conformal symmetry of the Helfrich–Willmore functional. The conformal modes of this functional are responsible for the anomalously strong fluctuations of the membranes. We begin with the intuitive, "physical" aspects of the conformal symmetry of the Helfrich–Willmore functional and then consider the mathematical aspects of this problem.

The main property of the Helfrich–Willmore functional is its invariance with respect to conformal transformations of the enveloping space R^3. The following assertion, which was proved back in the 1920s by Thomsen [10], is valid.

Statement 2. Let $M^2 \longrightarrow R^3$ be a smooth embedding (immersion) of the compact surface M^2. The Helfrich–Willmore functional [defined in Eq. (1)] is invariant with respect to the conformal group $G(R^3)$. The conformal group $G(R^3)$, according to the classical Liouville theorem, is generated by transformations of the motion: $O(3)$ rotations scale transformations and translations, and the "nontrivial" inversion transformation

$$\frac{\mathbf{r}'}{r'^2} \longrightarrow \frac{\mathbf{r}}{r^2} + \mathbf{a}, \tag{11}$$

where $\mathbf{a}$ is a constant vector.

The group $G(R^3)$ is isomorphic to the Lorentz group $O(4, 1)$.

Note. This result generalizes to the groups $G(R^n)$ $(n > 3)$: $G(R^n) \sim O(n + 1, 1)$.

It follows from Statement 2 that the number of geometrically nontrivial parameters of the group $G(R^3)$, i.e., those that alter the shape of the surface, is equal to four.

To estimate the total number of independent conformal modes of the Helfrich–Willmore functional, it is necessary to take into account the restrictions imposed by the physical conditions. Since the liquid inside a vesicle is incompressible, the condition

$$V = \text{const}, \tag{12}$$

where V is the volume of the vesicle, must hold for all admissible transformations. The amphiphilic molecules forming the vesicle are close-packed on the surface of the vesicle. Therefore, the area of the surface is also fixed:

$$A = \text{const}. \tag{13}$$

Using the arguments presented above, it is easy to comprehend the description of the conformal modes (and the conformal diffusion due to their existence).

4.1. General Case

We recall that the number of classes of conformally nonequivalent compact Riemann surfaces M_g^2 depends on the genus of the surface and is equal to

$$\begin{aligned} \dim_R M_g^2 &= 6g - 6 \quad \text{for} \quad g \geq 2, \\ \dim_R M_1^2 &= 2 \quad \text{for} \quad g = 1, \\ \dim_R M_0^2 &= 0 \quad \text{for} \quad g = 0. \end{aligned} \tag{14}$$

By comparing (14) with the Liouville theorem, we obtain the following result.

Statement 3. The number of (real) conformal modes k inducing conformal diffusion is equal to

$$\begin{aligned} k &= 0 \quad \text{for} \quad g = 0, \\ k &= 1 \quad \text{for} \quad g = 1, \\ k &= 6g - 8 \quad \text{for} \quad g \geq 2. \end{aligned}$$

4.2. Minimal Willmore Surfaces

For the Willmore surfaces obtained by projection from minimal surfaces embedded in S^3, the answer is considerably different. Conditions (12) and (13) are no

longer independent, and an estimate of the dimension of the space of moduli (parameters) m_g of the Willmore surface requires the use of a more detailed differential-geometric technique. In the case of compact minimal surfaces $\tilde{M}_g^2$ embedded in S^3, we can use the arguments given, e.g., in [21]. Let $\tilde{M}_g$ be the moduli space of the surfaces $\tilde{M}_g^2$. To estimate $\tilde{M}_g$ we proceed as follows.

Every surface M_g^2 admits a conformal metric and a holomorphic quadratic differential $Q(z)dz^2$. It is known that for any differential $Q(z)dz^2$there exists a conformal metric that defines the surface, including the minimal surface in S^3. The dimension of the space of such surfaces is $12g - 12$. This follows from an estimate of the dimension of the space of quadratic differentials $6g - 6$ and of the moduli space of the Riemann surface $6g - 6$. The number of different immersions of the surfaces M_g^2 is determined by the number of different spinor structures on M_g^2. More precisely, if φ_1 and φ_2 are different solutions of the Gauss–Weingarten equation for the surface M_g^2, then $\varphi_1^{-1}\varphi_2$ determines the same spinor structure on M_g^2. We choose on the surface $\tilde{M}_g^2$ a basis of homological cycles a_i, b_i ($i = 1, \ldots, g$), corresponding to the canonical conjugated lines of the surface. The spinor structure is determined by the monodromy transformation of the solutions φ_i ($i = 1, 2$):

$$\varphi_i \longrightarrow (-1)^{\alpha_{i+1}}\varphi_i,$$

$$\varphi_i \longrightarrow (-1)^{\beta_{i+1}}\varphi_i,$$

where $\alpha_i, \beta_i \in (0, 1)$.

The numbers $[\alpha, \beta]$, where $\alpha = (\alpha_1, \ldots, \alpha_g)$ and $\beta = (\beta_1, \ldots, \beta_g)$, are called the θ characteristics of the surface $\tilde{M}_g^2$ The θ characteristics $[\alpha, \beta]$ depend on the choices of bases (α, β) of the surface $\tilde{M}_g^2$. The parity of $[\alpha, \beta]$ is invariant:

$$[\alpha, \beta] = \sum_{i=1}^{g} \alpha_i\beta_i.$$

With the aid of parity one can classify the embeddings of the surfaces with respect to regular homotopies. Pinkal [22] showed that two surfaces, M_g^2 and N_g^2, can be transformed into each other by a smooth homotopy if and only if the parities of their θ characteristics are the same.

Using this result, it is easy to estimate the number of conditions that specify a unique embedding of the surface M_g^2. It is determined by the triviality of the monodromy of the solutions φ_i with respect to an arbitrary cycle c, since the monodromy is determined by the mapping of the fundamental group $\pi_1(M_g^2)$ of the surface $M_g^2: \pi_1(M_g^2) \longrightarrow SU(2)$. Since $\pi_1(M_g^2)$ is specified by generators with one commutation relation and the motion along each basis cycle is determined by the motion of the conformal group $CL(2C)$, we obtain conditions $6(2g - 1)$. Adding in the more general global transformation of S^3, we obtain a total of conditions $12g - 12$. Thus the number of equations and the number of parameters determining a compact minimal surface in are equal. This number is obviously finite for fixed S^3.[3] Known examples of minimal compact surfaces embedded in are the surfaces of Lawson [13] and of Karcher, Pinkal, and Sterling [14].

We note in concluding this Section that many physical characteristics of membranes and vesicles can be determined, e.g., from the distribution of the concentration of marker atoms or of the temperature on their surfaces. They are determined from the diffusion equation

$$\frac{\partial u}{\partial t} = D\Delta u,$$

where u is the physical field under study, D is the coefficient of diffusion, and Δ is the covariant Laplace-Beltrami operator, defined on the surface M^2. The stationary distribution of the field (e.g., the concentration or temperature) in turn is found from the spectrum of the operator Δ. The Laplace–Beltrami operator for a surface with a metric has the form

$$\Delta = \frac{1}{\sqrt{\tilde{q}}}\sum_k \partial_k\left(\sum_i g_{ik}\sqrt{\tilde{g}}\right)\partial_i,$$

where $\tilde{g}$ is the determinant of the metric tensor.

Determination of the shape and other geometric characteristics of a surface from the spectrum of a Laplacian operator defined on it for surfaces of constant Gaussian curvature is a classical mathematical problem. In particular, from the spectrum of the Laplacian one can recover (in the case of a compact surface) the lengths of the closed geodesics. Moreover, in that case one can say that the number of different surfaces of a fixed genus is finite.

In our case the Willmore surface is determined from the equation

$$\Delta H + H(H^2 - K) = 0$$

(H is the mean curvature and K is the Gaussian curvature) and need not be a surface of constant Gaussian curvature. However, as we have said, the number of different compact surfaces of genus $g \geq 2$ is clearly also finite and, consequently, the problem of recovering the main geometric and topological characteristics of Will-

[3] From our analyses it follows that toric minimal surfaces do not have conformal diffusion.

more surfaces from the spectrum of the Laplacian operator is also entirely sensible.

5. CONCLUSION

In this study we have proposed a new geometric approach to the description of phase transitions and fluctuations in membranes with nontrivial topology. The method is based on the possibility of representing real membranes and vesicles, defined in the space R^3, as minimal surfaces embedded in S^3. The change in the genus of the physical membrane corresponds to the formation of holes in the minimal surface. In the framework of mean field theory we have constructed a model for a phase transition that can be characterized as a crystallization of holes in S^3. In real membranes this corresponds to a phase transition from a cubic to a sponge phase. The formation of holes on the minimal surface in S^3 corresponds to an increase in the genus of the physical membrane (of the Willmore surface in R^3). Therefore the dilute hole phase can be used as a model for the lamellar phase L_α of membranes with infrequent "passages" between layers. The hole liquid phase models the sponge phase L_3. For a description of the ordered distributions of holes on the minimal surface (cubic phases formed by physical membranes) the free energy must include terms that depend not only on the mean density of holes but also on their distribution on the minimal surface. A periodic distribution of holes on the minimal surface corresponds in our approach to an infinite threefold-periodic Willmore surface having cubic symmetry. One of the possible realizations of such a cubic phase is the Schwarz surface [3, 10].

ACKNOWLEDGMENTS

This study was done as part of the "Statistical Physics" program of the Ministry of Science. One of the authors (E.I.K.) is grateful for grants from INTAS (grant no. 30234) and the Russian Foundation for Basic Research (project no. 00-02-17785), and the other author (M.I.M.) to the Russian Foundation for Basic Research (project nos. 99-01-00123, 99-01-17931) for financial support of this study.

REFERENCES

1. *Physics of Amphiphilic Layers*, Ed. by J. Meuner, D. Langevin, and N. Boccara (Springer-Verlag, Berlin, 1987); Springer Proceedings in Physics, Vol. 21.
2. S. A. Safran and N. A. Clark, *Physics of Complex and Supermolecular Fluids* (Wiley, New York, 1987).
3. D. Nelson, T. Pvian, and S. Weinberg, *Statistical Mechanics of Membranes and Surfaces* (World Scientific, New York, 1989).
4. A. M. Belloog, J. Piaif, and P. Bothorel, Adv. Colloid Interface Sci. **20**, 167 (1984).
5. G. Porte, J. Phys.: Condens. Matter **4**, 8649 (1992).
6. B. Fourcade, M. Mutz, and D. Bensimon, Phys. Rev. Lett. **68**, 2251 (1992).
7. F. Julicher, U. Seifert, and R. Lipowsky, Phys. Rev. Lett. **71**, 452 (1993).
8. E. I. Kats and V. V. Lebedev, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 57 (1995) [JETP Lett. **61**, 59 (1995)].
9. W. Helfrich, Z. Naturforsch. B **103**, 67 (1975).
10. T. J. Willmore, *Riemannian Geometry* (Clarendon Press, Oxford, 1993).
11. J. L. Weiner, Indiana Univ. Math. J. **27**, 19 (1978).
12. U. Pinkal, Invent. Math. **81**, 379 (1985).
13. H. B. Lawson, Ann. Math. **92**, 335 (1970).
14. H. Karcher, U. Pinkal, and I. Sterling, J. Diff. Geom. **28**, 169 (1988).
15. P. Li and S. T. Yau, Invent. Math. **69**, 269 (1982).
16. M. Kardar and R. Golestanian, Rev. Mod. Phys. **71**, 1233 (1999).
17. J. M. Ziman, *Principles of the Theory of Solids* (Cambridge Univ. Press, London, 1972, 2nd ed.; Mir, Moscow, 1966).
18. E. I. Kats, V. V. Lebedev, and A. R. Muratov, Phys. Rep. **228**, 1 (1993).
19. J. C. Lang and R. D. Morgan, J. Chem. Phys. **73**, 5849 (1980).
20. V. L. Golo, E. I. Kats, and G. Porte, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 575 (1996) [JETP Lett. **64**, 631 (1996)].
21. A. I. Bobenko, Usp. Mat. Nauk **46** (4), 3 (1991).
22. U. Pinkal, Topology **24**, 421 (1985).

Translation was provided by AIP