Cosmological effects of gravitational birefringence in a theory of gravitation with *C* and *P* violation

Yu. M. Loskutov*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 25 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1921–1929 (June 1998)

An explanation is given within a previously proposed theory of gravitation with broken conjugation and parity symmetry (*C* and *P* violation) for the recently detected effect of non-Faraday rotation of the polarization plane in the propagation of electromagnetic radiation over cosmological distances. © *1998 American Institute of Physics.* [S1063-7761(98)00106-1]

1. INTRODUCTION

An extensive body of experimental facts has been accumulated (see Ref. 1) on the rotation of the plane of polarization of the electromagnetic radiation emitted by distant radio galaxies (data have been obtained on 160 galaxies). From this body of data it clearly follows that the polarization plane undergoes not only Faraday rotation, but also an additional rotation. In contrast to Faraday rotation, this other rotation does not depend on the emission wavelength. The absolute value of the additional rotation angle β lies in the range from 0 to π (if we disregard the possible addition of multiples of π to the value). Recently Nodland and Ralston,[2] after mathematically processing these data,[1] came to a quite unexpected conclusion. According to them, the experimental points are most densely grouped about an ansatz for β specified by the expression

$$\beta=\frac{1}{2\Lambda}\, r\, \cos\, \gamma+\beta_0, \tag{1}$$

where r is the distance from the source to the observer, Λ is the measurement scale of r, and γ is the angle between the propagation direction of the ray and some unit vector **s**. In the case of a sampling of 71 galaxies with a redshift $z>0.3$, the best agreement between formula (1) and the observations is obtained if **s** is characterized by declination and right ascension equal to $0°\pm 20°$ and $21\,\mathrm{h}\pm 2\,\mathrm{h}$, respectively.

A comparison of formula (1) with the observations shows that most of the galaxies with $\cos\gamma>0$ have $\beta\sim\pi/2$ and that the number of galaxies decreases as β deviates to either side of $\pi/2$. Similarly, the galaxies with $\cos\gamma<0$ are scattered in β about the value $\beta\sim-\pi/2$. The value of β_0 for all galaxies turns out to be near $1°-2°$ (with a typical value of $1°$).

In essence, the ansatz (1) introduces two types of non-Faraday rotation: right-handed rotation for one group of galaxies and left-handed rotation for the other (therefore, the uncertainty in the determination of β for the two groups of galaxies will be equal to $+n\pi$ and $-n\pi$). If we agree that the ansatz (1) corresponds more accurately to the observations than any other ansatz, then this would be evidence of a violation of the homogeneity and isotropy of space, at least in our Metagalaxy—too presumptuous a conclusion not to expect objections. And such objections soon appeared in a paper by Carrol and Field (Ref. 3).[1)] In contrast to Nodland and Ralston,[2] Carrol and Field[3] assumed that the non-Faraday rotation of the polarization plane of the electromagnetic radiation from all galaxies is of the same type and does not appear as right and left-handed rotation in different groups of galaxies. Such an hypothesis at once excludes the possibility of introducing an asymmetrical ansatz of the type (1) and thus the possibility of a violation of the homogeneity and isotropy of space. According to this hypothesis, the rotation angle β (if we disregard its uncertainty, which is now equal only to $+n\pi$) for the majority of galaxies (with $z>0.3$) is scattered about $\pi/2$, and the values of $\beta=\pi/2+|\Delta\beta|$ correspond to a smaller number of galaxies, the larger is the value of $|\Delta\beta|$ (this is reflected in histogram *1* in Ref. 3). Carrol and Field[3] assert that the experimental points β (with consideration of the hypothesis of a single type of rotations) are grouped about the ansatz $\beta\simeq\pi/2$ more densely than around the ansatz (1). It is impossible to verify this due to the absence of details. However, if we judge from Fig. 2 in Ref. 3 and Fig. 1d in Ref. 2, which plot distributions of the numbers of galaxies (with $z>0.3$) in β as functions of $r\cos\gamma$, it is hard to come to this conclusion—on a purely visual level they seem to be saying the opposite (of course, a visual estimate is not always valid, since it depends, for example, on the scales). At the same time, the question of around which ansatz for the β values are the experimental points more densely grouped is decisive here. An additional independent mathematical treatment of the observation data[1] is clearly necessary in order to decide if it is at all possible to answer this question on the basis of the available data. At the current moment, an unequivocal decision in favor of Ref. 2 or Ref. 3 would seem premature, although the hypothesis in Ref. 3 is preferable from the physical standpoint.

It should be recalled that the experimental non-Faraday values of β were obtained from 160 galaxies, while the main conclusions in Refs. 2 and 3 were based on data from the 71 galaxies with $z>0.3$. What may be expected in the interpretations in Refs. 2 and 3 from the remaining 89 galaxies with $z<0.3$?

If the ansatz (1) is a reflection of an actual law of nature,

then it should also apply to nearby (with $z<0.3$) galaxies. This, however, is not observed—it is enough to look at Fig. 1c from Ref. 2, and even Nodland and Ralston[2] noted the poor correlation of (1) with the observed distribution over β for the nearby galaxies. Such a result places the ansatz (1) in serious doubt as a universal law.

If the ansatz $\beta\simeq\pi/2$ proposed in Ref. 3 reflects a law of nature, then the experimental β points of the nearby galaxies should reproduce histogram *1* in Ref. 3. This, however, they do not do. Therefore, it is still impossible to make a solid judgment about the universality of the ansatz introduced in Ref. 3.

If an additional analysis of the data from all 160 radio galaxies reveals that the distribution of the observed β values on the interval $[0,\pi]$ correlates very weakly with some ansatz or does not correlate with any, then this would indicate that the β values are distributed not on the interval $[0,\pi]$, but on the interval $[0,n\pi]$, where n is large. From a general standpoint, such a picture seems more likely. However, the deciding word must wait for an additional mathematical treatment of all the observation data.

An explanation for non-Faraday (gravitational) birefringence based on the theory of gravitation with broken conjugation and parity symmetry (C and P violation) proposed in Ref. 4 is presented below. Additional consequences of the treatments in Refs. 2 and 3 are analyzed. Some of these consequences can be interpreted as evidence in favor of the explanation of Ref. 3, although it is not theoretically possible to categorically reject the explanation of Ref. 2. The possibility of an uncorrelated distribution over the gravitational rotation angles β is also discussed.

2. COSMOLOGICAL EFFECTS OF GRAVITATIONAL BIREFRINGENCE AS A CONSEQUENCE OF THE THEORY OF GRAVITATION WITH *C* and *P* VIOLATION AND THE POSSIBILITY OF PRESERVING THE MODEL OF A HOMOGENEOUS, ISOTROPIC UNIVERSE

In Ref. 4 I constructed a theory of gravitation satisfying the principle of the universality of the law of the violation of conjugation (C) and parity (P) symmetry ($CP=$const) in nature. By virtue of a number of requirements, the structure of the Lagrangian of spinor particles and the structure of the Lagrangian of polarized photons in a gravitational field turn out to be uniquely defined. In the quasiclassical limit both Lagrangians lead to identical dynamical equations of polarized particles in a gravitational field. In particular, in the weak-field approximation the equation describing the propagation of circularly polarized photons in space with a static Riemannian metric $g_{\mu\nu}(\mathbf{x})$ has the form:[2)]

$$E^2g^{00}+\mathbf{p}^2g^{11}+2\zeta hC(\mathbf{p}\nabla)h_1^{(1)}=0. \tag{2}$$

Here $E\equiv\hbar\omega$ is the photon energy, $\mathbf{p}$ is its canonical momentum, $\zeta=\pm1$ corresponds to right circular (upper sign) and left circular (lower sign) polarization of the photon, $h_\nu^{(a)}$ are the associated tetrads ($h_{(a)}^{\nu}=g^{\nu\mu}h_{(a)\mu}$, $h_{(b)\nu}=h_\nu^{(a)}\eta_{ab}$, etc., where η_{ab} is the Minkowski metric tensor), and C is a dimensionless real constant of gravitational interactions violating C and P symmetry (if the units of measure are restored and the structure of $h_\nu^{(a)}$ is taken into account, the constant C will always appear in the equation together with the factor G/c^2).

The presence of the last term in Eq. (2) at once leads to birefringence effects in a gravitational field: an integrated effect of gravitational rotation of the polarization plane and a local effect of visible angular splitting of the rays with right and left circular polarization.

On the path from the source to the observer the polarization plane is rotated according to Eq. (2) by the angle (in radians)

$$\beta=C[h_1^{(1)}(\mathbf{r}_2)-h_1^{(1)}(\mathbf{r}_1)], \tag{3}$$

where $h_1^{(1)}(\mathbf{r}_2)$ and $h_1^{(1)}(\mathbf{r}_1)$ are the values of the tetrads at the signal reception and emission points, respectively. It also follows from Eq. (2) that for the same frequency ω pulses of photons with right ($\zeta=1$) and left circular ($\zeta=-1$) polarization will be different from the observation point. Within the approximation chosen of a weak gravitational field

$$\mathbf{p}_\zeta\simeq\mathbf{p}_0-\zeta\hbar C\nabla h_1^{(1)}(\mathbf{r}_2). \tag{4}$$

This means that in right- and left-circularly polarized rays the source will be seen at different angles from the observation point. The visible angular splitting will lie in the plane $(\mathbf{p}_0,\nabla h_1^{(1)}(\mathbf{r}_2))$ and will be equal to

$$\alpha\simeq\frac{C\lambda}{\pi}\,|[\boldsymbol{\kappa}^0\nabla h_1^{(1)}(\mathbf{r}_2)]|, \tag{5}$$

where λ is the emission wavelength and $\boldsymbol{\kappa}^0$ is a unit vector pointing from the source to the observer.

Imagine that a plane-polarized radio signal is emitted from a satellite traveling along a given elliptical trajectory about the Sun and is received on the Earth. Then

$$h_1^{(1)}(\mathbf{r})\simeq1-\frac{M_\odot}{r}-\frac{M_\oplus}{|\mathbf{r}-\mathbf{r}_e|}, \tag{6}$$

where $\mathbf{r}_e$ is the radius vector of the center of the Earth relative to the center of the Sun. Consequently,

$$\alpha\simeq\frac{C\lambda M_\oplus}{\pi R^2}\sin\chi,\quad \beta\simeq C\left[\frac{M_\odot}{r_1}-\frac{M_\odot}{|\mathbf{r}_e-\mathbf{R}|}-\frac{M_\oplus}{R}\right]. \tag{7}$$

Here $\mathbf{R}$ is the radius vector of the observer relative to the center of the Earth, and χ is the angle between $\mathbf{p}_0$ and $\mathbf{R}$. It has been taken into account in the expression for α that at the Earth's surface the gradient of the Sun's potential can be neglected in comparison with the gradient of the Earth's potential and the Earth's potential in the vicinity of the satellite has been dropped in the expression for β.

If the conditions of the experiment are such that before the start of the experiment the polarization plane of the source on the satellite is set at a certain angle (e.g., 90°) to the ecliptic plane at the antenna (and the orientation of the antenna is subsequently held fixed by gyroscopes), then by determining the orientation of the polarization plane at the observer, it is possible to determine the rotation angle β. If, however, before the start of the experiment the antenna is oriented so that the polarization plane at the observer would form a certain angle (e.g., 90°) with the ecliptic plane and

the variation of this angle would subsequently be measured as the ray approaches the Sun, then the quantity

$$\Delta\beta \simeq CM_\odot \left\{ \left[\frac{1}{r_1(t)} - \frac{1}{r_1(0)} \right] - \left[\frac{1}{r_e(t)} - \frac{1}{r_e(0)} \right] + \left[\frac{\mathbf{R}(t)\mathbf{r}_e(t)}{r_e^3(t)} - \frac{\mathbf{R}(0)\mathbf{r}_e(t)}{r_e^3(0)} \right] \right\} \tag{8}$$

should be compared with the observations.

Gravitational birefringence effects must be compared with the analogous effects of a magnetoplasma origin. Toward this end, we make use of the dispersion relation linking the frequency of a photon with its wave vector in a magnetized plasma. Under conditions in which the cyclotron frequency $\tilde{\Omega} \equiv e\mathscr{H}/mc$ (here $\mathscr{H}$ is the intensity of the external magnetic field, and e and m are the charge and mass of the electron) is much smaller than the photon frequency ω, the dispersion relation for right- and left-circularly polarized photons can be represented in the form

$$E^2 = \frac{4\pi N e^2 \hbar^2}{m} + c^2\mathbf{p}^2 + \zeta \frac{4\pi N e^3 \hbar}{m^2\omega^2} (\mathbf{p}\mathscr{H}), \tag{9}$$

where N is the electron density. Hence, for the visible angular splitting α_1 of the right-and left-circularly polarized rays at the observation point and the Faraday rotation angle β_1 after the ray has traveled from the source to the observer we obtain, respectively,

$$\alpha_1 \simeq \frac{e^3 \lambda^3 N(\mathbf{r}_2)\mathscr{H}(\mathbf{r}_2)}{2\pi^2 m^2 c^4} \sin\chi_1,$$

$$\beta_1 \simeq \frac{e^3\lambda^2}{2\pi m^2 c^4} \int_{\mathbf{r}_1}^{\mathbf{r}_2} N(s)\mathscr{H}(s)ds. \tag{10}$$

Here χ_1 is the angle between $\mathbf{p}$ and $\mathscr{H}$, $N(s)$ is the electron density on the path of the ray, and $\mathscr{H}(s)$ is the magnetic field component tangent to the ray.

A comparison of Eqs. (7) and (10) shows that, in contrast to Faraday rotation, the gravitational rotation of the polarization plane does not depend on the emission wavelength, which simplifies the problem of its identification. In addition, to measure β or $\Delta\beta$ it is not necessary that the ray approach the Sun, since the values of β and $\Delta\beta$ depend only on the gravitational potentials at the signal emission and reception points. Conversely, in order to mark off the Faraday rotation, it is better to make the measurements where the contribution of magnetoplasma effects is very small. As for the visible splitting angle, the functional structure of α and α_1 is also very different, and the splitting planes do not coincide. Therefore, distinguishing them experimentally does not present any great difficulty.

Rotation of the polarization plane upon the passage of rays near the Sun was studied experimentally in Refs. 5 and 6. The Pioneer-6, 7, and 8 spacecraft served as platforms for sources of plane-polarized 2292-MHz radio emission. The orientation of the antennas was rigidly maintained by gyroscopes, although it cannot be gathered from these papers whether the initial orientation of the polarization plane was fixed at the satellite (first variant) or at the observer (second variant). The authors did not observe any noticeable change in the orientation of the polarization plane of the rays at distances from the Sun greater than six solar radii. On the one hand, this means that the magnetoplasma effects are negligibly small at such distances. On the other hand, if the experiment went according to the first variant, it means that β, is small, as in the case[3)] for $C \ll 10^8$ (in the second variant this restriction is lifted). As the ray further approaches the Sun, quite rapid growth of the rotation angle is observed, but it is associated with Faraday rotation rather than gravitational rotation. Consequently, in the second variant $C < 10^{12}$.

It would be appropriate here to mention the experiments in Ref. 7, which have a bearing on the possibility of estimating the value of C. Harwit *et al.*[7] state (in Sec. 2 of their paper) that quasar 3C273 is seen at different angles in right- and left-circularly polarized rays. The upper value of the angle of visible splitting amounts to ''roughly 0.002 arc sec in the direction perpendicular to the radius vector to the Sun'' and ''roughly 0.0035 arc sec in the direction of the radius vector'' (the angular resolution was 10^{-4} arc sec). As can be seen, the plane of the observed splitting forms an angle of about 30° with the source-Sun-observer plane. Symmetry arguments cast grave doubts on the explanation of the splitting effect as a result of the influence of the Sun and the circumsolar plasma (on the days of the measurements 3C273 was 5.0° and 7.8° from the Sun). Harwit *et al.*[7] clearly observed a local effect of the splitting given by the values of α and α_1 from formulas (7) and (10).

A comparison of formulas (7) and (10) with the observations requires knowledge of the diurnal time of the measurements (which is lacking in Ref. 7), since both the orientation of the splitting plane and the values of the angles χ and χ_1 depend on it. However, the order of magnitude of the corresponding quantities can be estimated. To reconcile the observed data with α, it is necessary to assume a value of C of the order of 10^{10}. However, if these data are attributed to an interaction of the radiation with electrons of the radio interferometer itself in the Earth's magnetic field, then the number density N of such electrons in formula (10) should be of the order of 10^8 cm^{-3}, which is entirely possible. Thus, to arrive at any final conclusions would require a refinement of the experimental data, or better yet, a specially designed experiment. Taking into account the local character of the angular splitting effect, we could set up the experiment with any sufficiently distant object emitting in the radio-frequency range (in the case of very short waves the effect becomes small) and at any time of day.

Let us turn our attention now to the effect of cosmological gravitational rotation of the polarization plane. We, first of all, estimate the contribution of the gravitational potentials of the galaxies to the rotation angle β. By virtue of the great distance between the radiation sources and the observer and between different sources, this contribution is governed by the values of the potential of our Galaxy (the small potentials of the Earth and the Sun can be neglected) at the observation point and the potential of the respective galaxy in the region of formation of the radio emission. The contribution of these potentials to β is given by Eq. (3):

$$\beta_i \simeq C[\Phi_i(\mathbf{r}_1) - \Phi_0(\mathbf{r}_2)], \tag{11}$$

where $\Phi_i(\mathbf{r}_1)$ is the gravitational potential of the ith radio galaxy in the region of formation of the emission, and $\Phi_0(\mathbf{r}_2)$ is the gravitational potential of our Galaxy at the observation point ($\sim 6\times 10^{-7}$). As can be seen, this expression does not contain any asymmetry. Therefore, within the model of Ref. 2 all the β_i should be assigned to the second term in Eq. (1), i.e., to $\beta_0 \sim 1°$. The polarized component of the radio emission of galaxies is usually linked with magnetic braking (synchrotron) radiation near neutron stars. In the region of formation of the emission the potentials of neutron stars can be estimated as $(1-5)\times 10^{-2}$. For all the β_i as given by formula (11) to agree with the explanation given in Ref. 2, it is necessary to set $C \sim 1$.

If Carrol and Field[3] are closer to the truth, then $C \sim 10^2$; in this case the spread of the β_i should mirror the spread of the regions of formation of the emissions near neutron stars. If, on the other hand, there are no correlations in the distribution of the β_i, then $C \gg 10^2$.

For an asymmetrical term to appear in the expression for the gravitational rotation angle β, as in Eq. (1), it is necessary to assume the existence of a gravitational field with a preferred directionality in the Metagalaxy, as can be seen from Eq. (3). Since matter is distributed uniformly inside the Metagalaxy, according to the observations, internal sources cannot create a directed field. Consequently, we are left to assume the existence in the Universe of large-scale inhomogeneities outside the Metagalaxy. The homogeneity and isotropy of the Universe will not be lost if we average not only over the small-scale, but also over the large-scale inhomogeneities.

For example, the following two forms of inhomogeneities are permissible. 1) During several billion years a certain excess (positive or negative) of matter density above its mean formed in a layer of large radius (similar to a disk) are far from the Metagalaxy. This excess created a gravitational field in the Metagalaxy almost orthogonal to the layer (if the edges of the inhomogeneity disk are far from the Metagalaxy). 2) A density excess could have formed in a sphere of any radius far from the Metagalaxy. A gravitational field then appears in the Metagalaxy from this centrosymmetric source. If only the first or second field shows up in the Metagalaxy, this means that other inhomogeneities are located at significantly greater distances from it.

In the first case the inhomogeneity can be approximated by a layer with an effective width a, an effective radius b much greater than the radius of the Metagalaxy, and a constant excess density $\tilde{\rho}$. Then,

$$h_1^{(1)} \simeq 1 + 2\pi\tilde{\rho} a z, \tag{12}$$

where z is the distance from the middle of the layer to some chosen point in the Metagalaxy. We note in passing that the excess (positive or negative) mass of the layer $\tilde{M} = \pi\tilde{\rho} a b^2$ can be much less than the mass of the Metagalaxy. According to Eq. (3), the angle β of gravitational rotation of the plane of polarization on the path from a radio galaxy to an observer is equal to

$$\beta = -2\pi C\tilde{\rho} a r \cos\gamma. \tag{13}$$

Here r is the distance from the galaxy to the observer and γ is the angle between the propagation direction of the ray and the direction to the inhomogeneity layer. This is in complete agreement with the empirically introduced expression (1). A comparison of expression (13) with the numerical results of Ref. 2 gives

$$2\pi C a|\tilde{\rho}| \simeq 6.7 \text{ g/cm}^2. \tag{14}$$

In the second case

$$h_1^{(1)} \simeq 1 - \frac{\tilde{M}}{r}, \tag{15}$$

where $\tilde{M}$ is the excess (positive or negative) mass concentrated in a sphere with an effective radius $r_0 \ll r$ and r is the distance from the center of the sphere to the chosen point in the Metagalaxy. This leads to

$$\beta = C\tilde{M}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \simeq -\frac{C\tilde{M}}{r_2^2} r \cos\gamma. \tag{16}$$

Here r_2 is the distance from the center of the sphere to the observer, r is the distance from the galaxy to the observer, and γ is the angle between the propagation direction of the ray and the direction to the center of the inhomogeneity sphere. A comparison of (16) with the numerical results of Ref. 2 gives

$$\frac{C|\tilde{M}|}{r_2^2} \simeq 6.7 \text{ g/cm}^2. \tag{17}$$

Above we showed that $C \sim 1$ in the variant argued in Ref. 2. However, for such values of C the conditions on the parameters of the large-scale inhomogeneities following from Eqs. (14) and (17) become completely unacceptable. For example, the gravitational potentials of the inhomogeneities in the Metagalaxy in such a case amount to at least several tens of units, and the relative velocity of the Metagalaxy and an inhomogeneity due to their interaction turns out to be close to the speed of light at the present time. If we require that this velocity not exceed, say, 100 km/s, then $2\pi a|\tilde{\rho}|$ and $|\tilde{M}|/r_2^2$ should not exceed a value of the order of 10^{-3}. For $C \sim 1$ this would lead only to a very weak asymmetry in the distribution of the β values.

Thus, from the theoretical standpoint outlined above two versions remain tenable (if the plane-polarized component of galactic radio emission is formed largely in the vicinity of neutron stars). Either almost all of the β values are distributed on the interval $[0,\pi]$ about $\beta \sim \pi/2$ (and thus $C \sim 10^2$), or all the β values are distributed on an interval $[0,n\pi]$ with $n \gg 1$, which maps them onto a practically random distribution on the interval $[0,\pi]$ (and thus $C \gg 10^2$). Granted, the possibility of the asymmetrical contribution (13), (16) cannot be excluded theoretically in either case if large-scale inhomogeneities are present.

3. CONCLUSION

In summation, it can be said that the theory of gravitation with C and P violation proposed in Ref. 4 (the corre-

sponding interaction could be called ''graviweak'') affords an explanation for the non-Faraday (gravitational) rotation of the plane of polarization of electromagnetic radiation propagating over cosmological distances. The reason for gravitational rotation of the polarization plane is a postulated gravitational interaction, which depends on the spin states of the propagating photons—see Ref. 4 and Eq. (2).

If the plane-polarized component of the emission of radio galaxies is formed mainly in the vicinity of neutron stars, i.e., in regions with relatively large (of the order of 0.01–0.1) gravitational potentials, then it is not possible to reconcile the proposed theory with the ansatz (1). It is possible to obtain satisfactory agreement between the theory and the ansatz (1) only in the case in which practically the entire polarized component of the radiation is formed in regions with small (less than 10^{-4}) gravitational potentials. In this case it is also necessary to postulate the existence in the Metagalaxy of a gravitational field with a preferred directionality, which can arise due to, for example, large-scale cosmological inhomogeneities—see Eqs. (11), (13), and (16).

If it is confirmed that the experimental β values are more densely grouped about the ansatz $\beta \sim \pi/2$, as Carrol and Field[3] maintain, then in the case when the plane-polarized component of the galactic radio emissions is formed mainly in the vicinity of neutron stars this will require $C \sim 10^2$ from the proposed theory—see Eq. (11). If the polarized component of the radiation is formed in regions with lower gravitational potentials, then the value of C should be higher. There is then no need to hypothesize the existence of a gravitational field with a preferred orientation in the Metagalaxy. In this case the result in Ref. 3 is clearly preferable to the result in Ref. 2. However, such arguments in favor of one interpretation cannot be taken as decisive—only experiments can decide the issue.

Since it is still too early to speak of 100% confidence in the results of either Ref. 2 or Ref. 3, there is reason, in my opinion, to propose a third kind of experimental test: assembling the data from all 160 galaxies and then verifying whether the spread in β on the interval $[0,\pi]$ or the intervals $[-\pi,0]$ and $[0,\pi]$ is nearly random, i.e., either very weakly correlated with some ansatz or completely uncorrelated with any ansatz. If the picture turns out to be such, then from the standpoint of the proposed theory this will mean that $C \gg 10^2$.

In any case there is an urgent need for experiments to verify the gravitational birefringence effects predicted in Ref. 4 [see Eqs. (7) and (11)]. The identification of such gravitational effects would be assisted by their qualitative differentiation from magnetoplasma effects [compare Eqs. (7) and (10)]. The setting up of such experiments would be of tremendous significance, since they would provide answers to fundamental questions.

*)E-mail: loskutov@moldyn.phys.msu.su

1)I am grateful to the reviewer for directing my attention to Ref. 3, about which I was previously unaware. It totally changed the tone of the present work.

2)The system of units $c = G = 1$ has been chosen.

3)In Ref. 4 C was estimated on the basis of fragmentary excerpts (see the clarification in Ref. 4) from Ref. 7, which was lost at that time and found only recently. In these excerpts, the results of Ref. 7 turned out, as has now become clear, to be mixed together with the work in Ref. 6, to which considerable space was allotted in Ref. 7. For this reason, the estimate of C obtained in Ref. 4 turns out to be greatly overestimated (Ref. 4 draws attention to the necessity of refining it). Nevertheless, all of the fundamental and qualitative conclusions in Ref. 4 drawn from the value of C remain in force (only the quantitative estimates vary).

[1] F. F. Gardner and J. B. Whiteoak, Nature **197**, 1162 (1963); Annu. Rev. Astron. Astrophys. **4**, 245 (1966); G. Burbidge and A. H. Growne, Astrophys. J., Suppl. **40**, 583 (1979); H. Spinrad *et al.*, Publ. Astron. Soc. Pac. **97**, 932 (1985); J. N. Clarke, P. P. Kronberg, and M. Simard-Normandin, Mon. Not. R. Astron. Soc. **190**, 205 (1980); H. Alfvén and K. Herlofson, Phys. Rev. **78**, 616 (1950).
[2] B. Nodland and J. P. Ralston, Phys. Rev. Lett. **78**, 3043 (1997).
[3] S. M. Carrol and G. B. Field, Phys. Rev. Lett. **79**, 2394 (1997).
[4] Yu. M. Loskutov, Zh. Éksp. Teor. Fiz. **107**, 283 (1995) [JETP **80**, 150 (1995)]; Teor. Mat. Fiz. **105**, 324 (1995).
[5] G. S. Levy, T. Sato, B. L. Seidel, *et al.*, Science **166**, 596 (1969).
[6] C. T. Stelzried, G. S. Levy, T. Sato, *et al.*, Sol. Phys. **14**, 440 (1970).
[7] M. Harwit, R. V. E. Lovelace, B. Dennison, *et al.*, Nature **249**, 230 (1974).

Translated by Paul F. Schippnick
Edited by P. Shelnitz

The diffraction cone for exclusive vector meson production in deep inelastic scattering

J. Nemchik

Dipartimento di Fisica Teorica, Universitá di Torino and INFN, Sezione di Torino, I-10125 Torino, Italy; Institute of Experimental Physics, Slovak Academy of Sciences, 04353 Košice, Slovak Republic

N. N. Nikolaev

IKP(Theorie), KFA Jülich, 5170 Jülich, Germany; L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia

E. Predazzi

Dipartimento di Fisica Teorica, Universitá di Torino and INFN, Sezione di Torino, I-10125 Torino, Italy; Institute of Experimental Physics, Slovak Academy of Sciences, 04353 Košice, Slovak Republic

B. G. Zakharov

L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia

V. R. Zoller

Institute for Theoretical and Experimental Physics, 117218 Moscow, Russia

(Submitted 25 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1930–1962 (June 1998)

We develop the color dipole gBFKL phenomenology of a diffraction cone for photo- and electroproduction γ^*N-VN of heavy vector mesons (charmonium & bottomonium) at HERA and in fixed target experiments. We predict a substantial shrinkage of the diffraction cone from the CERN/FNAL to the HERA range of c.m.s. energy W. The Q^2-controlled sensitivity to the color dipole size (scanning phenomenon) is shown to lead to a decrease of the diffraction slope with Q^2 (which is supported by the available experimental data). We predict an approximate flavor independence of the diffraction slope in the scaling variable $Q^2+m_V^2$. For diffractive production of the radially excited $2S$ states (Ψ',Υ'), a counterintuitive inequality of diffraction slopes $B(2S)\lesssim B(1S)$ is predicted, which defies the common wisdom that diffraction slopes are larger for reactions with larger size particles. © *1998 American Institute of Physics.* [S1063-7761(98)00206-6]

1. INTRODUCTION

Diffractive real and virtual photoproduction of ground state $V(1S)$ and radially excited $V'(2S)$ vector mesons,

$$\gamma^*p\to V(1S)p,\quad V'(2S)p, \tag{1}$$

at high c.m.s. energy $W=\sqrt{s}$ is an ideal testing ground of ideas on QCD pomeron exchange. New experimental data on vector meson production coming from HERA and fixed target experiments provide unique insight into how the pomeron exchange evolves from the nonperturbative to semiperturbative and perturbative regimes with increasing virtuality of the photon Q^2 and/or increasing mass m_V of the produced vector meson, and have prompted intense theoretical discussions.[1–11]

The usual approach to the perturbative QCD (pQCD) pomeron is based on the BFKL equation[12,13] for the leading-logs (LLs) evolution of the gluon distribution, formulated in the scaling approximation of fixed QCD coupling $\alpha_S=$const and of infinite gluon correlation (propagation) radius R_c (massless gluons); it sums the ladder diagrams with reggeized t-channel gluon exchanges. More recently, a novel s-channel approach to the LLs BFKL equation has been developed[14,15] in terms of the color dipole cross section $\sigma(\xi,r)$ (hereafter r is the color dipole moment, $x_{\rm eff}=(m_V^2+Q^2)/(W^2+Q^2)$ and $\xi=\log(1/x_{\rm eff})$ is the rapidity variable). The color dipole approach, to be referred to as the running gBFKL approach, is favored because it incorporates consistently the two crucial properties of QCD: 1) asymptotic freedom (AF), i.e., the running QCD coupling $\alpha_S(r)$ and, 2) the finite propagation radius R_c of perturbative gluons.

AF and the running $\alpha_S(r)$ are an indispensable feature of the modern theory of deep inelastic scattering (DIS); without running $\alpha_S(r)$ it is impossible to match the leading $\log Q^2$ (LLQ^2) limit of the gBFKL equation with the conventional GLDAP equation[16] in the overlapping applicability region of the moderately small x,[14,17] see also Refs. 13, 18 and 19. The finite R_c is of great importance since the nonperturbative fluctuations in the QCD vacuum restrict the phase space for the soft perturbative (real and virtual) gluons, and there is strong evidence for finite R_c from the lattice QCD studies (for the review see Ref. 20 and references therein) and hadronic interactions.[21,22] In the infrared, one also needs to freeze $\alpha_S(r)$ in order not to run into the Landau pole, $\alpha_S(r)\leqslant\alpha_S^{(fr)}$. Of course, if in our running gBFKL

equation[14,15] one puts $R_c=\infty$ and $\alpha_S=$const, then the original scaling BFKL equation[12,13] is recovered.[17,23]

Being formulated in terms of real (rather than reggeized) gluon exchanges, the color dipole running gBFKL equation[14,15] readily incorporates the running $\alpha_S(r)$. The effect of finite R_c can be included by modifying the gluon propagator in the infrared, for instance, by introducing the effective gluon mass $\mu_g\approx 1/R_c$. Remarkably, in the LLs approximation, a finite R_c is consistent with QCD gauge invariance. The freezing of $\alpha_S(r)$ and the gluon correlation radius R_c are the nonperturbative parameters which describe the transition from the soft, infrared region, to the perturbative, hard region.

Purely perturbative pomeron exchange does not exhaust the scattering amplitude, and in the practical phenomenology of deep inelastic scattering one must add a certain soft nonperturbative exchange. It is significant that the color dipole picture and color dipole factorization for the proton structure function and for exclusive diffractive amplitudes do not require the validity of pQCD, and are also viable for soft pomeron exchange. Soft pomeron exchange is important only for sufficiently large color dipoles, $r>R_c$, and can readily be extracted from the experimental data on hadronic cross sections[24] and diffractive leptoproduction of light vector mesons.[9] On the other hand, pQCD (or hard) pomeron exchange can be related to the perturbative gluon structure function of the proton.[14,25]

Diffractive production of $V(1S)$ mesons is particularly interesting because of the so-called scanning phenomenon, whereby the production amplitude probes the color dipole cross section at the dipole size $r\sim r_S$, where

$$r_S\approx\frac{A}{\sqrt{Q^2+m_V^2}} \tag{2}$$

is the scanning radius.[3–5]

This scanning property follows from the color dipole factorization for production amplitudes and the shrinkage of the transverse size of the virtual photon with Q^2, and holds beyond the pQCD domain.[24] By varying Q^2, one can study the transition from large nonperturbative and semiperturbative dipole size r_S to the perturbative region of very short $r_S\ll R_c$ in a very well-controlled fashion.[9,10] Furthermore, the scanning radius r_S defines the transverse size of the γ^*V transition vertex, which contributes to the total interaction radius and to the diffraction slope $B(\gamma^*\to V)$.

Motivated by the remarkable phenomenological success of such a unified color dipole picture of hard and soft pomerons in application to the proton structure function[26,27] and vector meson production,[5,10] in this communication we develop the color dipole description of the forward diffraction cone $B(\gamma^*\to V)$ in exclusive diffractive DIS. We use our previous results for the energy dependence of the forward cone in color dipole scattering,[28] obtained from the solution of the running gBFKL equation for the diffraction slope.[29] Here the crucial point is that breaking of the scale invariance by AF, i.e., by running $\alpha_S(r)$, and finite R_c, dramatically alters the very nature of the BFKL pomeron from a fixed cut in the scaling approximation to a series of moving poles for the running gBFKL pomeron[29,27] (for early semiclassical analysis see also Ref. 13). As a result, in the running gBFKL approach one predicts a substantial Regge shrinkage of the diffraction slope in the vector meson production, which can be tested at HERA. In this paper we present for the first time a detailed analysis of the Q^2 dependence and of the Regge growth of the diffraction slope for the production of charmonium and bottomonium states.

Diffractive production of radially excited $V'(2S)$ mesons provides additional insight into the dipole size dependence of the diffraction slope. Because of the node in the radial wave function of the $V'(2S)$ states, there is strong cancellation among contributions to the production amplitude from dipoles r larger than or smaller than the node position r_n (the node effect[1,3,30]). The resulting strong suppression of diffractive production of $V'(2S)$ vs. $V(1S)$ has been confirmed experimentally in J/Ψ and Ψ' photoproduction at HERA and in fixed target experiments.[31–33] It also has interesting manifestations in the differential cross sections, which we discuss in the present paper for the first time.

Because the radius of the $V'(2S)$ state is larger than the radius of the ground state $V(1S)$, for the diffraction slopes one would naively expect the inequality $B(\gamma^*\to\Psi')>B(\gamma^*\to J/\Psi)$. However, in this paper we demonstrate that the node effect in conjunction with the color dipole factorization predicts the counterintuitive inequality $B(\gamma^*\to\Psi')\lesssim B(\gamma^*\to J/\Psi)$, which can be tested at HERA. Because the node effect is sensitive to the form of the dipole cross section and its variation with energy, we predict the specific energy dependence of the $V'(2S)/V(1S)$ production ratios, which can also be tested at HERA.

The presentation is organized as follows. The subject of the introductory Sec. 2 is color dipole factorization and the determination of the pQCD factorization scales for diffractive production. The running gBFKL formalism for the calculation of the color dipole scattering matrix and of the diffraction slope is presented in Sec. 3. The decomposition of the diffraction slope into perturbative and nonperturbative components, and the physics that controls the W^2, flavor, and Q^2 dependence of the diffraction slope is expounded in Sec. 4. In Sec. 5 we discuss in more detail the properties of soft pomeron exchange in the color dipole representation. In Sec. 6 we present the salient features of soft and hard exchange for VN total cross sections. Predictions from the running gBFKL dynamics for forward and t-integrated vector meson production cross sections are reported in Sec. 7. We find good agreement with the low-energy data and data from the HERA collider experiments. The subject of Sec. 8 is predictions for the forward cone in diffractive production of $V(1S)$ states, with special emphasis on flavor symmetry. Section 9 is concerned with the node effect in forward production of $V(2S)$ states. A summary and conclusions are presented in Sec. 10.

2. INTRODUCTION TO COLOR DIPOLE FACTORIZATION AND pQCD FACTORIZATION SCALES FOR DIFFRACTIVE AMPLITUDES

The Fock state expansion for the light-cone meson starts with the $q\bar{q}$ state, a snapshot of the relativistic meson as a $q\bar{q}$

color dipole. The probability amplitude of finding the $q\bar{q}$ color dipole of size $\vec{r}$ is precisely the $q\bar{q}$ wave function, $\Psi(\vec{r},z)$, where z is the fraction of the meson's light-cone momentum carried by a quark (the Sudakov light-cone variable). The interaction of the relativistic color dipole of the dipole moment $\vec{r}$ with the target nucleon is quantified by the energy-dependent color dipole cross section $\sigma(\xi,r)$. The effect of higher Fock states $q\bar{q}g\ldots$ is very important at high energy ν. In the LLs and/or LL1/x approximation it can be reabsorbed into the energy (rapidity) dependence of $\sigma(\xi,r)$, which is described by the running gBFKL equation.[14,15] The dipole cross section is flavor independent, and provides a unified description of various diffractive processes.

In the limit of high photon energy ν, the $q\bar{q}$-nucleon scattering matrix $\widetilde{\mathscr{M}}$ becomes diagonal in the mixed $(\vec{r},z)$ representation. This $(\vec{r},z)$ diagonalization derives from the large longitudinal coherence length,

$$l_{\rm coh}\sim\frac{2\nu}{Q^2+m_V^2}, \tag{3}$$

and holds if $l_{\rm coh}\gg R_p$, where R_p is a size of the target proton. Because the coherence length is a purely kinematic scale,[34] $(\vec{r},z)$ diagonalization does not require the applicability of pQCD and must hold also for soft pomeron exchange, i.e., even if the dipole size $\vec{r}$ is large. A necessary condition is that the longitudinal scale $l_{\rm soft}$ for soft pomeron exchange be small, $l_{\rm soft}\ll l_{\rm coh}$, which is the case, for instance, in the dual parton string model[35] or various models of exchange by nonperturbative gluons.[11,36,37] For the phenomenological success of a unified color dipole picture of vector meson production, see Refs. 9 and 10.

Taking advantage of the $(\vec{r},z)$ diagonalization of the scattering matrix $\widetilde{\mathscr{M}}$, the amplitude for real (virtual) photoproduction of vector mesons with momentum transfer $\vec{q}$ can be represented in the color dipole factorized form

$$\mathscr{M}(\gamma^*\to V,\xi,Q^2,\vec{q})=\langle V|\widetilde{\mathscr{M}}|\gamma^*\rangle$$
$$=\int_0^1 dz\int d^2r\ \Psi_V^*(\vec{r},z)\times\mathscr{M}(\xi,r,z,\vec{q})\Psi_{\gamma^*}(\vec{r},z). \tag{4}$$

Our normalization is such that $(d\sigma/dt)_{t=0}=|\mathscr{M}|^2/16\pi$. In Eq. (4), $\Psi_{\gamma^*}(\vec{r},z)$ and $\Psi_V(\vec{r},z)$ represent the probability amplitudes to find a color dipole of size $\vec{r}$ in the photon and quarkonium (vector meson), respectively (for the sake of brevity we suppress spin indices), and $\mathscr{M}(\xi,r,z,\vec{q})$ is the amplitude for elastic scattering of the color dipole on the target nucleon. The color dipole distribution in (virtual) photons was derived in Refs. 24 and 14.

The color dipole cross section $\sigma(\xi,r)$ depends only on the dipole size r, and not on the q-$\bar{q}$ relative momentum z. Because $z\approx 1/2$ in nonrelativistic heavy quarkonium, at small $\vec{q}$ in the diffraction cone one can safely neglect the z-dependence of $\widetilde{\mathscr{M}}$ and set $z=1/2$. Hereafter we will suppress the argument z, and either ξ, $x_{\rm eff}$, or x will be used whenever convenient (x is the Bjorken variable for inclusive DIS; straightforward analysis of the relevant Sudakov variables yields $x_{\rm eff}\approx 2x$).

We focus on calculating the imaginary part of the scattering amplitude, for which there is a simple representation in terms of the gluon density matrix (see below). The small real part can easily be reconstructed from analyticity considerations[38]:

$$\mathrm{Re}\ \mathscr{M}(\xi,r,\vec{q})=\frac{\pi}{2}\frac{\partial}{\partial\xi}\,\mathrm{Im}\ \mathscr{M}(\xi,r,\vec{q}). \tag{5}$$

We forgo discussion of $\mathrm{Re}\,\mathscr{M}$, which is consistently included in all numerical results.

The details of calculation of the diffractive amplitude have been presented elsewhere.[5,10] For the $Vq\bar{q}$ vertex function we assume the Lorentz structure $\Gamma\widetilde{\Psi}\gamma_\mu\Psi V_\mu$. For the s-channel helicity conservation at small $\vec{q}$, transversely (T) polarized photons produce transversely polarized vector mesons and longitudinally (L) polarized photons (to be more precise, scalar photons) produce longitudinally polarized vector mesons. One finds

$$\mathrm{Im}\ \mathscr{M}_T(x_{\rm eff},Q^2,\vec{q})=\frac{N_cC_V\sqrt{4\pi\alpha_{em}}}{(2\pi)^2}\int d^2r\,\sigma(x_{\rm eff},r,\vec{q})\times\int_0^1\frac{dz}{z(1-z)}\{m_q^2K_0(\varepsilon r)\phi(\vec{r},z)-[z^2+(1-z)^2]\varepsilon K_1(\varepsilon r)\partial_r\phi(\vec{r},z)\}, \tag{6}$$

$$\mathrm{Im}\ \mathscr{M}_L(x_{\rm eff},Q^2,\vec{q})=\frac{N_cC_V\sqrt{4\pi\alpha_{em}}}{(2\pi)^2}\frac{2\sqrt{Q^2}}{m_V}\times\int d^2r\,\sigma(x_{\rm eff},r,\vec{q})\int_0^1 dz\ K_0(\varepsilon r)\times\{[m_q^2+z(1-z)m_V^2]\phi(\vec{r},z)-\partial_r^2\phi(\vec{r},z)\}, \tag{7}$$

where

$$\varepsilon^2=m_q^2+z(1-z)Q^2, \tag{8}$$

α_{em} is the fine structure constant, and $N_c=3$ is the number of colors. $C_V=1/\sqrt{2}$, $1/3\sqrt{2}$, 1/3, 2/3, 1/3 are the charge-isospin factors for ρ^0, ω^0, ϕ^0, J/Ψ, and Υ production, respectively, and $K_{0,1}(x)$ are modified Bessel functions. A detailed discussion and parametrization of the light-cone radial wave function $\phi(r,z)$ of the $q\bar{q}$ Fock state of the vector meson are given in Ref. 10. For heavy quarkonia one can safely identify the current and constituent quarks. The terms $\propto K_0(\varepsilon r)\phi(\vec{r},z)$ and $\propto\varepsilon K_1(\varepsilon r)\partial_r\phi(\vec{r},z)$ for T polarization, and $K_0(\varepsilon r)\partial_r^2\phi(\vec{r},z)$ for L polarization, correspond to the helicity conserving and helicity-flip transitions in the $\gamma^*\to q\bar{q}$ and $V\to q\bar{q}$ vertices, respectively. In nonrelativistic heavy quarkonia, the helicity flip transitions are the relativistic corrections, which become important only at large Q^2. Equation (7) corrects a slight mistake in the relativistic correction to the amplitude for production of longitudinally po-

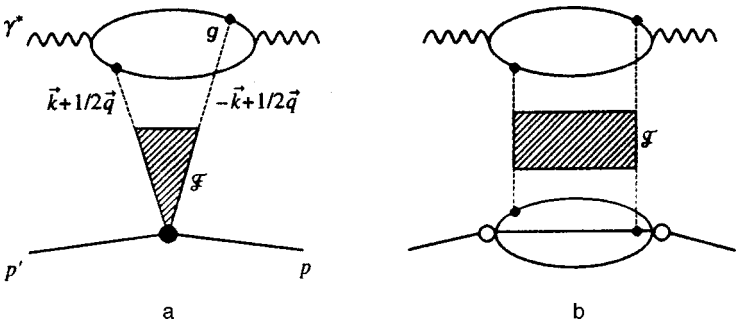


FIG. 1. The perturbative QCD diagrams for vector meson perturbation.

larized photons made in Ref. 5. The numerical results of Ref. 5 for the J/Ψ are only marginally different from those to be reported in this paper.

The representation for $\sigma(x,r,\vec{q})$ in terms of the gluon density matrix (see Fig. 1) is

$$\sigma(x,r,\vec{q})=\frac{4\pi}{3}\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)\left[J_0\left(\frac{1}{2}qr\right)-J_0(kr)\right]\mathscr{F}\left(x,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}\right),\tag{9}$$

where $J_0(x)$ is the usual Bessel function. AF dictates that at the gluon-color dipole vertex, the QCD running coupling must be taken at the largest relevant virtuality, $\kappa^2=(\max\{\vec{k}^2,C^2r^{-2}\})$, where $C\approx 1.5$ (Ref. 24), ensuring numerically similar results of calculations in both the mixed (r,z) and momentum representations. The gluon density matrix $\mathscr{F}(x,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2)$ is proportional to the imaginary part of the nonforward gluon–nucleon scattering amplitude; at $\vec{q}=0$ it equals the unintegrated gluon structure function of the nucleon $\mathscr{F}(x,\vec{k},-\vec{k})=\partial G(x,k^2)/\partial\log k^2$. Equation (9) generalizes to the nonforward case $\vec{q}\neq 0$ the formula[25,14] for the dipole cross section

$$\sigma(x_{\rm eff},r,\vec{q}=0)=\sigma(x_{\rm eff},r)=\frac{\pi^2r^2}{3}\times\int\alpha_S(\kappa^2)\,\frac{dk^2}{k^2}\,\frac{4[1-J_0(kr)]}{(kr)^2}\,\frac{\partial G(x_{\rm eff},k^2)}{\partial\log k^2}.\tag{10}$$

Because the function $f(y)=4[1-J_0(y)]/y^2$ can be qualitatively approximated by the Heaviside step function, $f(y)\approx\theta(A_\sigma-y)$, where $A_\sigma\approx 10$ (Ref. 39) for small $r\ll R_c$ one readily finds

$$\sigma(x,r)=\frac{\pi^2}{3}\,r^2\alpha_s(r)G(x,q_\sigma^2),\tag{11}$$

where the gluon structure function enters at the pQCD factorization scale $q_\sigma^2\sim A_\sigma/r^2$.[14,25,39] For large dipoles, $r\gtrsim R_c$, one can neglect $J_0(kr)$ in the integrand, and the dipole cross section saturates,

$$\sigma(x_{\rm eff},r\gtrsim R_c)=\frac{4\pi^2}{3}\int\alpha_S(k^2)\,\frac{dk^2}{k^4}\,\frac{\partial G(x_{\rm eff},k^2)}{\partial\log k^2}.\tag{12}$$

Next, notice that the integrands in Eqs. (6) and (7) are smooth at small r and vanish exponentially at $r>1/\varepsilon$ due to $K_{0,1}(\varepsilon r)$. Because of the behavior $\sigma(x,r)\propto r^2$ in (11), the amplitudes (6) and (7) are dominated by the contribution from the dipole size $r\approx r_S$ given by Eq. (2)—the scanning phenomenon.[3–5] The scanning property is best quantified in terms of the weight functions $W_{T,L}(Q^2,r^2)$ defined by

$$\mathscr{M}_T(x_{\rm eff},Q^2,\vec{q})=\frac{C_V}{(m_V^2+Q^2)^2}\times\int\frac{dr^2}{r^2}\,\frac{\sigma(x_{\rm eff},r,\vec{q})}{r^2}\,W_T(Q^2,r^2),\tag{13}$$

$$\mathscr{M}_L(x_{\rm eff},Q^2,\vec{q})=\frac{C_V}{(m_V^2+Q^2)^2}\,\frac{2\sqrt{Q^2}}{m_V}\times\int\frac{dr^2}{r^2}\,\frac{\sigma(x_{\rm eff},r,\vec{q})}{r^2}\,W_L(Q^2,r^2),\tag{14}$$

where in a somewhat abbreviated form ($i=T,L$; for the exact integrands see Eqs. (6) and (7))

$$W_i(Q^2,r^2)=\frac{\pi}{C_V}\,r^4(m_V^2+Q^2)^2\times\int_0^1 dz\,\Psi^*_{V_i}(r,z)\Psi_{\gamma_i^*}(r,z).\tag{15}$$

For $1S$ mesons, to a good approximation the $W_{T,L}(Q^2,r^2)$ thus defined are sharply peaked functions of the natural variable $y=\log[r^2(Q^2+m_V^2)]$. The height and width of the peak in the y-distribution vary only weakly with Q^2 and the flavor, and the peak position defines the scanning radius $r_S\approx A_{T,L}/\sqrt{Q^2+m_V^2}$. Consequently, the leading twist terms in the expansion over the relevant short-distance parameter $r_S^2\propto 1/(Q^2+m_V^2)$ are of the form (here we assume $\vec{q}=0$)

$$\mathrm{Im}\,\mathscr{M}_T\propto\frac{1}{Q^2+m_V^2}\,\sigma(x_{\rm eff},r_S)\propto\frac{1}{(Q^2+m_V^2)^2}\,G(x_{\rm eff},q_T^2),\tag{16}$$

$$\mathrm{Im}\,\mathscr{M}_L\approx\frac{\sqrt{Q^2}}{m_V}\,\mathscr{M}_T\propto\frac{\sqrt{Q^2}}{m_V}\,\frac{1}{(Q^2+m_V^2)^2}\,G(x_{\rm eff},q_L^2).\tag{17}$$

By virtue of (11), here the pQCD scale $q_{T,L}^2=\tau_{L,T}(Q^2+m_V^2)$, where the scale parameter $\tau_{T,L}$ can be estimated as

$$\tau_{T,L}\approx\frac{A_\sigma}{A_{T,L}^2}. \tag{18}$$

For the more direct evaluation of the pQCD factorization scales $q_{T,L}^2$, it is convenient to substitute Eq. (10) into (13) and (14), which then take a form reminiscent of the k-factorization formulas for $F_2(x,Q^2)$[24,40]:

$$\mathrm{Im}\,\mathscr{M}_T(x_{\rm eff},Q^2,\vec{q}=0)=\frac{C_V\alpha_S(Q^2+m_V^2)}{(m_V^2+Q^2)^2}\times\int\frac{dk^2}{k^2}\frac{\partial G(x_{\rm eff},k^2)}{\partial\,\log k^2}\,\Theta_T(Q^2,k^2), \tag{19}$$

$$\mathrm{Im}\,\mathscr{M}_L(x_{\rm eff},Q^2,\vec{q}=0)=\frac{C_V\alpha_S(Q^2+m_V^2)}{(m_V^2+Q^2)^2}\,\frac{2\sqrt{Q^2}}{m_V}\times\int\frac{dk^2}{k^2}\frac{\partial G(x_{\rm eff},k^2))}{\partial\,\log k^2}\,\Theta_L(Q^2,k^2), \tag{20}$$

where

$$\Theta_{T,L}(Q^2,k^2)=\frac{\pi^2}{3}\int\frac{dr^2}{r^2}\frac{\alpha_S(\kappa^2)}{\alpha_S(Q^2+m_V^2)}\frac{4[1-J_0(kr)]}{(kr)^2}\times W_{T,L}(Q^2,r^2). \tag{21}$$

Because of properties of $f(y)$ and the sharp peaking of $W_{T,L}(Q^2,r^2)$ at $r\approx r_S$, the weight functions $\Theta_{T,L}(Q^2,k^2)$ are similar to the step function,

$$\Theta_{T,L}(Q^2,k^2)\propto\theta(q_{T,L}^2-k^2), \tag{22}$$

and

$$\int\frac{dk^2}{k^2}\frac{\partial G(x_{\rm eff},k^2))}{\partial\,\log k^2}\,\Theta_i(Q^2,k^2)=G(x_{\rm eff},q_i^2)\int\frac{dk^2}{k^2}\,\Theta_i(Q^2,k^2))=G(x_{\rm eff},q_i^2)I_i(Q^2), \tag{23}$$

where the factors

$$I_{T,L}(Q^2)=\frac{\pi^2}{3}\int\frac{dr^2}{r^2}\frac{\alpha_S(\kappa^2)}{\alpha_S(Q^2+m_V^2)}\,W_{T,L}(Q^2,r^2) \tag{24}$$

exhibit only a marginal dependence on Q^2.

For small Q^2 the scale parameters $A_{T,L}$ are close to the nonrelativistic estimate $A\sim 6$, which follows from $r_S=3/\varepsilon$ with the nonrelativistic choice $z=1/2$. In general $A_{T,L}\geqslant 6$, and they increase slowly with Q^2 (Ref. 5); for heavy quarkonia $A_{T,L}(\Upsilon)\simeq 6$ at $Q^2\leqslant 100\ \mathrm{GeV}^2$, $A_{T,L}(J/\Psi)\sim 6$ at $Q^2=0$, and $A_{T,L}(J/\Psi)\simeq 7$ at $Q^2=100\ \mathrm{GeV}^2$, which shows that the relativistic corrections in the charmonium and bottomonium electroproduction are small. The corollary of the large scanning radius r_S and large values of $A_{T,L}$ is a very small scale factor $\tau_{T,L}$ in the pQCD factorization scale[5]: $\tau_{T,L}(J/\Psi)\approx 0.20$, $\tau_L(\rho^0)\approx 0.15$, and $\tau_T(\rho^0)\approx 0.07-0.10$ for $Q^2\sim 10-100\ \mathrm{GeV}^2$, which are substantially smaller than $\tau\approx 0.25$ suggested in Ref. 2 and $\tau\approx 1$ suggested in Ref. 6. Consequently, the moderate values of Q^2 attainable at HERA do, at best, correspond to the nonperturbative and semiperturbative values of $q_{T,L}^2$, the soft contribution to the vector meson production must be substantial, and one must be careful with the interpretation of the vector meson production data in terms of the gluon structure function. The point that at $Q^2\lesssim m_{J/\Psi}^2$ the scanning radius r_S is comparable to the radius of the J/Ψ is overlooked in Ref. 2, and the formulas of Ref. 2 for the J/Ψ production amplitudes in terms of the J/Ψ wave function at the origin are too crude. Strictly speaking, Eqs. (19), (20), and (23) were derived for hard pQCD exchange when $r_S\lesssim R_c$ and/or for perturbatively large $q_{T,L}^2$. However, because the color dipole factorization is true beyond pQCD, one can extend Eq. (10) to the soft pomeron and regard this relationship as an operational definition of the nonperturbative gluon distribution in the proton. To the same extent, Eqs. (19), (20), and (23) can serve as a unique basis for extracting the whole gluon distribution, perturbative plus nonperturbative, at small x from the experimental data on diffractive vector meson electroproduction at HERA.

The dominance of the longitudinal amplitude (17) at $Q^2\gtrsim m_V^2$ follows, as a matter of fact, from electromagnetic gauge invariance, and as such it is true in any reasonable model of vector meson production, the familiar vector dominance model (VDM) included. The Q^2 dependence of $\mathscr{M}_{T,L}$ differs drastically from the VDM prediction

$$\mathscr{M}_T(\mathrm{VDM})\propto\frac{1}{m_V^2+Q^2}\,\sigma_{\rm tot}(\rho N);$$

instead of $\sigma_{\rm tot}(\rho N)$ in (16) one has $\sigma(x_{\rm eff},r_S)\propto r_S^2\propto 1/(Q^2+m_V^2)$.

3. DIFFRACTION CONE IN THE COLOR DIPOLE gBFKL APPROACH

In the familiar impact-parameter representation for amplitude of elastic scattering of the color dipole

$$\mathrm{Im}\,\mathscr{M}(\xi,r,\vec{q})=2\int d^2b\,\exp(-i\vec{q}\vec{b})\Gamma(\xi,\vec{r},\vec{b}), \tag{25}$$

the diffraction slope $B=-2(d\,\log\mathrm{Im}\,\mathscr{M}/dq^2)_{q=0}$ equals

$$B(\xi,r)=\frac{1}{2}\,\langle\vec{b}^2\rangle=\frac{\lambda(\xi,r)}{\sigma(\xi,r)}, \tag{26}$$

where

$$\lambda(\xi,r)=\int d^2b\vec{b}^2\Gamma(\xi,\vec{r}\vec{b}). \tag{27}$$

Then, the generalization of the color dipole factorization formula (4) to the diffraction slope of the reaction $\gamma^*p\to Vp$ reads

$$B(\gamma^*\to V,\xi,Q^2)\,\mathrm{Im}\,\mathscr{M}(\gamma^*\to V,\xi,Q^2,\vec{q}=0)=\int_0^1 dz\int d^2r\lambda(\xi,r)\Psi_V^*(r,z)\Psi_{\gamma^*}(r,z). \tag{28}$$

We sketch here the running gBFKL equation[28] for $\lambda(\xi,r)$. The running gBFKL equation for the energy dependence of the color dipole cross section reads[14,15]

$$\frac{\partial\sigma(\xi,r)}{\partial\xi}=\mathscr{K}\otimes\sigma(\xi,r)=\frac{3}{8\pi^3}\int d^2\rho_1\mu_G^2 \times\left|g_S(R_1)K_1(\mu_G\rho_1)\,\frac{\vec{\rho}_1}{\rho_1}-g_S(R_2)K_1(\mu_G\rho_2)\,\frac{\vec{\rho}_2}{\rho_2}\right|^2 \times[\sigma(\xi,\rho_1)+\sigma(\xi,\rho_2)-\sigma(\xi,r)]. \tag{29}$$

Here the kernel $\mathscr{K}$ is related to the squared wave function of the color-singlet $q\bar{q}g$ state with the Weizsäcker–Williams (WW) soft gluon, in which $\vec{r}$ is the $\bar{q}$-q separation and $\vec{\rho}_{1,2}$ are the q-g and $\bar{q}$-g separations in the two-dimensional impact parameter plane. The quantity

$$\vec{\mathscr{E}}(\vec{\rho})=\mu_G g_S(\rho)K_1(\mu_G\rho)\,\frac{\vec{\rho}}{\rho}=-g_S(\rho)\nabla_\rho K_0(\mu_G\rho),$$

where $K_\nu(x)$ is the modified Bessel function, describes the Yukawa screened transverse chromoelectric field of the relativistic quark, and

$$\mu_G^2\left|g_S(R_1)K_1(\mu_G\rho_1)\,\frac{\vec{\rho}_1}{\rho_1}-g_S(R_2)K_1(\mu_G\rho_2)\,\frac{\vec{\rho}_2}{\rho_2}\right|^2 =|\vec{\mathscr{E}}(\vec{\rho}_1)-\vec{\mathscr{E}}(\vec{\rho}_2)|^2 \tag{30}$$

describes the flux (the modulus of the Poynting vector) of WW gluons in the $q\bar{q}g$ state. The asymptotic freedom of QCD uniquely prescribes that the chromoelectric field be computed with the running QCD charge $g_S(r)=\sqrt{4\pi\alpha_S(r)}$ taken at the shortest relevant distance, $R_i=\min\{r,\rho_i\}$ in the $q\bar{q}g$ system. The particular combination of the three color dipole cross sections,

$$\Delta\sigma(\rho_1,\rho_2,r)=\frac{9}{8}\,[\sigma(\xi,\rho_1)+\sigma(\xi,\rho_2)-\sigma(\xi,r)], \tag{31}$$

which emerges on the right-hand side of the gBFKL equation, is precisely the change in the color dipole cross section in the presence of the WW gluon[14] in the $q\bar{q}g$ state.

At short distances, $r,\rho_{1,2}\ll R_c=1/\mu_G$, the kernel $\mathscr{K}$ does not depend on the infrared cutoff R_c. The Yukawa cutoff of the long-range chromoelectric field, which was used in Eqs. (29) and (30), is the simplest phenomenological option. In the LL($1/x$) approximation, this cutoff is consistent with gauge invariance. If one sacrifices AF, putting $g_S=\text{const}$, and lifts the infrared cutoff by letting $R_c\to\infty$, one recovers the scale-invariant kernel $\mathscr{K}$. Both the finite R_c and running α_S break the scale invariance; a detailed discussion of the consequences is found in Refs. 15, 17, 27, and 29. The principal phenomenon is that because of the lack of strong $\log r^2$ ordering in the BFKL equation, there is an intrusion from hard scattering to the regime of soft interactions and vice versa, and the effect of the soft region is especially enhanced by AF. In the numerical analysis[15] an infrared freezing $\alpha_S(q^2)\leqslant\alpha_S^{(fr)}=0.82$ has been imposed on the three-flavor, one-loop $\alpha_S(q^2)=4\pi/9\log(k^2/\Lambda^2)$ with $\Lambda=0.3$ GeV. With $R_c=0.27$ fm, i.e., $\mu_G=0.75$ GeV, we found $\Delta_{\mathrm{P}}=0.4$.[15] The calculation of Regge trajectories of subleading pomeron singularities is reported in Ref. 27; the emerging successful description of the proton structure function at small x is published in Refs. 26 and 27.

In Ref. 29 the gBFKL equation (29) is generalized to the profile function $\Gamma(\xi,\vec{r},\vec{b})$, where the impact parameter $\vec{b}$ is defined with respect to the center of the dipole:

$$\frac{\partial\Gamma(\xi,\vec{r},\vec{b})}{\partial\xi}=\mathscr{K}\otimes\Gamma(\xi,\vec{r},\vec{b}) =\frac{3}{8\pi^3}\int d^2\rho_1\mu_G^2\left|g_S(R_1)K_1(\mu_G\rho_1)\,\frac{\vec{\rho}_1}{\rho_1} -g_S(R_2)K_1(\mu_G\rho_2)\,\frac{\vec{\rho}_2}{\rho_2}\right|^2\left[\Gamma\left(\xi,\vec{\rho}_1,\vec{b}+\frac{1}{2}\,\vec{\rho}_2\right)+\Gamma\left(\xi,\vec{\rho}_2,\vec{b}+\frac{1}{2}\,\vec{\rho}_1\right)-\Gamma(\xi,\vec{r},\vec{b})\right]. \tag{32}$$

The calculation of the impact parameter integral (25) reduces Eq. (32) to Eq. (29). The calculation of the moment (27) leads to the integral equation for $\lambda(\xi,r)$. It is convenient to separate from the diffraction slope $B(\xi,r)$ the purely geometrical term $r^2/8$ related to the elastic form factor of the color dipole of the dipole moment r, and to discuss instead of $\lambda(\xi,r)$ the function

$$\eta(\xi,r)=\lambda(\xi,r)-\frac{1}{8}\,r^2\sigma(\xi,r),$$

which satisfies the inhomogeneous integral equation

$$\frac{\partial\eta(\xi,r)}{\partial\xi}=\frac{3}{8\pi^3}\int d^2\rho_1\mu_G^2\left|g_S(R_1)K_1(\mu_G\rho_1) \times\frac{\vec{\rho}_1}{\rho_1}-g_S(R_2)K_1(\mu_G\rho_2)\,\frac{\vec{\rho}_2}{\rho_2}\right|^2 \times\left\{\eta(\xi,\rho_1)+\eta(\xi,\rho_2)-\eta(\xi,r)+\frac{1}{8} \times(\rho_1^2+\rho_2^2-r^2)[\sigma(\xi,\rho_2)+\sigma(\xi,\rho_1)]\right\} =\mathscr{K}\otimes\eta(\xi,r)+\beta(\xi,r), \tag{33}$$

where the inhomogeneous term is

$$\beta(\xi,r)=\mathscr{L}\otimes\sigma(\xi,r) =\frac{3}{64\pi^3}\int d^2\rho_1\mu_G^2\left|g_S(R_1)K_1(\mu_G\rho_1)\,\frac{\vec{\rho}_1}{\rho_1} -g_S(R_2)K_1(\mu_G\rho_2)\,\frac{\vec{\rho}_2}{\rho_2}\right|^2(\rho_2^1+\rho_2^2-r^2) \times[\sigma(\xi,\rho_2)+\sigma(\xi,\rho_1)]. \tag{34}$$

Because the homogeneous part of Eq. (33) coincides with the gBFKL equation (29), asymptotically the dipole cross section $\sigma(\xi,r)$ and the solution $\eta(\xi,r)$ of homogeneous Eq. (33) have identical energy dependence. Consequently, the solutions of the homogeneous Eq. (33) give the

asymptotically constant contribution to the diffraction cone and if $\sigma_a(\xi,r)$ is a solution of Eq. (29) and $\eta_a(\xi,r)$ is a solution of Eq. (33) with the diffraction slope $B_a(\xi,r)$, then $\eta_b(\xi,r)=\eta_a(\xi,r)+\Delta b\cdot\sigma_a(\xi,r)$, where $\Delta b=$const, is also a solution of Eq. (33) with the diffraction slope $B_b(\xi,r)=B_a(\xi,r)+\Delta b$. It is the inhomogeneous term, $\beta(\xi,r)$, which gives rise to $\eta(\xi,r)\propto\xi\sigma(\xi,r)$, i.e., to the asymptotic Regge growth of the diffraction slope, $B(\xi,r)=B(\xi_0,r)+2\alpha'_{\mathbb{P}}\xi$, and the Regge term $2\alpha'_{\mathbb{P}}\xi$ does not depend on the size r of the dipole. Parametrically, $\alpha'_{\mathbb{P}}\propto\alpha_S(R_c)R_c^2$ times a small numerical factor. With the infrared parameters specified above, $\alpha'_{\mathbb{P}}\approx 0.072\ \mathrm{GeV}^{-2}$ was found in Ref. 28; for slopes of subleading trajectories see Ref. 29.

4. BEAM, TARGET AND EXCHANGE DECOMPOSITION OF THE DIFFRACTION SLOPE

To obtain more insight into the dipole-size dependence of the diffraction slope, it is useful to look at the scattering amplitude $\sigma(\xi,r,\vec{q})$ in terms of the gluon density matrix. For our purposes, it is sufficient to treat the color structure of the proton in terms of the three valence (constituent) quarks. Then, as illustrated graphically in Fig. 1b, the unintegrated density matrix of gluons can be written

$$\begin{aligned}\mathscr{F}&\left(x,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}\right)\\&=\frac{4}{\pi}\int d^2k_1\mathscr{F}\left(\xi,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q},\vec{k}_1\right.\\&\left.+\frac{1}{2}\vec{q},-\vec{k}_1+\frac{1}{2}\vec{q}\right)\alpha_S(k_1^2)\Big[G_1(q^2)\\&-G_2\left(\vec{k}_1+\frac{1}{2}\vec{q},-\vec{k}_1+\frac{1}{2}\vec{q}\right)\Big],\end{aligned}\tag{35}$$

where $G_1(q^2)$ and $G_2(\vec{\kappa}_1,\vec{\kappa}_2)$ are the single- and two-quark form factors of the proton probed by gluons and $\mathscr{F}(\xi,\vec{k}+\vec{q},-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q},-\vec{k}_1+\vec{q}/2)$ denotes the propagation function of two t-channel gluons. In the Born approximation,

$$\begin{aligned}\mathscr{F}&\left(\xi,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q},\vec{k}_1+\frac{1}{2}\vec{q},-\vec{k}_1+\frac{1}{2}\vec{q}\right)\\&=\frac{\delta(\vec{k}-\vec{k}_1)}{[(\vec{k}+\vec{q}/2)^2+\mu_G^2][(\vec{k}-\vec{q}/2)^2+\mu_G^2]}.\end{aligned}\tag{36}$$

Splitting the color dipole vertex function into two pieces,

$$\begin{aligned}V_d(q,r)&=\left[J_0\left(\frac{1}{2}qr\right)-J_0(kr)\right]\\&=\left[J_0\left(\frac{1}{2}qr\right)-1\right]+[1-J_0(kr)],\end{aligned}$$

we obtain the useful decomposition

$$\begin{aligned}\sigma(\xi,r,\vec{q})=\frac{4\pi}{3}&\left[J_0\left(\frac{1}{2}qr\right)-1\right]\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)\mathscr{F}\left(x,\vec{k}\right.\\&\left.+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}+\frac{4\pi}{3}\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)\right.\\&\left.\times[J_0(kr)-1]\mathscr{F}\left(x,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}\right)\right.\end{aligned}\tag{37}$$

Because of the property defined by Eq. (14), the second term has the typical logarithmic k^2 integration. It comprises the contributions to the q dependence from the target and exchanged gluons. In contrast, such a logarithmic k^2 integration is absent from the first term; here the k^2 integration converges at finite $k^2\sim R_c^{-2}$.

The resulting representation

$$\begin{aligned}\sigma(\xi,r,\vec{q})=\frac{4\pi}{3}&\left[J_0\left(\frac{1}{2}qr\right)-1\right]\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)\\&\times\mathscr{F}\left(x,\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}\right)+\frac{16}{9}\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)\\&\times[1-J_0(kr)]\int d^2k_1\mathscr{F}\left(\xi,\vec{k}+\vec{q},-\vec{k}+\frac{1}{2}\vec{q},\vec{k}_1\right.\\&\left.+\frac{1}{2}\vec{q},-\vec{k}_1+\frac{1}{2}\vec{q}\right)\alpha_S(k_1^2)\Big[G_1(q^2)\\&-G_2\left(\vec{k}_1+\frac{1}{2}\vec{q},-\vec{k}_1+\frac{1}{2}\vec{q}\right)\Big]\end{aligned}\tag{38}$$

nicely illustrates how the three relevant size parameters in the problem give rise to the three major components of the diffraction slope. The q dependence coming from the proton vertex function $V_p(\vec{k}_1,\vec{q})=G_1(q^2)-G_2(\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)$ is controlled by the proton size. The q dependence coming from the color dipole vertex function $V_d=J_0(qr/2)-1$ is controlled by the color dipole size r. The q dependence coming from $\mathscr{F}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)$ depends on the effective k^2 and k_1^2, which contribute to the scattering amplitude, and on the gluon propagation radius R_c. The latter scale remains important even at large k, because the properties of the running gBFKL pomeron are controlled by interactions at $r\sim R_c$. In the asymptotic BFKL regime, at small x, the $\vec{k}_1$ and $\vec{k}$ become azimuthally uncorrelated.

To proceed further, one needs a model for $G_1(q^2)$ and $G_2(\vec{\kappa}_1,\vec{\kappa}_2)$. The radius R_N of the proton probed by the gluon can be different from the charge radius R_{ch}; still R_{ch} serves as a useful scale. The two-quark form factor $G_2(\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2)$ is a steep function of k^2 and a smoother function of q^2.[41] For instance, for the oscillator wave function of the 3-quark proton one readily finds

$$G_2\left(\vec{k}+\frac{1}{2}\vec{q},-\vec{k}+\frac{1}{2}\vec{q}\right)=G_1\left(\frac{1}{4}q^2\right)G_1(3k^2).\tag{39}$$

A straightforward differentiation yields the following transparent decomposition of $d\sigma(\xi,r,\vec{q})/dq^2$ into four terms:

$$\left.\frac{d\sigma(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}$$
$$=\sum_{i=1}^{4}\left.\frac{d\sigma^{(i)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}=-\frac{16}{3}\int\frac{d^2k}{k^4}\,\alpha_S(\kappa^2)$$
$$\times\left\{\frac{1}{16}\,r^2\int d^2k_1\mathscr{T}(\xi,\vec{k},-\vec{k},\vec{k}_1,-\vec{k}_1)\alpha_S(k_1^2)\right.$$
$$\times[1-G_2(\vec{k}_1,\vec{k}_1)]-[1-J_0(kr)]\int d^2k_1\alpha_S(k_1^2)$$
$$\times[1-G_2(\vec{k}_1,\vec{k}_1)]$$
$$\times\left.\frac{\partial\mathscr{T}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)}{\partial q^2}\right|_{q^2=0}$$
$$+\frac{1}{6}\,R_N^2[1-J_0(kr)]\int d^2k_1\mathscr{T}(\xi,\vec{k},-\vec{k},\vec{k}_1,-\vec{k}_1)\alpha_S(k_1^2)$$
$$-\frac{1}{24}\,R_N^2[1-J_0(kr)]\int d^2k_1\mathscr{T}(\xi,\vec{k},-\vec{k},\vec{k}_1,-\vec{k}_1)$$
$$\left.\times\alpha_S(k_1^2)G_2(\vec{k}_1,\vec{k}_1)\right\}. \tag{40}$$

The following properties of $\mathscr{T}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)$ are important in (40). First, in the infrared-regulated QCD it is nonsingular at $k^2=0$; cf. Eq. (36). Second (modulo logarithmic scaling violations), its large-k^2 asymptotic behavior is similar to that of the Born term (36), $\mathscr{T}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)\propto 1/k^{-4}$. Third, in the Born approximation (after azimuthal averaging)

$$\left.\frac{\partial\mathscr{T}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)}{\partial q^2}\right|_{q^2=0}$$
$$=-\frac{R_c^2}{(1+R_c^2k^2)^2}\,\mathscr{T}\left(\xi,\vec{k}+\frac{1}{2}\,\vec{q},-\vec{k}+\frac{1}{2}\,\vec{q},-\vec{k}_1+\frac{1}{2}\,\vec{q},-\vec{k}_1+\frac{1}{2}\,\vec{q}\right). \tag{41}$$

Fourth, finding the asymptotic Regge growth of the diffraction slope in Ref. 28 implies that in the high-energy limit $\xi\to\infty$ and for all $\vec{k}$ and $\vec{k}_1$,

$$\left.\frac{\partial\mathscr{T}(\xi,\vec{k}+\vec{q}/2,-\vec{k}+\vec{q}/2,\vec{k}_1+\vec{q}/2,-\vec{k}_1+\vec{q}/2)}{\partial q^2}\right|_{q^2=0}$$
$$=-[\alpha'_{\mathbb{P}}(\xi-\xi_0)+O(R_c^2)]\mathscr{T}\left(\xi,\vec{k}+\frac{1}{2}\,\vec{q},-\vec{k}+\frac{1}{2}\,\vec{q},\vec{k}_1+\frac{1}{2}\,\vec{q},-\vec{k}_1+\frac{1}{2}\,\vec{q}\right). \tag{42}$$

Consider first the decomposition of the diffraction slope for large dipoles, $r\gtrsim R_c$. In this limit, the cross section (38) is dominated by the contribution from $k^2\sim\mu_G^2=R_c^{-2}\gg R_N^{-2}$, so that $J_0(kr),G_1(3k^2)\ll 1$; they can be neglected, and (cf. Eq. (9))

$$\sigma(\xi,r)=\frac{4\pi}{3}\int d^2k\,\alpha_S^2(k^2)\mathscr{T}(\xi,\vec{k},\vec{k}).$$

The first term in the expansion (40) can then be evaluated as

$$\left.\frac{d\sigma^{(1)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}=-\frac{1}{2}\,\Delta b_1\sigma(\xi,r)=-\frac{1}{16}\,r^2\sigma(\xi,r). \tag{43}$$

Similarly,

$$\left.\frac{d\sigma^{(3)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}=-\frac{1}{2}\,\Delta b_3\sigma(\xi,r)=-\frac{1}{6}\,R_N^2\sigma(\xi,r). \tag{44}$$

The integrand of the fourth term in (40) contains the steeply decreasing two-body form factor $G_2(\vec{k},-\vec{k})$, which cuts off the integration at $k^2\lesssim R_N^2$. Consequently, one must distinguish between the cases $r\lesssim R_N$ and $r\gtrsim R_N$. A simple estimate, which interpolates between these limiting cases, is

$$\left.\frac{d\sigma^{(4)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}=-\frac{1}{2}\,\Delta b_4\sigma(\xi,r)$$
$$=-\frac{1}{24}\,R_c^2\sigma(\xi,r)\,\frac{r^2}{r^2+R_N^2}. \tag{45}$$

The bottom line is that $\Delta b_4\ll\Delta b_3$. Finally, making use of (42), the second term in (40) can be estimated as

$$\left.\frac{d\sigma^{(2)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0}=-\frac{1}{2}\,\Delta b_2\sigma(\xi,r)=-[\alpha'_{\mathbb{P}}(\xi-\xi_0)+O(R_c^2)]\sigma(\xi,r). \tag{46}$$

At low energy, in the Born approximation, Eq. (41) gives $\Delta b_2=2R_c^2$. The salient feature of the resulting diffraction slope

$$B(\xi,r)=\sum_i\,\Delta b_i=\frac{1}{8}\,r^2+\frac{1}{3}\,R_N^2+2\alpha'_{\mathbb{P}}(\xi-\xi_0)+O(R_c^2), \tag{47}$$

is the presence of the geometrical contributions $\Delta b_1=r^2/8$ and $\Delta b_3=R_N^2/3$.

For large dipoles, $r\gtrsim R_c$, one recovers a sort of additive quark model, in which uncorrelated gluonic clouds build up around the beam and target quarks and antiquarks, and the terms $O(R_c^2)$ and $2\alpha'_{\mathbb{P}}(\xi-\xi_0)$ describe the familiar Regge growth of diffraction slope for quark–quark scattering. The opposite limit of small dipoles, $r\ll R_c$, is somewhat more tricky. In the second and third term in (40), the k^2 integration is cut off by $1-J_0(kr)$ and extends up to A_σ/r^2, precisely as in the dipole cross section (10). Consequently, their contributions to the derivative (40) are still given by Eqs. (46) and (44), respectively, so that the Regge term and the contribution from the target proton size to expansion (47) are retained. The contribution from the first term in (40), i.e., from the size of the color dipole, changes dramatically and will no longer have the geometric form $r^2/8$. Indeed, as we discussed following Eq. (37), the k^2 integration in the first

term in (40) converges at $k^2 \lesssim R_c^2$. Consequently, in this limit $\kappa^2 = C^2/r^2$, and one can factor out $\alpha_S(\kappa^2) = \alpha_S(r)$ from the integrand. This leads to the estimate

$$\left.\frac{d\sigma^{(1)}(\xi,r,\vec{q})}{dq^2}\right|_{q^2=0} \approx -\frac{r^2}{16}\frac{\alpha_S(r)}{\alpha_S(R_c)}\,\sigma(\xi,R_c)$$

$$\approx -\frac{r^2}{16}\frac{\pi^2}{3}\,\alpha_S(r)R_c^2 G\left(\xi,\frac{A_\sigma}{R_c^2}\right) \quad (48)$$

and, after making use of (11), to

$$\Delta b_1 = \frac{R_c^2}{8}\frac{G(\xi,A_\sigma/R_c^2)}{G(\xi,A_\sigma/r^2)}. \quad (49)$$

Similar considerations yield an estimate for the contribution to the diffraction slope from the fourth term in (40), which is a negligible correction to Δb_1:

$$\Delta b_4 = \frac{R_c^2}{12}\frac{R_c^2}{R_N^2}\frac{G(\xi,A_\sigma/R_c^2)}{G(\xi,A_\sigma/r^2)}. \quad (50)$$

More comments on Δb_1 are in order. At asymptotically large ξ and/or asymptotically small x, the running gBFKL approach predicts the universal x dependence of the gluon structure function[17]

$$G(x,Q^2) \propto \left[\frac{1}{\alpha_S(Q^2)}\right]^\gamma \left(\frac{1}{x}\right)^{\Delta_{\mathrm{P}}}, \quad (51)$$

where $\gamma = 12\Delta_{\mathrm{P}}/\beta_0$ and $\beta_0 = 11 - 2n_f/3$. Consequently, in the well-developed BFKL regime, Δb_1 will not depend on energy:

$$\Delta b_1 = \frac{R_c^2}{8}\left[\frac{\alpha_S(r)}{\alpha_S(R_c)}\right]^\gamma. \quad (52)$$

However, at moderately small x values, the x dependence of the gluon structure function exhibits strong dependence on the factorization scale, the ratio $G(\xi,A_\sigma/R_c^2)/G(\xi,A_\sigma/r^2)$ has substantial x dependence, and Δb_1 contributes to the energy dependence of the diffraction cone. Specifically, it makes the slope of the effective Regge trajectory α'_{eff} substantially larger than the true slope of the leading Pomeron trajectory α'_{P}.[28]

To summarize, the geometrical contribution to the diffraction slope from the target proton size, $\Delta b_3 = (1/3)R_N^2$, persists for all dipole sizes (the term Δb_4, which is also associated with the proton size, is negligible in all cases). Although the nonperturbative parameter R_N^2 is not calculable from first principles, its contribution to the diffraction slope varies neither with energy nor with dipole size and can eventually be fixed using accurate experimental data.

5. SOFT POMERON AND DIFFRACTIVE SCATTERING OF LARGE COLOR DIPOLES

The need for a soft pomeron contribution in addition to the gBFKL dipole cross section described previously is brought about by phenomenological considerations. A viable gBFKL phenomenology of the rising component of the proton structure function over the whole range of Q^2 studied at HERA (real photoabsorption included) is obtained if one starts with the Born dipole cross section $\sigma_B(r)$ as a boundary condition for the gBFKL evolution at $x_0 = 0.03$.[26,27] However, such a $\sigma_B(r)$ falls short of the interaction strength at $r \gtrsim R_c$; roughly speaking, for the phenomenological value $R_c = 0.27$ fm one finds $\sigma_B(r \gtrsim 1\ \mathrm{fm}) \sim 5$ mb, whereas for the description of soft processes one rather needs the dipole cross section ~ 50 mb at $r \gtrsim 2$ fm. Therefore, at $r \gtrsim R_c$, the perturbative gBFKL dipole cross section described above (which hereafter we denote by the subscript ''pt''), $\sigma_{\mathrm{pt}}(\xi,r)$, must be complemented by the contribution from the nonperturbative soft pomeron, $\sigma_{\mathrm{npt}}(\xi,r)$. Because in all the cases studied the contribution from $\sigma_{\mathrm{pt}}(\xi,r)$ exhausts the rise of the total cross sections and/or of the proton structure function, in Refs. 26 and 5 we have modeled the soft nonperturbative pomeron by the energy-independent $\sigma_{\mathrm{npt}}(\xi,r) = \sigma_{\mathrm{npt}}(r)$. For the lack of better theoretical and experimental information as well as simplicity, we make the simplest possible assumption that the eikonals for the perturbative and soft interactions are additive, which to lowest order amounts to additivity of the dipole cross sections $\sigma(\xi,r) = \sigma_{\mathrm{pt}}(\xi,r) + \sigma_{\mathrm{npt}}(r)$.

The direct determination of the total dipole cross section $\sigma(\xi,r)$ from the experimental data on photo- and leptoproduction of vector mesons is reported in Ref. 9, and supports the flavor independence of $\sigma(\xi,r)$. Other constraints for $\sigma_{\mathrm{npt}}(r)$ include real photoproduction,[5,10] hadronic diffractive scattering,[24] nuclear shadowing in deep inelastic scattering,[42] diffractive deep inelastic scattering at HERA,[43,44] nuclear attenuation in photoproduction of light vector mesons and the onset of color transparency in leptoproduction of vector mesons[4] and the proton structure function at moderate and small Q^2.[26,27] All the results are consistent with the form of the dipole cross section suggested in Refs. 24, 26, and 5, a convenient parameterization for which is

$$\sigma_{\mathrm{npt}}(r) = \sigma_0\left[1 - \sum_{i=1}^{2} A_i \exp\left(-\frac{r^2}{a_i^2}\right)\right] \times \left[1 + \sum_{i=1} D_i \exp\left(-\frac{(r-b_i)^2}{c_i^2}\right)\right] \quad (53)$$

with $\sigma_0 = 41.2$ mb, $A_1 = 1.45$, $A_2 = -0.45$, $a_1 = 1.30$ fm, $a_2 = 0.75$ fm, $D_1 = 0.80$, $D_2 = 0.36$, $b_1 = 0.88$ fm, $b_2 = 2.08$ fm, $c_1 = 0.53$ fm, and $c_2 = 1.14$ fm. For a somewhat cruder fit with $D_i = 0$ we find $\sigma_0 = 51.6$ mb, $A_1 = 1.82$, $A_2 = -0.82$, $a_1 = 1.05$ fm, and $a_2 = 0.72$ fm. For small dipoles, $r \ll R_c$, this cross section is poorly known because it is swamped by $\sigma_{\mathrm{pt}}(\xi,r)$.

There is nothing unusual in the concept of a nonperturbative cross section. The conventional gluon structure function of the photon,

$$G(x,Q^2) = \int_0^{Q^2} \frac{dk^2}{k^2}\,\mathscr{F}(x,\vec{k},-\vec{k}),$$

always contains a contribution from gluons with soft transverse momenta $k^2 < Q_0^2 \lesssim 1\ \mathrm{GeV}^2$, which persists at all Q^2 and equals precisely $G(x,Q_0^2)$, the familiar input to the conventional GLDAP analysis of the Q^2 evolution of parton densities. One is perfectly content with the strong sensitivity

of GLDAP evolution to this unknown soft input $G(x,Q_0^2)$, which is routinely fixed from fits to the experimental data. In the color dipole approach to DIS, our soft dipole cross section $\sigma_{\rm npt}(r)$ plays exactly the same role as the gluon (quark) structure functions at a soft scale Q_0^2. Furthermore, it is tempting to reinterpret this soft dipole cross section $\sigma_{\rm npt}(r)$ in terms of the nonperturbative gluon distribution in the spirit of Eq. (10). Models of soft scattering via polarization of the nonperturbative QCD vacuum[36,11] belong to this category and yield a $\sigma_{\rm npt}(r)$ very similar to our parameterization (53). In the interesting region of $r\lesssim 1-1.5$ fm, a conservative estimate of uncertainties in $\sigma_{\rm npt}(r)$ is 10–20%, the major source of uncertainty being due to absorption corrections. For heavy quarkonia the absorption corrections are negligible.[5]

We assume the conventional Regge rise of the diffraction slope for the soft pomeron,

$$B_{\rm npt}(\xi,r)=\Delta B_d(r)+\Delta B_N+2\alpha'_{\rm npt}(\xi-\xi_0),$$

where $\Delta B_d(r)$ and ΔB_N denote the contribution of the beam dipole and target nucleon size, and $\xi_0=\log(1/x_0)$. As guidance we take the experimental data on pion–nucleon scattering,[45] which suggest $\alpha'_{\rm npt}=0.15\ {\rm GeV}^{-2}$ (for small $\alpha'_{\rm npt}$ descriptions of nucleon–nucleon scattering see Ref. 46). A plausible guess for the proton size contribution is

$$\Delta B_N=\Delta b_3=\frac{1}{3}\,R_N^2. \tag{54}$$

In energy-independent soft exchange for small dipoles, $\Delta B_d(r)$ is likely to follow the geometric law $\Delta B_d(r)\approx r^2/8$, as in Eq. (43). Extension of this law to large dipoles is questionable. The large-r saturation of $\sigma_{\rm npt}(r)$ as parametrized by (53) is a simplifying assumption; what happens at $r\gg 1$ fm is immaterial, because even in hadrons the probability of finding large dipoles, $r\gg 1$ fm, is negligible. However, the diffraction slope is more sensitive to the large dipole contribution. For instance, if scattering of large dipoles of size $r\gtrsim R_N$ is modeled by scattering of thin classical strings by a strongly absorbing target nucleon of radius R_N, then for large dipoles (strings), $r\gtrsim 2R_N$, one readily finds the profile function

$$\Gamma(b,r)\approx\theta(R-b)+\frac{2}{\pi}\,\theta(b-R)\,\theta\left(R+\frac{1}{2}\,r-b\right)\sin^{-1}\frac{R}{b},$$

which yields the large-r behavior $\sigma_{\rm npt}(r)\approx 2R_N r$ and the tamed rise of the diffraction slope $\Delta B_d(r\gg 1\ {\rm fm})\sim r^2/24$. This consideration suggests an interpolation formula of the form

$$B_d(r)=\frac{r^2}{8}\,\frac{r^2+aR_N^2}{3r^2+aR_N^2}, \tag{55}$$

where a is a phenomenological parameter, $a\sim 1$.

Such a taming of the growth of $B_d(r)$ is supported by the phenomenology of πN scattering. Let us take for the pion the oscillator wave function, and assume that the gluon-probed radius of the pion equals the charge radius. Then, the contribution of the pion size to the diffraction slope for the purely geometrical form (43) of $B_d(r)$ yields the unacceptably large value

$$\Delta B_\pi=\frac{1}{8}\,\frac{\langle\pi|(r^2/8)[\sigma_{\rm pt}(x_0,r)+\sigma_{\rm npt}(r)]|\pi\rangle}{\langle\pi|[\sigma_{\rm pt}(x_0,r)+\sigma_{\rm npt}(r)]|\pi\rangle}\approx 9.7\ \ {\rm GeV}^{-2}.$$

Taking for the contribution of the proton size ΔB_N the estimate (54) we end up with $B_{\pi N}\approx 15\ {\rm GeV}^{-2}$, which substantially exceeds the experimental result $B_{\pi N}(\nu=200\ {\rm GeV})=9.9\pm 0.1\ {\rm GeV}^{-2}$.[45] The discrepancy increases further if one adds to the above theoretical estimate the Regge term $2\alpha'_{\rm npt}(\xi-\xi_0)\approx 1\ {\rm GeV}^{-2}$ evaluated using the relationship between $x_{\rm eff}$ and the pion energy, $x_{\rm eff}\approx m_V^2/2\nu m_p$.

What is the origin of this discrepancy? If $\sigma(\xi,r)$ were r-independent and if the gluon-probed and charge radii of the pion were identical, then one would find from (5) the familiar

$$\Delta B_\pi(\xi_0)=\frac{1}{3}\,\langle R_{ch}^2\rangle_\pi\approx 4\ \ {\rm GeV}^{-2}. \tag{56}$$

With our parametrization (53), the soft dipole cross section keeps rising at $r\sim 1$ fm, and for this reason the matrix element (5) is dominated by an r^2 larger than in the charge radius of the pion, and we end up with ΔB_π larger than the expectation (56) based on the charge radius of the pion. The matrix element (5) can be made smaller and compatible with the experiment at the expense of rapid saturation of the soft cross section for large dipoles, $\sigma_{\rm npt}(r\gtrsim 1\ {\rm fm})\approx\sigma_{\rm tot}(\pi N)$, whereupon one recovers the estimate (56). This solution must be rejected, because it would lead to negligible fluctuations of the soft dipole cross section, in conflict with experimental data on diffraction dissociation of pions, which require

$$\frac{\langle\pi|\sigma^2(\nu_0,r)|\pi\rangle-\langle\pi|\sigma(\nu_0,r)|\pi\rangle^2}{\langle\pi|\sigma(\nu_0,r)|\pi\rangle^2}\approx 0.5. \tag{57}$$

An attempt to retain the geometrical law and still agree with experiment at the expense of taking $\Delta B_N\sim 0$ must be rejected as well. We believe that the string-model suggested taming of $B_d(r)$ Eq. (55) is a more acceptable solution. Hereafter we take $\Delta B_N=\Delta b_3=4.8\ {\rm GeV}^{-2}$. Then the pion–nucleon diffraction slope is reproduced with reasonable values of the parameter a in Eq. (55): $a=1.2$ for $\alpha'_{\rm npt}=0.1\ {\rm GeV}^{-2}$ and $a=0.9$ for $\alpha'_{\rm npt}=0.15\ {\rm GeV}^{-2}$. Hereafter we use the latter set of parameters.

6. SOFT–HARD DECOMPOSITION OF TOTAL CROSS SECTIONS FOR *VN* SCATTERING

We present our results starting with an evaluation of the vector meson–nucleon total cross section

$$\sigma_{\rm tot}(VN)=\frac{N_c}{2\pi}\int_0^1\frac{dz}{z^2(1-z)^2}\int d^2r\{m_q^2\phi(r,z)^2+[z^2+(1-z)^2][\partial_r\phi(r,z)]^2\}\sigma(x_{\rm eff},r). \tag{58}$$

For the parametrization of light-cone wave functions $\phi(r,z)$ of vector mesons, see Ref. 10. The results for $x_{\rm eff}\leqslant x_0=0.03$ are shown in Fig. 2 (left box). The smaller the radius

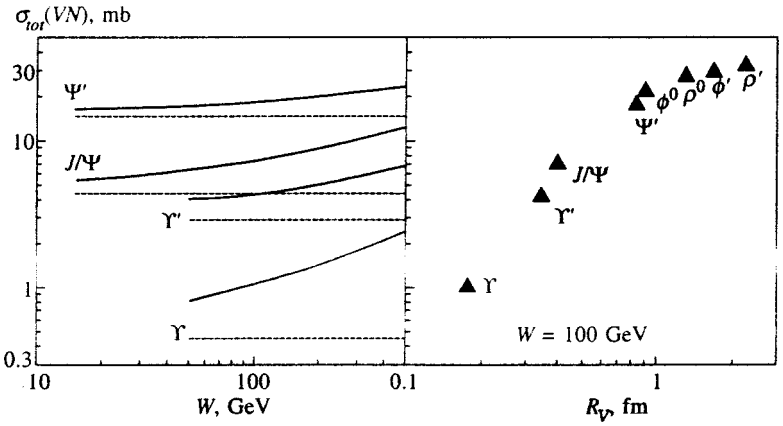


FIG. 2. The left box corresponds to the color dipole model predictions for the total cross section $\sigma_{\rm tot}(VN)$ for the interaction of the heavy vector mesons J/Ψ, Ψ', Υ, and Υ' with the nucleon target as a function of c.m.s. energy W. The dashed curve represents the soft pomeron contribution. The right box shows the color dipole model predictions for the total cross section $\sigma_{\rm tot}(VN)$ vs. radius R_V of vector mesons ρ^0, ρ', ϕ^0, ϕ', J/Ψ, Ψ', Υ, and Υ'.

of the vector meson V, the smaller the total cross section $\sigma_{\rm tot}(VN)$; to a crude approximation, $\sigma_{\rm tot}(VN)\propto R_V^2$, excepting the radial excitations ϕ',ρ'.

In Fig. 2 (left box) we show separately the soft pomeron contribution to $\sigma_{\rm tot}(VN)$. For the J/Ψ the radius is large, $R_{J/\Psi}\approx 0.4$ fm$>R_c=0.27$ fm, and the soft contribution is substantial; for the Υ the soft contribution is a small correction to the dominant perturbative contribution. At subasymptotic energies, the gBFKL approach predicts a steeper rise with energy for smaller dipoles (cf. Eq. (11)), a trend that is clearly visible in Fig. 2 (left box). At asymptotic energies the contribution from the rising gBFKL cross section takes over for all channels. In Ref. 47 it is observed that for the ''magic'' radius $r_\Delta\sim 0.15$ fm$\sim R_c/2$ the gBFKL color dipole cross section exhibits the precocious asymptotic energy dependence $\sigma_{\rm pt}(x,r_\Delta)\propto x^{-\Delta_{\rm P}}$. Because $R_\Upsilon\approx 0.18$ fm is very close to the ''magic'' radius r_Δ, the predicted energy dependence of the perturbative contribution to $\sigma_{\rm tot}(\Upsilon N)$ is very close to $\propto W^{2\Delta_{\rm P}}=W^{0.8}$.

The case of the Ψ' is interesting for its large radius $R_{\Psi'}\approx 0.8$ fm and large soft contribution. Because the Ψ' and ϕ^0 have very similar radii, a useful comparison is with $\sigma_{\rm tot}(\phi N)$. For small W, $\sigma_{\rm tot}(\Psi' N)$ of the present paper is indeed numerically very close to $\sigma_{\rm tot}(\phi^0 N)$ calculated in Ref. 10, but the rise of $\sigma_{\rm tot}(\Psi' N)$ by $\sim$50% from $W\sim 10$ GeV to $W\sim 500$ GeV is much weaker than the rise of $\sigma_{\rm tot}(\phi N)$ by almost a factor 2 over the same energy range. With our energy-independent $\sigma_{\rm npt}(r)$, the rise of $\sigma_{\rm tot}(\Psi' N)$ is entirely due to the perturbative gBFKL cross section $\sigma_{\rm pt}(\xi,r)$, which rises with energy more steeply at small r. Although the Ψ' and the ϕ^0 have similar mean square radii, because of the node effect the relative contribution of small r for the case of $\Psi' N$ is smaller than for the case of ϕN, and this explains the counterintuitive difference of the energy dependence of the two cross sections.

7. DIFFRACTIVE PRODUCTION CROSS SECTIONS FOR THE 1*S* STATES *J*/Ψ AND Υ

We now turn to vector meson production. The strong point about color dipole factorization equations (13), (14), (19), and (20) is that apart from the trivial factors C_V and C_V/m_V, the production amplitudes are flavor-independent when considered as a function of the scanning radius r_S

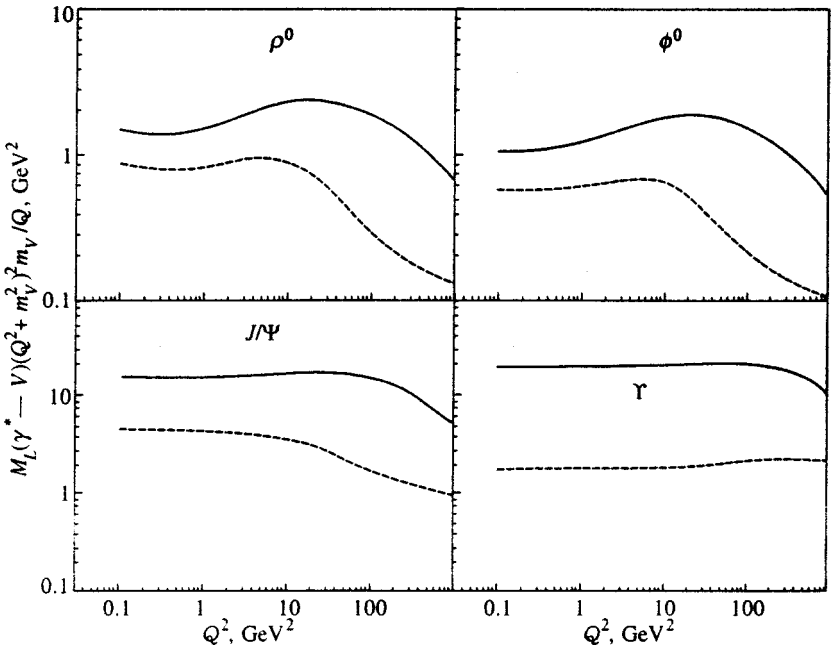


FIG. 3. Decomposition of production amplitude for longitudinally polarized vector mesons into soft (dashed curves) and perturbative+soft (solid curves) contribution as a function of $Q^2+m_V^2$. The nonmonotonic Q^2 dependence is due to an increase in $x_{\rm eff}$ at fixed $W=150$ GeV.

and/or $Q^2+m_V^2$.[3–5,28,10] To this end, Eqs. (13), (14), (19), and (20) represent the leading twist terms and the correct twist expansion goes in powers of $1/(Q^2+m_V^2)$ rather than in powers of $1/Q^2$. For instance, in Ref. 10 we have shown how the ratio of the J/Ψ and ρ production cross sections becomes remarkably constant when the two cross sections are taken at equal $Q^2+m_V^2$, in contrast to a variation by about three orders of magnitude when the two cross sections are compared at equal Q^2. For this reason we strongly advocate the presentation of the experimental data as a function of the flavor-symmetry restoring variable $Q^2+m_V^2$ rather than Q^2, and whenever appropriate we present our results in terms of this scaling variable.

The soft/hard decomposition of production amplitudes depends on the relationship between r_S and R_c. The hard contribution dominates at $r_S \lesssim R_c$, i.e., if

$$Q^2+m_V^2 \gtrsim \frac{A^2}{Rc^2} \sim 30 \ \ \mathrm{GeV}^2, \tag{59}$$

which holds better for the heavier vector mesons and the larger Q^2. Our phenomenological soft interaction, as well as other models for the soft pomeron,[36,11] extends well into $r \lesssim R_c$. Arguably, with better understanding of the perturbative gBFKL amplitude, one can eventually use vector meson production to better assess the effect of soft interactions at short distances. In Fig. 3 we show our decomposition of the production amplitudes into a hard (perturbative) and soft contribution as a function of Q^2 for various vector mesons at the typical HERA energy $W=150$ GeV. Because the pQCD scale parameter is smaller for the lighter mesons, $\tau_L(\rho) < \tau_L(J/\Psi) < \tau_L(\Upsilon)$, the soft contribution is somewhat larger for the lighter quarkonia.

For comparison with experimental data, the most straightforward theoretical predictions are for forward production, and we calculate $(d\sigma/dt)_{t=0}$ and $B(t=0)$. The experimental determination of these quantities requires extrapolation of $d\sigma/dt$ to $t=0$, which is not always possible, and one often reports the t-integrated production cross sections. The principal lesson from high precision $\pi^{\pm}N$ scattering experiments is that the diffraction slope $B(t)$ depends strongly on the region of t, and for the average $\langle t \rangle \sim 0.1-0.2\ \mathrm{GeV}^2$ which dominate the integrated total cross section, the diffraction slope is less than at $t=0$ by $\sim 1\ \mathrm{GeV}^{-2}$.[45] We take these πN scattering data for guidance, and for more direct comparison with the presently available experimental data, instead of the directly calculated $B(t=0)$, in all cases we report

$$B = B(t=0) - 1 \ \ \mathrm{GeV}^{-2}, \tag{60}$$

which we also use to evaluate the t-integrated production cross section from the theoretically calculated $(d\sigma/dt)_{t=0}$:

$$\sigma(\gamma^* \to V) = \frac{1}{B} \left. \frac{d\sigma(\gamma^* \to V)}{dt} \right|_{t=0}. \tag{61}$$

The uncertainties in the value of B and with the evaluation (61) presumably do not exceed 10%, and can be reduced when more accurate data become available.

We begin by presenting our results and comparing them with the available experimental data on real photoproduction of the J/Ψ in Fig. 4. The agreement with experimental data from the fixed target experiments (EMC;[48] E516;[49] E401;[50] E687[51]) and from the HERA experiments (ZEUS;[52,53] H1[54–56]) is good with regard to both the magnitude and energy dependence of the cross section. For the J/Ψ and $Q^2=0$ the scanning radius is still large, $r_s \approx 0.4$ fm, and at lower energies $W \simeq 15$ GeV the soft contribution comprises $\sim 50\%$ of the photoproduction amplitude. Still, it is smaller

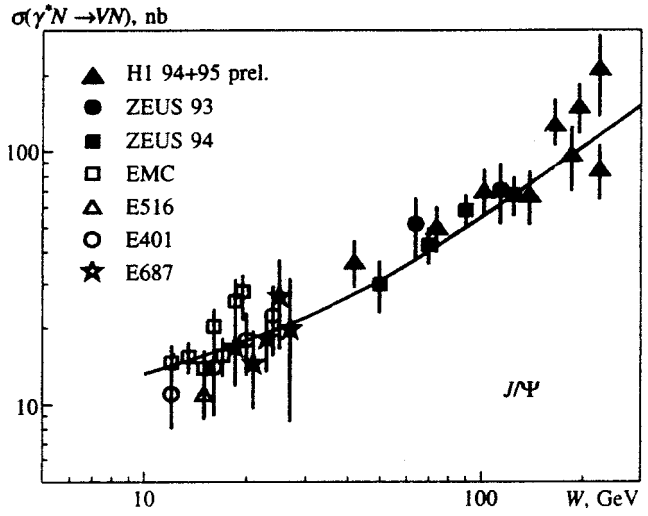


FIG. 4. The color dipole model predictions for the W dependence of the real photoproduction cross section $\sigma(\gamma^* \to V)$ for J/Ψ production vs. the low-energy EMC,[48] E516,[49] E401,[50] E687[51] and high-energy ZEUS[52,53] and H1[54–56] data.

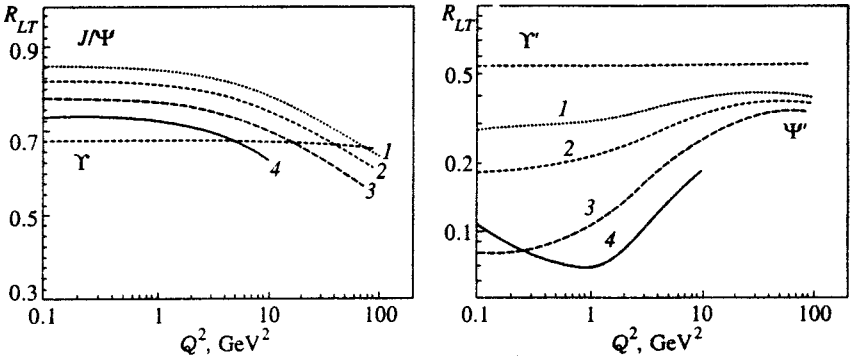


FIG. 5. The color dipole model predictions for the Q^2 and W dependence of the ratio of the longitudinal and transverse differential cross sections in the form given by Eq. (62) for $W=500$ GeV (curve *1*), 150 GeV (curve *2*), 50 GeV (curve *3*), and 15 GeV (curve *4*).

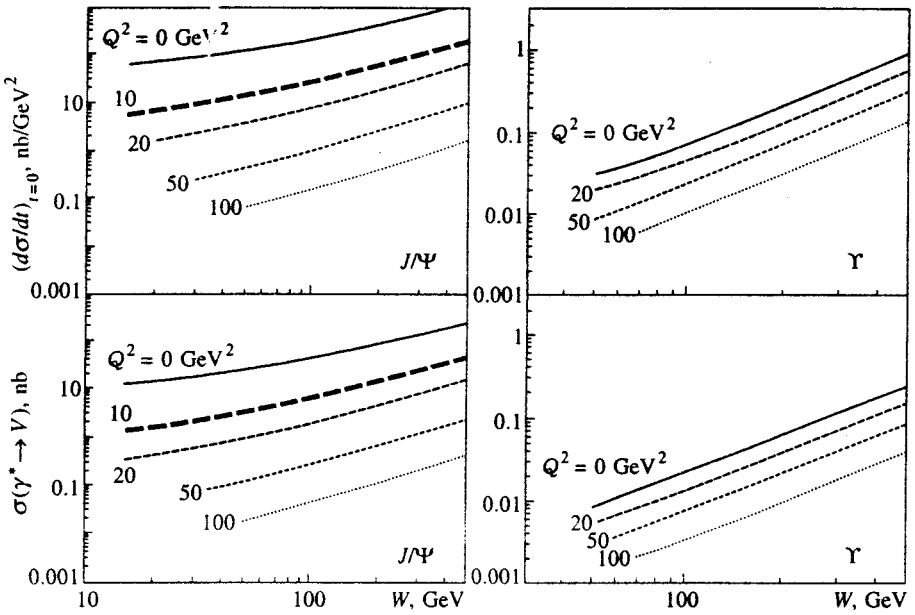


FIG. 6. The color dipole model predictions for the polarization-unseparated forward differential cross section (top boxes) $[d\sigma(\gamma^*\to V)/dt]_{t=0}=[d\sigma_T(\gamma^*\to V)/dt]_{t=0}+[d\sigma_L(\gamma^*\to V)/dt]_{t=0}$ for J/Ψ and Υ production as a function of the c.m.s. energy W at various values of Q^2. The bottom boxes are predictions for the polarization-unseparated t-integrated cross section $\sigma(\gamma^*\to V)=\sigma_T(\gamma^*\to V)+\sigma_L(\gamma^*\to V)$.

than in $\sigma_{\rm tot}(J/\Psi N)$ and $\sigma(\gamma\to J/\Psi)$ rises much faster that $\sigma^2_{\rm tot}(J/\Psi N)$, which is one example of the failure of the vector dominance model for processes with heavy quarkonia. Recall that VDM predicts $\sigma(\gamma\to J/\Psi)\propto\sigma^2_{\rm tot}(J/\Psi N)$.

The relationship (14) (and also (7)) is to a large extent model independent, and predicts the dominance of σ_L at large Q^2. It is convenient to present the results for $R=\sigma_L/\sigma_T$ in the form of the ratio

$$R_{LT}=\frac{m_V^2}{Q^2}\left.\frac{d\sigma_L(\gamma^*\to V)}{d\sigma_T(\gamma^*\to V)}\right|_{t=0} \tag{62}$$

shown in Fig. 5 (m_V is the mass of the vector meson). The point made in Refs. 4 and 5 and repeated in Ref. 6 in somewhat different form is that compared to $\mathscr{M}_L$ the transverse amplitude $\mathscr{M}_T$ receives a larger contribution from large-r asymmetric end-point configurations with $z(1-z)\ll 1$. For this reason $R_{LT}<1$, and it decreases steadily with Q^2. The steeper rise of $\sigma_{\rm pt}(x,r)$ at smaller r makes the end-point contributions less important at higher energies, and R_{LT} rises with energy, although very weakly. The above predictions for $R=d\sigma_L/d\sigma_T$ must be tested at $t=0$, the present experimental data on R are for the t-integrated cross sections. In Ref. 8 it is argued that at large t $\sigma_T\gg\sigma_L$ instead, so that the ratio R measured experimentally for the t-integrated cross sections can be somewhat smaller than our predictions for $t=0$. The calculation of the t dependence of R_{LT} is an interesting subject in its own right, but it lies outside the scope of the present analysis.

In the typical HERA kinematics the longitudinal polarization parameter $\epsilon\approx 1$, and as our prediction for polarization-unseparated production cross section we have

$$\sigma(\gamma^*\to V)=\sigma_T(\gamma^*\to V)+\sigma_L(\gamma^*\to V).$$

In Fig. 6 we show our predictions for J/Ψ and Υ production. The shorthand representation of the same results is in terms of the exponent of the energy dependence of the t-integrated $\sigma(\gamma^*\to V)\propto W^\delta=W^{4\Delta_{\rm eff}}$ and/or $d\sigma/dt|_{t=0}\propto W^\delta=W^{4\Delta_{\rm eff}}$. The exponent δ for the t-integrated cross section is slightly smaller because of the shrinkage of the diffraction cone. The effective intercept $\Delta_{\rm eff}$ depends on the range of W covered by the fit (a more detailed discussion of this issue is found in Ref. 5); in Fig. 7 we present our evaluations for $W=100$ GeV. For the sake of completeness, we show $\Delta_{\rm eff}$ in the same plot for light vector mesons evaluated from cross sections calculated in Ref. 10. Slight departures from exact flavor symmetry are due to slight differences in the pQCD

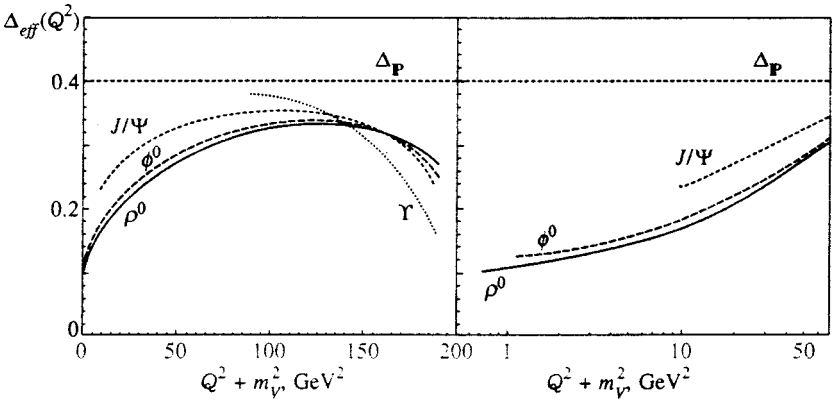


FIG. 7. The Q^2 dependence of the effective intercept $\Delta_{\rm eff}(Q^2)$ for the forward production of ρ^0, ϕ^0, J/Ψ, and Υ at $W=100$ GeV.

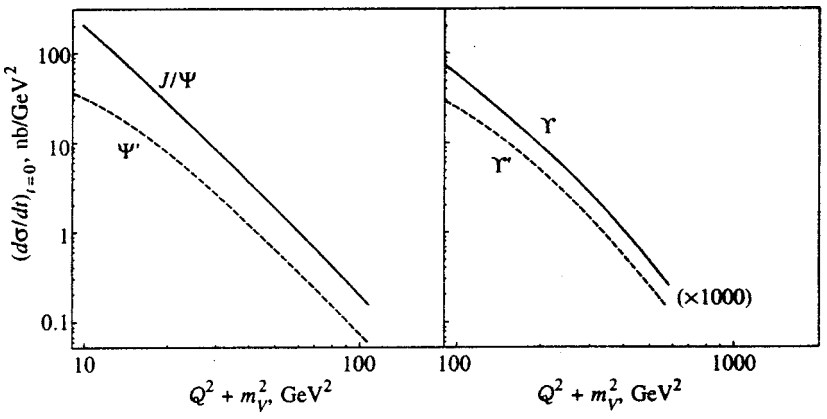


FIG. 8. The color dipole model predictions for the dependence on the scaling variable $Q^2+m_V^2$ of the polarization-unseparated $[d\sigma(\gamma^*\to V)/dt]_{t=0}=[d\sigma_T(\gamma^*\to V)/dt]_{t=0}+[d\sigma_L(\gamma^*\to V)/dt]_{t=0}$ at the HERA energy $W=100$ GeV.

scale factors $\tau(V)$ for different vector mesons. The predicted downward turn of $\Delta_{\rm eff}$ at very large Q^2 is due to the increase in $x_{\rm eff}$ at fixed W.

The real photoproduction of Υ offers one of the best determinations of the intercept $\Delta_{\rm p}$ of the gBFKL pomeron, because in this case one has the magic scanning radius $r_S\sim R_c/2$, and we indeed find $\Delta_{\rm eff}\approx\Delta_{\rm p}=0.4$. The usual fits to the experimental data are of the form $\sigma(\gamma^*p\to Vp)\propto W^{\delta}=W^{4\Delta_{\rm eff}}$. The evaluated value of $\delta\simeq 0.9$ from Fig. 4 in the range $40<W<140$ GeV is in good agreement with the value $\delta=0.92\pm0.14({\rm stat.})\pm0.10({\rm syst.})$ extracted from the data on elastic J/Ψ photoproduction.[53] Analogous estimation of $\delta\sim0.82$ from Fig. 4 in the range $30<W<240$ GeV is in good agreement with the value $\delta=0.77\pm0.13$ presented in Ref. 56. The recent H1 data on elastic virtual photoproduction of J/Ψ[56] reported the values $\delta=1.07\pm0.57$ at $Q^2=3.7$ GeV^2 and 1.22 ± 0.52 at $Q^2=16$ GeV^2 in the energy range $40<W<150$ GeV, which correspond to our results $\delta=0.98$ and $\delta=1.15$, respectively.

The salient features of the Q^2 dependence are best seen when cross sections are plotted as a function of the flavor symmetry restoring variable $Q^2+m_V^2$, whereupon the J/Ψ and Υ production cross sections exhibit very similar behavior (Fig. 8). For $R_{LT}\approx1$, the theory predicts

$$\left.\frac{d\sigma}{dt}\right|_{t=0}\sim\frac{1}{(Q^2+m_V^2)^3}\,G^2(x_{\rm eff},\tau(V)(Q^2+m_V^2)). \qquad (63)$$

If one fits (63) to the behavior $(Q^2+m_V^2)^{-n}$ and neglects the Q^2 dependence coming from the gluon structure function, then $n\approx3$. The effect of the gluon structure function is twofold. At fixed $x_{\rm eff}$, i.e., when the energy varies with Q^2 according to $W^2=(Q^2+m_V^2)/x_{\rm eff}$, the gluon structure function rises with Q^2, which reduces the fitted exponent n: $n\lesssim3$. On the other hand, experimentally one usually studies the Q^2 dependence at fixed energy W, when $x_{\rm eff}=(Q^2+m_V^2)/W^2$ increases with Q^2. Because the gluon structure function decreases toward large x, this induced Q^2 dependence increases the exponent n.

The exponent n depends on the range of Q^2 the fit is performed in. For instance, in J/Ψ production at a typical HERA energy $W=100$ GeV, we predict $n\approx2.8$ for the semiperturbative region of $Q^2\lesssim10$ GeV^2 and $n\approx3.2$ if the fit is performed over the range $15\lesssim Q^2\lesssim100$ GeV^2. We recall that for ρ^0 production we found $n\approx2.4$ for $Q^2\lesssim10$ GeV^2 and $n\approx3.2$ for $15\lesssim Q^2\lesssim100$ GeV^2.[10] The results for the Υ are similar to the large-Q^2 result for the J/Ψ. The departures from exact flavor symmetry due to $R_{LT}\neq1$ and the slight flavor dependence of the pQCD scale $\tau(V)$ are marginal for all practical purposes.

The experimental data on virtual photoproduction of charmonium states are still sparse, and there are as yet no data on the photoproduction of bottomonium. In Fig. 9 we present a summary of the experimental data on the J/Ψ production from the fixed target EMC experiment[48] and the ZEUS[52,57] and H1[54,56,58] experiments at HERA. The theoretical results are for $W=15$ GeV, appropriate for the EMC experiment (dashed curve), and for $W=100$ GeV, appropriate for the HERA experiments; theory and experiment are in reasonable agreement.

One of the outstanding experimental problems at large Q^2 is a separation of elastic reaction $\gamma^*p\to V+p$ from the inelastic background $\gamma^*p\to V+X$; the low-energy EMC data are well known to have been plagued by the inelastic background. The contribution of the inelastic background to the experimental cross section may be the reason why we underestimate the experimental data. One more argument in favor of this point will be presented in the discussion of the diffraction slope below.

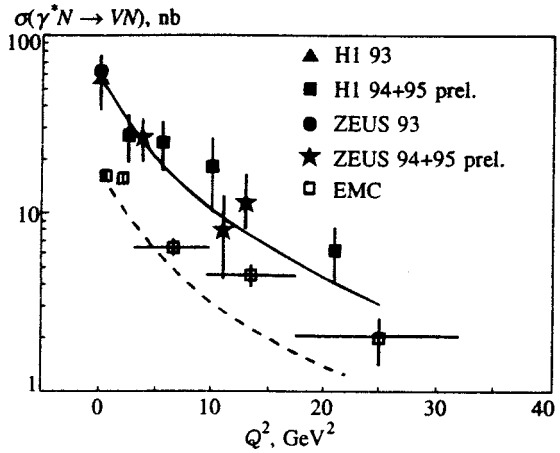


FIG. 9. The color dipole model predictions for the Q^2 dependence of the observed cross section $\sigma(\gamma^*\to V)=\sigma_T(\gamma^*\to V)+\epsilon\sigma_L(\gamma^*\to V)$ of exclusive J/Ψ production vs. the low-energy (EMC[48]) and high-energy (ZEUS,[52,57] H1[54,56,58]) data.

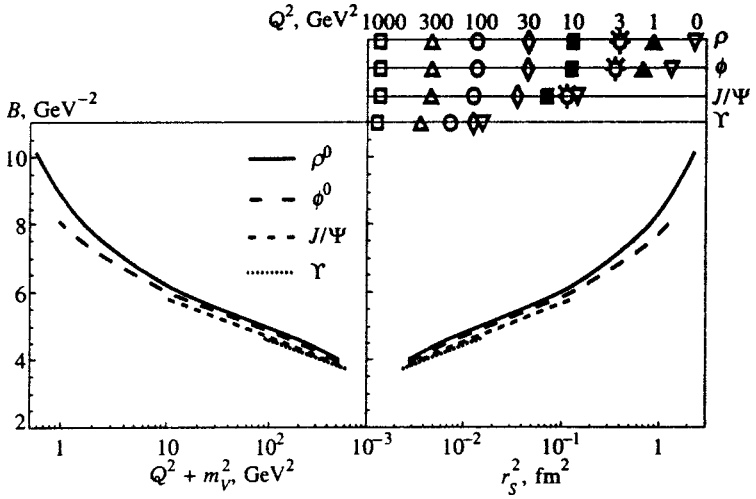


FIG. 10. The color dipole model predictions for the diffraction slope in production of different vector mesons as a function of the scaling variable $Q^2+m_V^2$ (left box) and scanning radius r_S (right box) at fixed c.m.s. energy $W=100$ GeV. The scales of Q^2 on the top of the right box show the values of Q^2 that correspond to the scanning radii shown on the bottom axis.

8. DIFFRACTION CONE FOR THE $V(1S)$ STATES

Evidently, the contribution to the diffraction slope from the $\gamma^*\to V$ transition vertex decreases with decreasing scanning radius r_S, i.e., with rising Q^2.[28] At fixed energy W the value of $x_{\rm eff}$ rises and the rapidity ξ decreases, which also reduces the diffraction slope because the Regge term becomes smaller, which is an important component of the Q^2 dependence at fixed W. In this section we report evaluations of the diffraction slope based on Eq. (28). We use the results of Ref. 28 for the energy and dipole size dependence of $B(\xi,r)$ for gBFKL color dipole amplitude. For the soft pomeron contribution, we use the parametrizations (53) and (55). According to Fig. 3, the nonperturbative contribution to the J/Ψ and Υ production amplitudes is small, and our results for the diffraction slope are insensitive to the soft pomeron effects. Our definition of the diffraction slope is Eq. (60) in Sec. 7, and is meant to correspond to the experimentally measured slope $B(t)$ at $t\simeq 0.1-0.15$ GeV^{-2}.

As shown in Ref. 28, at subasymptotic energies the diffraction slope for the gBFKL amplitude grows rather rapidly, by $\simeq 1.4$ GeV^{-2} when W grows by one order of magnitude from the fixed-target energy $W=15$ GeV to the HERA energy $W=150$ GeV. This corresponds to an effective shrinkage rate $\alpha'_{\rm eff}\approx 0.15$ GeV^{-2}; only at very high energies beyond the HERA range will the shrinkage follow the true slope of the Regge trajectory for the rightmost gBFKL singularity $\alpha'_{\rm I\!P}=0.07$ GeV^{-2}. The values of $\alpha'_{\rm I\!P}$ and $\alpha'_{\rm eff}$ are very sensitive to the gluon propagation radius R_c, and can eventually be used to fix this poorly known parameter. For now, we explore the major properties of the solution for $R_c=0.27$ fm.

One interesting prediction from color dipole phenomenology of a diffraction cone is that the diffraction slope for J/Ψ production at $Q^2=100$ GeV^2 nearly coincides with that for real photoproduction of the Υ. This is still another example of flavor symmetry restoration, because the scanning radii r_S for the two reactions are very similar.

The flavor symmetry properties of the diffraction cone can be seen in Fig. 10 (left box). The curves for $B(\gamma^*\to V)$ of all the vector mesons converge as a function of $Q^2+m_V^2$; slight departures from exact flavor symmetry can be attributed to a difference in the pQCD scale factors $\tau(V)$ for light and heavy vector mesons. At fixed W, the calculated Q^2 dependence is an interplay of the changing scanning radius r_S and the decrease in the Regge component with increasing $x_{\rm eff}$. For the light vector mesons and $Q^2\lesssim 10$ GeV^2, the soft pomeron is substantial, and high-precision experimental data on ρ^0 and ϕ^0 in this range of Q^2 are indispensable to a better understanding of the soft pomeron. In Fig. 10 (right box) the same results are presented as a function of the scanning radius r_S as defined by Eq. (2) with $A=6$. Crude estimates for the Q^2 dependence of $B(\gamma^*\to V)$ reported in Ref. 28 are close to the present results.

We can suggest useful empirical parametrizations for the diffraction slope. For production of heavy quarkonia, $V=J/\Psi$, Υ, the Q^2 dependence of the diffraction slope at $W=100$ GeV and in the considered range of $Q^2\lesssim 500$ GeV^2 can be approximated by

$$B(\gamma^*\to V)\approx\beta_0-\beta_1\,\log\frac{Q^2+m_V^2}{m_{J/\Psi}^2} \tag{64}$$

with slope $\beta_1\approx 1.1$ GeV^{-2} and intercept $\beta_0\approx 5.8$ GeV^{-2}. Although (64) must be regarded as only a purely empirical crude parametrization, the logarithmic term (64) is suggestive of a substantial role of the term (50) in the diffraction slope at high energy. We recall that the constant β_0 is subject to the choice of the t range; it is the slope β_1 that is more closely related to gBFKL dynamics. For the light vector mesons, a somewhat better approximation to the results shown in Fig. 10 is

$$B(\gamma^*\to V)\approx\beta_0-\beta_1\,\log\left(\frac{Q^2+m_V^2}{m_{J/\Psi}^2}\right)+\frac{\beta_2}{Q^2+m_V^2} \tag{65}$$

with the same $\beta_1=1.1$ GeV^{-2} as above and with $\beta_0=7.1$ GeV^{-2}, $\beta_2=1.6$ for ρ^0 production and $\beta_0=7.0$ GeV^{-2}, $\beta_2=1.1$ for ϕ^0 production.

Experimental studies of both the Q^2 and energy dependence of the diffraction slope are in the formative stage. In the heavy quarkonium sector, only photoproduction of the J/Ψ has been studied to any extent. The experimental situa-

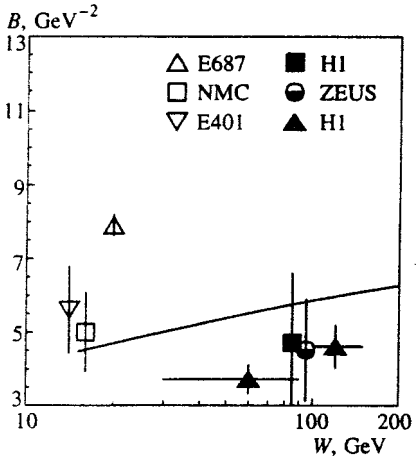


FIG. 11. Comparison of the color dipole model prediction for c.m.s. energy W dependence of the diffraction slope for photoproduction of the J/Ψ with the E401,[50] NMC,[59] E687,[32] H1[54,55] and ZEUS[52,53] data; $Q^2=0$.

tion is summarized in Fig. 11, at both fixed-target[32,50,59] and HERA energies[52–55]; the error bars are too large for definitive conclusions about the presence or lack of shrinkage of the diffraction cone to be drawn. On the experimental side, determinations of the diffraction slope are very sensitive to cancellation of the inelastic background. Only the E401 experiment[50] used a technique that allowed direct selection of purely elastic events. The E401 result $B(W=15\ \text{GeV},\ Q^2=0)=5.6\pm1.2\ \text{GeV}^{-2}$ is consistent with the NMC result $B(W=15\ \text{GeV},\ Q^2=0)=5.0\pm1.1\ \text{GeV}^{-2}$.[59] The recent high-statistics Fermilab E687 experiment[32] used a nuclear target, and its determination of the diffraction slope for quasielastic scattering, $B(W=20\ \text{GeV},\ Q^2=0)=7.99\pm0.23\ \text{GeV}^{-2}$, is subject to the model-dependent separation of the coherent and quasielastic production on nuclei. At HERA, the first H1 data gave $B(W=90\ \text{GeV},\ Q^2=0)=4.7\pm1.9\ \text{GeV}^{-2}$[54] and the first ZEUS data gave $B(W=90\ \text{GeV},\ Q^2=0)=4.5\pm1.4\ \text{GeV}^{-2}$,[52] updated with higher statistics to $B(W=90\ \text{GeV},\ Q^2=0)=4.6\pm0.6\ \text{GeV}^{-2}$.[53] In 1996 the H1 collaboration[55] found weak evidence for shrinkage of the diffraction cone: $B(W\sim60\ \text{GeV},\ Q^2=0)=3.7\pm0.2\pm0.2\ \text{GeV}^{-2}$ and $B(W\sim120\ \text{GeV},\ Q^2=0)=4.6\pm0.3\pm0.3\ \text{GeV}^{-2}$.

For virtual production of J/Ψ the H1 collaboration[58] reported in 1996 the first data:

$$B(W=90\ \text{GeV},\langle Q^2\rangle=18\ \text{GeV}^2)=3.8\pm1.2(\text{stat.})^{+2.0}_{-1.6}(\text{syst.})\ \text{GeV}^{-2}.$$

Recently the ZEUS collaboration[57] published a value of the diffraction slope at $Q^2=6\ \text{GeV}^2$:

$$B(W=90\ \text{GeV},\langle Q^2\rangle=6\ \text{GeV}^2)=4.5\pm0.8(\text{stat.})\pm1.0(\text{syst.})\ \text{GeV}^{-2}.$$

We predict a decrease in the diffraction slope from $Q^2=0$ to $Q^2=18\ \text{GeV}^2$ by a mere $\approx0.5\ \text{GeV}^{-2}$, too small an effect to be seen at the present experimental accuracy.

The end-point contribution from asymmetric large-size dipoles with $z(1-z)\ll1$ is different for the production of the T and L polarized vector mesons, and makes the average scanning radius somewhat larger in the case of T polarization. Consequently, one would expect inequality of the diffraction slopes, $B_T>B_L$, for the polarization states. But the numerical difference between them is negligible even for the charmonium states, because in nonrelativistic quarkonium the end-point effects are strongly suppressed. For bottomonium states, B_T-B_L is absolutely negligible.

9. WHAT IS SPECIAL ABOUT DIFFRACTIVE PRODUCTION OF THE RADIALLY EXCITED STATES $V(2S)$?

In diffractive production of radially excited $2S$ states (Ψ',Υ') the watchword is the node effect. The radial wave function of the $2S$ state changes sign at $r\sim R_V(1S)$, and there are cancellations of contributions to the production amplitude from large dipoles, $r\gtrsim R_V(1S)$, and small dipoles, $r\lesssim R_V(1S)$, which were noted for the first time in Refs. 1 and 3. Manifestations of the node effect for diffractive production of light vector mesons off nuclei have been discussed in Refs. 30 and 60. A detailed analysis of the forward real and virtual photoproduction of light $2S$ states (ρ',ϕ') at high energies is presented in Ref. 10. The major subject of the present discussion is new manifestations of the node for the diffraction cone.

The cancellation pattern obviously depends on the relationship between r_S and the position of the node r_n, which is close to the radius of the $1S$ state, $r_n\sim R_V(1S)$. If $r_S\ll R_V(1S)$, then the wrong-sign contribution to the production amplitudes from dipoles with $r\gtrsim r_n$ is small and cancellations are weak (the undercompensation scenario of Ref. 30). If $r_S\gtrsim R_V(1S)$, then the production amplitude can even be dominated by the wrong-sign contribution from r above the node position (the overcompensation scenario). For heavy quarkonia the scanning radius r_S is sufficiently small even at $Q^2=0$, and the undercompensation scenario is realized.

At fixed target energies, the node effect is sufficiently strong and suppresses the ratio $R_{21}(t=0)=[d\sigma(\Psi')/d\sigma(J/\Psi)]_{t=0}$ by almost an order of magnitude (Fig. 12). Evidently, the smaller the scanning radius, the smaller the large-r contribution and the weaker the node effect, so that the ratio $[d\sigma(\Psi')/d\sigma(J/\Psi)]_{t=0}$ rises with Q^2 as shown in Fig. 12. When the node effect is strong, which is the case for the Ψ' at $Q^2=0$, then even slight variations of the scanning radius r_S can change the strength of the node effect substantially. For this reason one must not be surprised that at fixed target energies the ratio $[d\sigma(\Psi')/d\sigma(J/\Psi)]_{t=0}$ changes with Q^2 quite rapidly, on a scale of Q^2 smaller than the natural scale m_V^2. The predicted energy dependence of $[d\sigma(\Psi')/d\sigma(J/\Psi)]_{t=0}$ derives from the faster growth with energy of the dipole cross section for smaller dipoles, which also diminishes the node effect. In Fig. 13 (left box) we show in more detail for the HERA energy $W=100\ \text{GeV}$ the Q^2 dependence of the ratio of the t-integrated cross sections $\sigma(2S)/\sigma(1S)$, evaluated using the diffraction slope $B(2S)$ described below. The predicted Q^2 and W dependences of

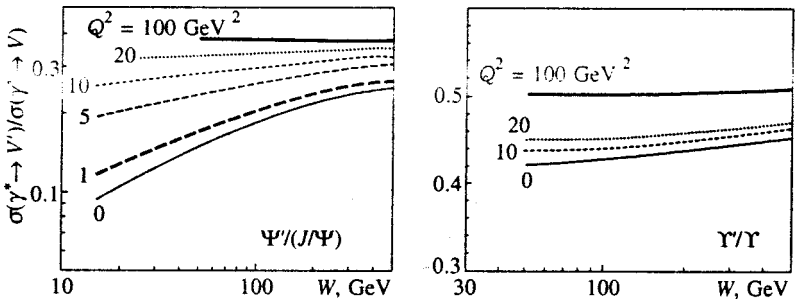


FIG. 12. The color dipole model predictions for the Q^2 and W dependence of the ratios $d\sigma(\gamma^*\to\Psi'(2S))/d\sigma(\gamma^*\to J/\Psi)$ and $d\sigma(\gamma^*\to\Upsilon'(2S))/d\sigma(\gamma^*\to\Upsilon)$ for the polarization-unseparated forward differential cross sections.

the node effect are both sufficiently strong to be observed at HERA. Because the scanning radius is substantially less than $R_V(1S)$ for the heavier b quarks than for the charmed quarks, the node effect in bottomonium production is much weaker, and the ratio $[d\sigma(\Upsilon')/d\sigma(\Upsilon)]_{t=0}$ is larger and exhibits much weaker Q^2 and W dependences than for the charmonium states (Fig. 12).

The node effect is slightly different for the T and L polarizations. This is best seen in Fig. 5, which shows the ratio $R_{LT}(2S)$ for $V'(2S)$ production, which is different from $R_{LT}(1S)$ both in magnitude and Q^2, W dependence.

The new effect that we focus on here is a nontrivial impact of the node effect on the diffraction cone. In the conventional situation, the larger the radii of the participating particles, the larger the diffraction slope, and for real photoproduction we have the clear hierarchy

$$B(\gamma\to\rho^0)>B(\gamma\leftarrow\phi^0)>B(\gamma\to J/\Psi)>B(\gamma\to\Upsilon), \tag{66}$$

which follows the hierarchy of radii $R_{\rho^0}>R_{\phi^0}>R_{J/\Psi}>R_\Upsilon$. Although the rms radius of the Ψ' is about twice $R_{J/\Psi}$, the color dipole approach uniquely predicts $B(\gamma\to\Psi')<B(\gamma\to J/\Psi)$ in striking disagreement with the hierarchy (66), which has the following origin. Let $\mathscr{M}_<$ and $\mathscr{M}_>$ be the moduli of contributions to the $V(2S)$ production amplitude from color dipoles with size r below and above the position of the node in the $2S$ radial wave function, and let $B_<$ and $B_>$ be the diffraction slopes for the corresponding contributions. Because of the hierarchy (66), we have the strong inequality

$$B_>>B_<. \tag{67}$$

For production of the $V(1S)$ state $B(1S)\approx B_<$. Now, the total $V(2S)$ production amplitude is $\mathscr{M}(2S)=\mathscr{M}_<-\mathscr{M}_>$, and for the diffraction slope we find

$$B(2S)=\frac{B_<\mathscr{M}_<-B_>\mathscr{M}_>}{\mathscr{M}_<-\mathscr{M}_>}$$

$$=B_<-(B_>-B_<)\,\frac{\mathscr{M}_>}{\mathscr{M}_<-\mathscr{M}_>}, \tag{68}$$

which yields the estimate

$$B(2S)-B(1S)\approx-(B_>-B_<)\,\frac{\mathscr{M}_>}{\mathscr{M}_<-\mathscr{M}_>}<0. \tag{69}$$

The weaker the node effect, the smaller the difference of diffraction slopes $B(2S)-B(1S)$. The typical color dipole sizes r that enter $M_<$ and $M_>$ differ by $\sim R_V(1S)$, and the difference of slopes $B_>-B_<$ can be evaluated as a variation of the diffraction slope $B(1S)$ when the scanning radius r_S changes by a factor ~ 2 from its value at $Q^2=0$. Then the parametrization yields an estimate $B_>-B_<\sim(1-2)\beta_1\sim 1\ \mathrm{GeV}^{-2}$ for heavier quarkonia. Equation (69) shows that the splitting $B(2S)-B(1S)$ is further suppressed if the node effect is weak, i.e., if $\mathscr{M}_>\ll\mathscr{M}_<$.

The results for $B(1S)-B(2S)$ are presented in Fig. 14. For the bottomonium family the node effect is negligibly weak, cf. Fig. 12, whereas for the charmonium family the chances of the experimental observation of the inequality $B(2S)<B(1S)$ are nonzero, at least in real photoproduction and in the fixed-target experiments. The difference of diffraction slopes $B(1S)-B(2S)$ is larger for L polarization, consistent with a stronger node effect for L polarization. As discussed above and shown in Fig. 12, the node effect diminishes with energy, and the difference of diffraction slopes $B(1S)-B(2S)$ drops by a factor ~ 2 from the fixed-target energy to HERA energy. It vanishes at large Q^2 following the demise of the node effect in Fig. 12; the remarks on the

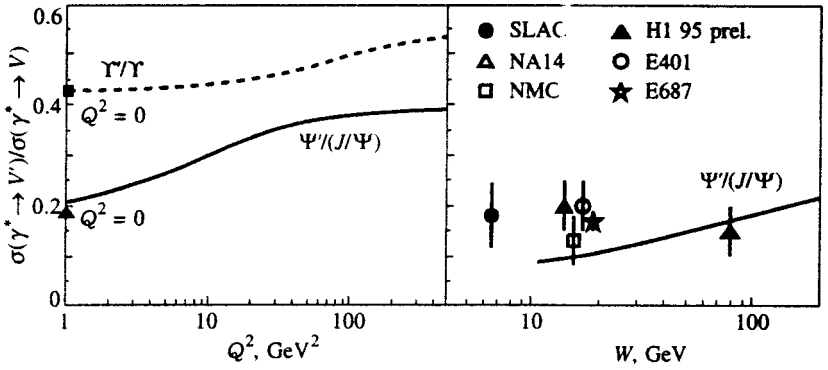


FIG. 13. The left box shows the color dipole model predictions for the Q^2 dependence of the ratio of the t-integrated polarization-unseparated cross sections $\sigma(\gamma^*\to\Psi'(2S))/\sigma(\gamma^*\to J/\Psi)$ and $\sigma(\gamma^*\to\Upsilon'(2S))/\sigma(\gamma^*\to\Upsilon)$ at c.m.s. energy $W=100$ GeV. The right box shows comparison of the color dipole model prediction for c.m.s. energy W dependence of the ratio $\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\Psi)$ at $Q^2=0$ with the E401,[50] NMC,[31] E687,[32] NA14,[62] SLAC[63] and H1[33] data.

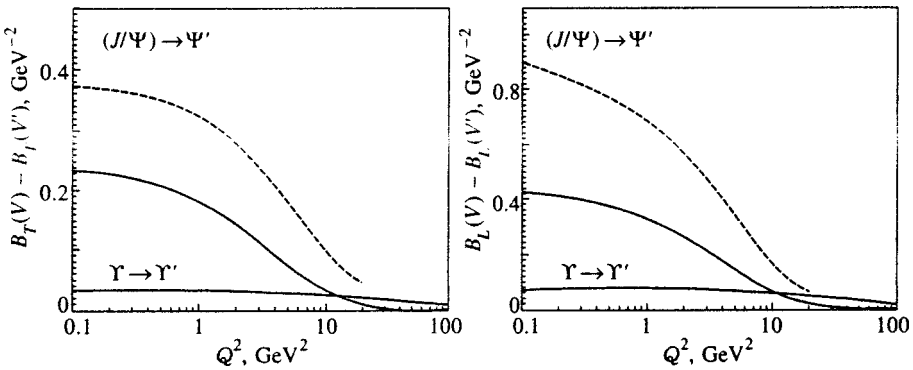


FIG. 14. The color dipole model predictions for the difference of diffraction slopes $B(1S)-B(2S)$ vs. Q^2 at c.m.s. energy $W=15$ GeV (dashed lines) and $W=100$ GeV (solid lines) for T and L polarization.

rapid variation of the node effect on a scale of Q^2 smaller than m_V^2 at fixed-target energies $W\sim 15$ GeV are fully relevant to $B(1S)-B(2S)$ as well.

Another demonstration of the node effect leading to the inequality $B(2S)<B(1S)$ is presented in Fig. 15 in the form of the t dependence of the differential cross section ratio

$$R_{V'/V}(t)=\frac{d\sigma(\gamma\to V')/dt}{d\sigma(\gamma\to V)/dt} \tag{70}$$

for real photoproduction. The ratio $R_{\Psi'/(J/\Psi)}(t)$ rises with t at $W=15$ GeV as a consequence of the node effect, whereas at $W=100$ GeV Fig. 15 shows an essentially constant t dependence of $R_{\Psi'/(J/\Psi)}(t)$ and $R_{\Upsilon'/\Upsilon}(t)$, corresponding to a weaker node effect at higher energy and for heavier vector mesons, respectively; see also Fig. 14.

There is solid experimental evidence for the node effect in real photoproduction of the Ψ'. In 1996 the H1 collaboration reported the first observation of real photoproduction of the Ψ' at HERA, with the result $R_{21}=\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\psi)=0.15\pm 0.05$.[33] In fixed target experiments, only E401 used a hydrogen target, with the result $\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\psi)=0.20\pm 0.05$ at $W=17$ GeV. Nuclear targets were used in all other experiments. Evaluation of the cross section ratio for the hydrogen target from these data requires corrections for nuclear shadowing in J/Ψ production and nuclear antishadowing in Ψ' production; there are also systematic uncertainties in the separation of coherent and incoherent production. Specifically, for the same color dipole model as used in this paper, it has been shown[1] that the ratio $R_{21}=\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\psi)$ is enhanced in incoherent production off nuclei by the factor 1.26, 1.55, and 2.16 for Be, Fe, and Pb nuclei, respectively. For a relatively dilute ^{6}Li target, the enhancement factor can be estimated to be $\approx$1.1. The E687 result $R_{21}(\mathrm{E687})=0.21\pm 0.02$ for a Be target at $W=19$ GeV[32] then corresponds to $R_{21}(\mathrm{E687};N)=0.17\pm 0.02$ for a free nucleon target; the NMC result $R_{21}=0.20\pm 0.05(\mathrm{stat.})\pm 0.07(\mathrm{syst.})$[31] after correction for the last value[61] of the branching ratio $BR(J/\Psi\to\mu^+\mu^-)=5.97\pm 0.25\%$ gives $R_{21}=0.17\pm 0.04(\mathrm{stat.})\pm 0.04(\mathrm{syst.})$ for a passive concrete absorber at $W=15$ GeV, which corresponds to $R_{21}(\mathrm{NMC};N)=0.13\pm 0.05$ for a free nucleon target. The NA14 result 0.22 ± 0.05 for a Li target at $W=14$ GeV[62] corresponds to $R_{21}(\mathrm{NA14};N)=0.2\pm 0.05$ for a free nucleon target, and the SLAC result 0.22 ± 0.08 for a Be target at $W\simeq 6.5$ GeV[63] corresponds to $R_{21}(\mathrm{SLAC};N)=0.18\pm 0.07$ for a free nucleon target.

In Fig. 13 (right box) we compare our prediction for $R_{21}(N)=\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\psi)$ for real photoproduction on protons with the H1 and E401 data for a proton target and the above evaluations of $R_{21}(N)$ from nuclear-target data. The overall agreement between theory and experiment is satisfactory. In view of the steady collection of data at HERA, higher precision fixed-target data on a hydrogen target are highly desirable to check unambiguously the predicted rise of $R_{21}(N)$ with energy.

10. SUMMARY AND CONCLUSIONS

The major focus of this work has been on the forward cone for diffractive real and virtual photoproduction of ground $(1S)$ and radially excited $(2S)$ states of heavy quarkonia in the framework of the color dipole running gBFKL approach. We presented a detailed discussion of the color dipole factorization for diffractive amplitudes, and of the relevant pQCD factorization scales, with strong emphasis on restoration of the flavor symmetry in the variable $Q^2+m_V^2$. We based our analysis on solutions of the gBFKL equations for the dipole cross section[26,5] and for the diffraction slope for the color dipole scattering amplitude.[28] Starting from the same dipole cross section that provides a good

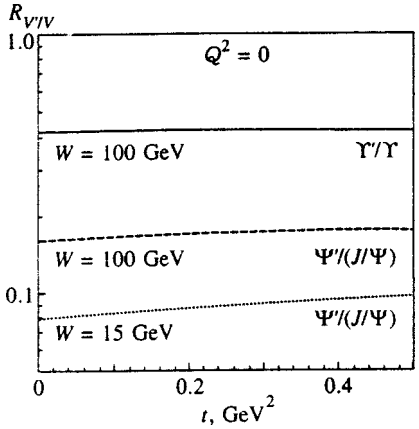


FIG. 15. The color dipole model predictions for t dependence of the ratio differential cross section $R_{V'/V}(t)$ given by (70) for the $\Psi'/(J/\Psi)$ and Υ'/Υ real photoproduction.

quantitative description of the rise in the proton structure function at small x,[26,27] we found encouraging agreement with experimental data on the Q^2 and energy dependence of diffractive J/Ψ production. There are many interesting predictions to be tested for J/Ψ production, for instance the Q^2 dependence of the effective intercept $\Delta_{\rm eff}$.

A detailed analysis of the energy and Q^2 dependence of the diffraction slope $B(\gamma^*\to V)$ for vector meson production is presented here for the first time. Of primary interest is the shrinkage of the diffraction cone, which follows from the finding[28] that the gBFKL pomeron is a set of moving poles. We identified various sources of the Q^2 dependence of the diffraction slope. Based on the solution[28] of the running gBFKL equation for the diffraction slope, we presented detailed calculations of the Q^2 and W dependence of $B(\gamma^*\to V)$. The present experimental data on $B(\gamma^*\to V)$ for J/Ψ production are not yet accurate enough to rule in or rule out our predictions for shrinkage of the diffraction cone.

Diffractive production of radially $2S$ mesons (Ψ',Υ') is a subject on its own. The key new feature of production of the $2S$ states is the node effect, the destructive interference of contributions to the production amplitude from small and large color dipoles because of the node in the radial wave function of $2S$ radial excitations. The resulting strong suppression of the Ψ' photoproduction agrees with the available experimental data.

An interesting prediction to be tested from color dipole dynamics is a rise in the cross section ratio $\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\Psi)$ by a factor of 2 from the CERN–FNAL to HERA energies. A new consequence of the node effect that we discussed in this paper is a counterintuitive inequality of diffraction slopes, $B(\gamma\to\Psi')<B(\gamma\to J/\Psi)$, to be contrasted with the familiar rise in the diffraction slope for elastic scattering processes with increasing radius of the beam and target particles. The scanning phenomenon allows one to control the node effect by varying the scanning radius with Q^2, and we present the corresponding predictions for the Q^2 dependence of the cross section ratio $\sigma(\gamma\to\Psi')/\sigma(\gamma\to J/\Psi)$ and the difference of diffraction slopes $B(\gamma\to\Psi')-B(\gamma\to J/\Psi)$.

The predicted effects for the charmonium family are within reach of modern experiments. The present analysis of diffractive production of heavy mesons provides a useful benchmark for future applications to light vector mesons. The experimental comparison of virtual and real photoproduction of vector mesons will shed light on the transition between soft pomeron exchange, which dominates ρ^0,ω^0,ϕ^0 production at small and moderate Q^2, and gBKFL pomeron exchange at higher Q^2 and/or heavy vector mesons.

This work was partly supported by INTAS Grant 93-0239ext. B.G.Z. and V.R.Z. thank Prof. J. Speth for hospitality at IKP, Jülich, where this work was partly carried out.

[1] B. Z. Kopeliovich and B. G. Zakharov, Phys. Rev. D **44**, 3466 (1991).
[2] M. G. Ryskin, Z. Phys. C **57**, 89 (1993).
[3] N. N. Nikolaev, Comments Nucl. Part. Phys. **21**, 41 (1992); B. Z. Kopeliovich, J. Nemchik, N. N. Nikolaev, and B. G. Zakharov, Phys. Lett. B **309**, 179 (1993).
[4] B. Z. Kopeliovich, J. Nemchik, N. N. Nikolaev, and B. G. Zakharov, Phys. Lett. B **324**, 469 (1994).
[5] J. Nemchik, N. N. Nikolaev, and B. G. Zakharov, Phys. Lett. B **341**, 228 (1994).
[6] S. J. Brodsky, L. L. Frankfurt, J. F. Gunion, A. H. Mueller, and M. Strikman, Phys. Rev. D **50**, 3134 (1994).
[7] J. R. Forshaw and M. G. Ryskin, Phys. Lett. B **375**, 301 (1996).
[8] D. Yu. Ivanov, Phys. Rev. D **53**, 3564 (1996); I. F. Ginzburg and D. Yu. Ivanov, Phys. Rev. D **54**, 5523 (1996).
[9] J. Nemchik, N. N. Nikolaev, E. Predazzi, and B. G. Zakharov, Phys. Lett. B **374**, 199 (1996).
[10] J. Nemchik, N. N. Nikolaev, E. Predazzi, and B. G. Zakharov, Z. Phys. C **75**, 71 (1997).
[11] H. G. Dosch, T. Gousset, G. Kulzinger, and H. J. Pirner, Phys. Rev. D **55**, 2602 (1997).
[12] E. A. Kuraev, L. N. Lipatov, and S. V. Fadin, Zh. Éksp. Teor. Fiz. **71**, 840 (1976) [Sov. Phys. JETP **44**, 443 (1976)]; Zh. Éksp. Teor. Fiz. **72**, 377 (1977) [Sov. Phys. JETP **45**, 199 (1977)]; I. Balitsky and L. N. Lipatov, Yad. Fiz. **28**, 1597, (1978) [Sov. J. Nucl. Phys. **28**, 822 (1978)].
[13] L. N. Lipatov, Zh. Éksp. Teor. Fiz. **90**, 1536 (1986) [Sov. Phys. JETP **63**, 904 (1986)]; L. N. Lipatov, in *Perturbative Quantum Chromodynamics*, A. H. Mueller (ed.), World Scientific (1989), p. 411.
[14] N. Nikolaev and B. G. Zakharov, Zh. Éksp. Teor. Fiz. **105**, 1117 (1994) [JETP **78**, 598 (1994)]; Z. Phys. C **64**, 631 (1994).
[15] N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, JETP Lett. **59**, 6 (1994); Zh. Éksp. Teor. Fiz. **105**, 1498 (1994) [JETP **78**, 806 (1994)]; Phys. Lett. B **328**, 486 (1994).
[16] V. N. Gribov and L. N. Lipatov, Yad. Fiz. **15**, 781 (1972) [Sov. J. Nucl. Phys. **15**, 438 (1972)]; L. N. Lipatov, Yad. Fiz. **20**, 340 (1974) [Sov. J. Nucl. Phys. **20**, 181 (1974)]; Yu. L. Dokshitser, Zh. Éksp. Teor. Fiz. **73**, 1216 (1977) [Sov. Phys. JETP **46**, 641 (1977)]; G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
[17] N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **327**, 157 (1994).
[18] E. M. Levin, Nucl. Phys. B **453**, 303 (1995).
[19] K. D. Anderson, D. A. Ross, and M. G. Sotiropoulos, preprint SHEP-97-09 and UM-TH-97-12 (1997), hep-ph/9705466.
[20] E. Schuryak, Rev. Mod. Phys. **65**, 1 (1993).
[21] E. Gotsman and S. Nussinov, Phys. Rev. D **22**, 624 (1980).
[22] A. B. Migdal, JETP Lett. **46**, 256 (1987).
[23] A. H. Mueller, Nucl. Phys. B **415**, 373 (1994).
[24] N. N. Nikolaev and B. G. Zakharov, Z. Phys. C **49**, 607 (1991); Z. Phys. C **53**, 331 (1992).
[25] V. Barone, M. Genovese, N. N. Nikolaev, E. Predazzi, and B. G. Zakharov, Phys. Lett. B **326**, 161 (1994).
[26] N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **327**, 149 (1994).
[27] N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, JETP Lett. **66**, 138 (1997).
[28] N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, Phys. Lett. B **366**, 337 (1996).
[29] N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, JETP Lett. **60**, 694 (1994).
[30] J. Nemchik, N. N. Nikolaev, and B. G. Zakharov, Phys. Lett. B **339**, 194 (1994).
[31] NMC Collab.: P. Amaudruz, M. Arneodo, A. Arvidson *et al.*, Nucl. Phys. B **371**, 553 (1992).
[32] P. Frabetti, V. S. Paolone, P. M. Yager *et al.*, presented at Int. Europhys. Conf. on HEP, Brussels, July 27–August 2, 1995.
[33] H1 Collab.: C. Adloff, S. Aid, M. Anderson *et al.*, preprint DESY 97-228 (1997), hep-ex/9711012.
[34] V. N. Gribov, Zh. Éksp. Teor. Fiz. **30**, 1306 (1970) [Sov. Phys. JETP **30**, 709 (1970)].
[35] G. Veneziano, Phys. Lett. B **52**, 220 (1974); Nucl. Phys. B **74**, 365 (1974); S. Nussinov, Phys. Rev. Lett. **34**, 1286 (1975); G. F. Chew and C. Rosenzweig, Phys. Rep. **41**, 264 (1978); A. Cappela and J. Tran Thanh Van, Phys. Lett. B **114**, 450 (1982).
[36] P. V. Landshoff and O. Nachtmann, Z. Phys. C **35**, 405 (1987).
[37] A. Donnachie and P. V. Landshoff, Phys. Lett. B **185**, 403 (1987); J. R. Cuddell, Nucl. Phys. B **336**, 1 (1990).
[38] V. N. Gribov and A. A. Migdal, Yad. Fiz. **8**, 1213 (1969) [Sov. J. Nucl. Phys. **8**, 703 (1969)].
[39] N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **332**, 184 (1994).
[40] S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991).
[41] B. G. Zakharov, Yad. Fiz. **49**, 1386 (1989) [Sov. J. Nucl. Phys. **49**, 860 (1989)]; J. F. Gunion and D. F. Soper, Phys. Rev. D **15**, 2617 (1977).

[42] V. Barone, M. Genovese, N. N. Nikolaev *et al.*, Z. Phys. C **58**, 541 (1993).
[43] M. Genovese, N. N. Nikolaev, and B. G. Zakharov, Zh. Éksp. Teor. Fiz. **108**, 1141, 1155 (1995) [JETP **81**, 625, 633 (1995)].
[44] N. N. Nikolaev, W. Schaefer, and B. G. Zakharov, preprint KFA-IKP-TH-1996-06 (1996), hep-ph/9608338.
[45] A. Schiz, L. A. Fajardo, R. Majka *et al.*, Phys. Rev. D **24**, 26 (1981); J. P. Burq, M. Chemarin, M. Chevallier *et al.*, Phys. Lett. B **109**, 111 (1982).
[46] B. Z. Kopeliovich, N. N. Nikolaev, and I. K. Potashnikova, Phys. Rev. D **39**, 769 (1989).
[47] N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **333**, 250 (1994).
[48] EMC Collab.: J. J. Aubert, G. Bassompiere, K. H. Becks *et al.*, Nucl. Phys. B **213**, 1 (1983); EMC Collab.: J. Ashman, B. Badelek, G. Baum *et al.*, Z. Phys. C **39**, 169 (1988).
[49] E516 Collab.: B. H. Denby, V. K. Bharadwai, D. J. Summers *et al.*, Phys. Rev. Lett. **52**, 795 (1984).
[50] E401 Collab.: M. Binkley, C. Bohler, J. Butler *et al.*, Phys. Rev. Lett. **48**, 73 (1982).
[51] E687 Collab.: P. L. Frabetti, V. S. Paolone, P. M. Yager *et al.*, Phys. Lett. B **316**, 197 (1993).
[52] ZEUS Collab.: M. Derrick, D. Krakauer, S. Magill *et al.*, Phys. Lett. B **350**, 120 (1995).
[53] ZEUS Collab.: J. Breitweg, M. Derrick, D. Krakauer *et al.*, Z. Phys. C **75**, 215 (1997).
[54] H1 Collab.: T. Ahmed, S. Aid, V. Andreev *et al.*, Phys. Lett. B **338**, 507 (1994); H1 Collab.: S. Aid, V. Andreev, B. Andrieu *et al.*, Nucl. Phys. B **472**, 3 (1996).
[55] H1 Collab.: S. Aid, V. Andreev, B. Andrieu *et al.*, *Diffractive and Non-Diffractive Photoproduction of J/Ψ at H1*, paper pa02-085, submitted to the 28th Intern. Conf. on High Energy Physics, Warsaw, Poland, 25–31 July 1996.
[56] H1 Collab.: C. Adloff, S. Aid, M. Anderson *et al.*, *Elastic Production of J/Ψ Mesons in Photoproduction and at High Q^2 at HERA*, paper No 242, submitted to the Intern. Europhys. Conf. on High Energy Physics, HEP97, Jerusalem, Israel, August 1997.
[57] ZEUS Collab.: J. Breitweg, M. Derrick, D. Krakauer *et al.*, *Exclusive Vector Meson Production in DIS at HERA*, paper No 639, submitted to the Intern. Europhys. Conf. on High Energy Physics, HEP97, Jerusalem, Israel, August, 1997.
[58] H1 Collab.: S. Aid, V. Andreev, B. Andrieu *et al.*, Nucl. Phys. B **468**, 3 (1996).
[59] NMC Collab.: M. Ameodo, A. Arvidson, B. Badelek *et al.*, Phys. Lett. B **332**, 195 (1994).
[60] O. Benhar, B. G. Zakharov, N. N. Nikolaev *et al.*, Phys. Rev. Lett. **74**, 3565 (1995); O. Benhar, S. Fantoni, N. N. Nikolaev *et al.*, Zh. Éksp. Teor. Fiz. **111**, 769 (1997) [JETP **84**, 421 (1997)].
[61] *Review of Particle Properties*, Phys. Rev. D **50**, 1177 (1994).
[62] NA14 Collab.: R. Barate, P. Bareyre, D. Bloch *et al.*, Z. Phys. C **33**, 505 (1987).
[63] SLAC Collab.: U. Camerini, J. G. Learned, R. Prepost *et al.*, Phys. Rev. Lett. **35**, 483 (1975).

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

Measurement of the antineutrino escape asymmetry with respect to the spin of the decaying neutron

A. P. Serebrov, I. A. Kuznetsov,*) I. V. Stepanenko, A. V. Aldushchenkov, and M. S. Lasakov

B. P. Konstantinov St. Petersburg Nuclear Physics Institute, Russian Academy of Sciences, 188350 Gatchina, Russia

Yu. A. Mostovoĭ

Kurchatov Institute, 123182 Moscow, Russia

B. G. Erozolimskiĭ

Harvard University, Cambridge, Massachusetts 02138, U.S.A.

M. S. Dewey and F. E. Wietfeldt

National Institute of Standards and Technology, Gaithersburg, Maryland 20899, U.S.A.

O. Zimmer and H. Börner

Institut Laue–Langevin, F-38042 Grenoble Cedex 9, France
(Submitted 16 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1963–1978 (June 1998)

The result of a new run of measurements of the antineutrino escape asymmetry with respect to the spin of the decaying neutron is presented. The asymmetry coefficient is $B=0.9821\pm0.0040$, which is consistent with the prediction of the standard model of weak interactions. The prospects for increasing the accuracy of the measurements are discussed.
 [S1063-7761(98)00306-0]

1. INTRODUCTION

The measurement of the antineutrino spin asymmetry coefficient B in neutron β decay is a test for left-right-symmetric models of the weak interaction.[1] The value of B within a simple left-right-symmetric model with light right-handed neutrinos (B_{LR}) should differ from the value of B_{V-A} obtained within the standard $V-A$ variant of the theory:

$$B_{LR}=B_{V-A}(1-2\delta^2-1.21\zeta^2-2.42\delta\zeta), \tag{1}$$

where δ is the ratio between the squares of the masses M_1^2 and M_2^2 for the mass states $W_1=W_L\cos\zeta-W_R\sin\zeta$ and $W_2=W_R\cos\zeta+W_L\sin\zeta$, and ζ is the mixing angle of W_L and W_R.

The value of B expected within the pure $V-A$ theory can be calculated on the basis of the values of the ratio between the axial and vector weak coupling constants λ known from measurements of the neutron lifetime and electron-spin correlation:[2–8]

$$B_{V-A}=2\,\frac{\lambda^2-\lambda}{1+3\lambda^2}. \tag{2}$$

The very weak dependence of B on λ,

$$\partial B/\partial\lambda=0.075, \tag{3}$$

makes such a calculation very accurate, so that although the spread in λ in the experiments in Refs. 2–8 amounts to 10^{-2}, they give $B_{V-A}=0.988$ to within 10^{-3}.

Thus, examining the contribution of right-handed currents to the decay probability is useful to the 10^{-3} level, and with respect to the mass of the right-handed vector boson W_R it is useful to the 540 GeV/c^2 level for a zero mixing angle or to $\zeta\leqslant3\times10^{-2}$ when $\delta=0$. The mass limit for W_R from muon decay is 406 GeV/c^2,[9] and the limit obtained for the mass of the additional vector boson W' from direct collider experiments recently reached the 652 GeV/c^2 level.[10] However, more complicated variants of the theoretical models with heavy right-handed neutrinos, nonidentical coupling constants ($g_R\neq g_L$), and inequivalent left- and right-handed Kobayashi–Maskawa matrices make improving the accuracy of measurements of the antineutrino spin asymmetry in neutron β decay a timely problem.

Of course, achieving the 10^{-3} accuracy level is a very complex experimental task involving measurement of the experimental asymmetry and measurement of the neutron beam polarization to appropriate accuracy. However, it is possible to approach this accuracy level. Recent measurements[11] performed at the St. Petersburg Nuclear Physics Institute (Gatchina) on the polarized cold neutron source of the VVR-M reactor achieved an accuracy of 8×10^{-3}. They were discontinued because of the breakdown of the cold neutron source, and the accuracy of the measurements was limited by the statistics of a 94-h series of measurements.

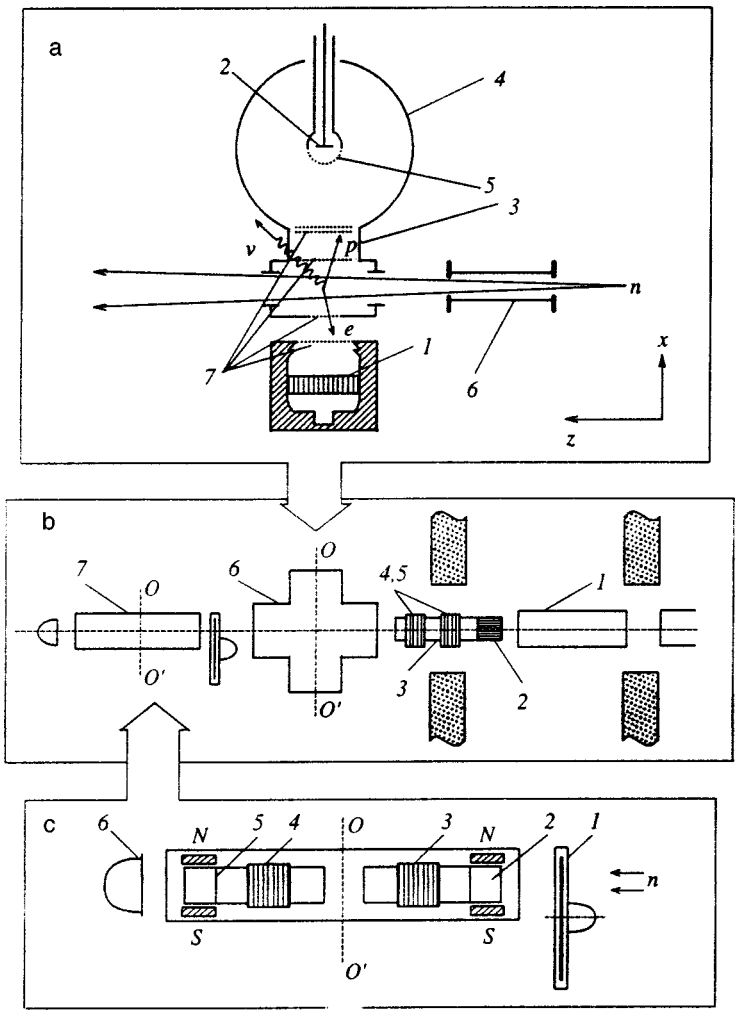


FIG. 1. Overall diagram of the experimental devices and their arrangement in the beam (viewed from above). a) Experimental apparatus for measuring asymmetry: *1* — electron detector, *2* — proton detector, *3* — time-of-flight electrode, *4* — spherical electrode, *5* — spherical grid, *6* — exit neutron guide, *7* — shielding grid. b) Setup of the in-beam experimental equipment: *1* — additional neutron guide section, *2* — polarizer, *3* — neutron guide with nonmagnetic coating, *4* — flipper F_1, *5* — flipper F_1', *6* — experimental apparatus for measuring asymmetry, *7* — apparatus for measuring polarization. c) Apparatus for measuring polarization: *1* — beam shutter, *2* analyzer A_1, *3* — flipper F_2, *4* — flipper F_2', *5* — analyzer A_2, *6* — neutron detector.

In this paper, we present the results of measurements continued at the reactor of the Institut Laue–Langevin (ILL) in Grenoble, where an accuracy of 4×10^{-3} was achieved, and we discuss the prospects for increasing that accuracy.

2. GENERAL EXPERIMENTAL SETUP. POLARIZED BEAM PRODUCTION

The general setup of the experiment performed at the reactor of the Institut Laue–Langevin (Grenoble, France) is shown in Fig. 1. It includes an apparatus for measuring the experimental asymmetry (Fig. 1a) and an apparatus for measuring the neutron beam polarization (Fig. 1c). Measurements of the neutron beam polarization were performed during the measurements of the experimental asymmetry, or the stability of the polarization and the efficiency of the flipper were monitored. The flipper provided for periodic reversal of the sign of the neutron beam polarization.

The first task was to produce a polarized neutron beam of maximum possible intensity. For this purpose, a neutron guide system consisting of an additional neutron guide section, a multislit supermirror polarizer, and a nonmagnetic neutron guide with two radio-frequency flippers was assembled in the PF1 beam from the cold-neutron source of the reactor (see Fig. 1b). Additional neutron guide section *1* with a length of 2 m and cross section 6×12 cm^2 made it possible to bring the polarizer closer to the site of the apparatus. Neutron guide *3* with a nonmagnetic ^{58}NiMo coating ($V_{\rm lim}=7.8$ m/s) is designed to transport the polarized beam to the apparatus and to accommodate the spin flippers. It had a length of 1 m and the same cross section (3×5 cm^2) as multislit supermirror polarizer *2*. The system of the two flippers *4* and *5* permitted reversal of the sign of the neutron beam polarization by either of the flippers and variation of the spin-flip efficiency. The resulting beam of polarized cold neutrons had a flux density of 2×10^8 n/cm$^2\cdot$s at the entrance to the apparatus, which corresponds to a flux density of thermal neutrons equal to 6.8×10^8 n/cm$^2\cdot$s.

Although the resulting flux density of polarized cold neutrons was one-third the density in the VVR-M reactor at Gatchina, the statistical accuracy of the measurements at the ILL reactor in Grenoble was 2.5 times greater, owing to a better signal-to-background ratio, higher degree of polarization of the beam, and successful collection of statistics over the course of 13 days.

3. MEASUREMENT OF NEUTRON BEAM POLARIZATION

The required polarization measurement accuracy in this endeavor is 0.1–0.2%. Determining the properties of the analyzer in measurements of the polarization from the secondary reflection[12] calls for a special approach, which takes into

account a very subtle effect, viz., the depolarization accompanying the interaction of neutrons with analyzing devices. In the general case, an analyzing device can be described by a transmission matrix. The off-diagonal elements of such a matrix correspond to depolarization processes:

$$\hat{A}=\begin{pmatrix} I & B \\ B & C \end{pmatrix}, \quad A=\frac{I-C}{I+2B+C}, \tag{4}$$

$$\hat{P}=\begin{pmatrix} J_1 \\ J_2 \end{pmatrix}, \quad P=\frac{J_1-J_2}{J_1+J_2}, \tag{5}$$

where $\hat{A}$ is the analyzer matrix, A is the analyzing power, I and C are the transmission coefficients of the different spin components of the neutron beam without no change in polarization, B is the transmission coefficient with a change in sign of the polarization, $\hat{P}$ is the polarization vector of the neutron beam, P is the value of the polarization, and J_1 and J_2 are the intensities of the spin components of the neutron beam.

A method for measuring the polarization with consideration of depolarization effects was proposed in our preceding paper (Ref. 13), which was devoted to this question. This method was carefully studied during the measurements performed.

The method is based on a design for an analyzing device consisting of two analyzers with a flipper between them (Fig. 1c) (a double flipper system was used to measure the flipper efficiency). Rotating the analyzing device by 180° about the horizontal axis $O-O'$ enables us to obtain the required number of independent measurements and to determine the beam polarization, as well as the efficiency of both analyzers under the assumption that their efficiency is conserved after the reversal operation.

The neutron beam polarization can be calculated from the formula[13]

$$P_1^2\equiv\frac{N_{-+}\tilde{N}_{-+}}{N_{++}\tilde{N}_{+-}}=P^2\frac{A_1}{\tilde{A}_1}, \tag{6}$$

$$P'^2_1\equiv\frac{N_{-+}\tilde{N}_{-+}}{\tilde{N}_{++}N_{+-}}=P^2\frac{\tilde{A}_2}{A_2}, \tag{7}$$

$$P_A=\sqrt{P_1P'_1}, \tag{8}$$

where N_{++}, N_{+-}, N_{-+}, and N_{--} are linear combinations of the measurement results for the forward position of the analyzing device (A_1,A_2):

$$\begin{aligned} N_{++}&=(N_{00}+N_{10})+(N_{01}+N_{11}),\\ N_{+-}&=(N_{00}+N_{10})-(N_{01}+N_{11}),\\ N_{-+}&=(N_{00}-N_{10})+(N_{01}-N_{11}),\\ N_{--}&=(N_{00}-N_{10})-(N_{01}-N_{11}). \end{aligned} \tag{9}$$

Similar expressions can be written for the linear combinations $\tilde{N}_{++}$, $\tilde{N}_{+-}$, $\tilde{N}_{-+}$, and $\tilde{N}_{--}$ for the backward position of the analyzing device (A_2,A_1). Here N_{00}, N_{10}, N_{01}, N_{11}, $\tilde{N}_{00}$, $\tilde{N}_{10}$, $\tilde{N}_{01}$, and $\tilde{N}_{11}$ are the detector counting rates for different states of the flippers in both the forward and backward positions of the analyzing device. The tilde corresponds to the backward position. The indices on N reflect the states of flipper F_1 and flipper F_2, respectively. Formulas (9) are written in the simplified form for flippers with an efficiency equal to unity. In the general case, the flipper efficiencies appear as correction factors in (9) without altering their structure.

The condition of equality of the analyzer efficiencies after reversal of the analyzing device is reflected in (6) and (7). The physical properties of the analyzers remain absolutely unchanged, but their analyzing power depends on the angle of incidence of the neutron beam. Therefore, in the present problem the necessary condition is satisfied if the beam divergence remains unchanged along the entire analyzing device, i.e., when the analyzers and the neutron guide form a system of parallel planes without breaks and bends. Unfortunately, to obtain high efficiency, the analyzer assemblies must be bent to avoid direct flight without interaction with the mirrors. Bent systems quickly alter the beam divergence and cause a spurious effect in the measurements. The spurious effect appearing in a polarization measurement has the form of sinusoidal dependence on the inclination of the device relative to the beam.

This effect was detected in the experiment and confirmed in Monte Carlo calculations. The geometry of the apparatus and the previously measured dependences of the reflection coefficients of the supermirrors for both spin components of the neutron beams were incorporated into the calculations. The calculated and experimental results are presented in Fig. 2. Spurious variation of the neutron beam polarization is observed. The effect has a sinusoidal character with an amplitude of 0.5%. When the apparatus is rotated about the beam axis (see Fig. 2a), the $P(\varphi)$ curve takes the form of its mirror image. This attests to the appearance of the spurious dependence under consideration as a consequence of the asymmetric geometry of the apparatus. Fortunately, the spurious effect has an alternating sign, and is compensated to a considerable extent when the mean value of the beam polarization is measured, i.e., when integration is carried out over the angles. As the Monte Carlo calculations show, the compensation is not complete. In the integrated value of the polarization the spurious effect causes underestimation of the mean value by 0.1%; therefore, an appropriate correction must be introduced. In our measurements the statistical accuracy for the mean beam polarization was very high (0.01%). However, we estimate the final accuracy of the polarization measurements to be at the 0.25% level, fearing that the asymmetric angular distribution of the beam can cause a deviation from the results obtained in the Monte Carlo model.

We noted during these investigations that depolarization takes place when the neutron beam interacts with the analyzing devices. The off-diagonal elements of the matrix describing the analyzing devices were equal to 2.5×10^{-3}. This is the spin-flip probability. This finding is important, since the assumption that there are no depolarization effects and the use of a simplified measurement scheme (without rotation of the analyzing device about axis *1*) would lead to a 0.5% error in the polarization measurements. One possible reason for

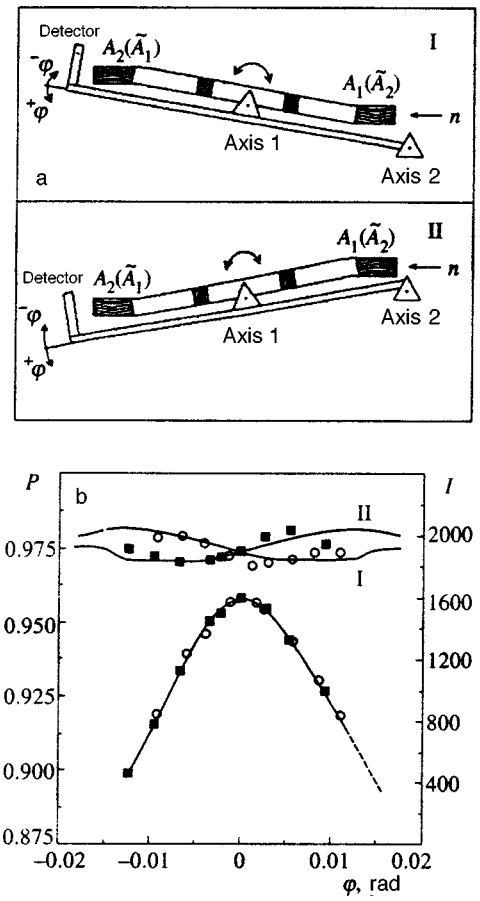


FIG. 2. a) Polarization measurement schemes in two geometries: normal (I) and mirror-image (II) (lateral view). In each of the geometries measurements were performed for the forward position of the analyzing device (A_1,A_2) and for the backward position of the analyzing device $(\tilde{A}_2,\tilde{A}_1)$, The position was changed from forward to backward and vice versa by rotation about axis *1*. The angular dependence of the polarization was measured by rotating the entire device about axis *2*. b) Dependence of the polarization $P(\varphi)$ and the intensity I on the angle of inclination of the analyzing device relative to the beam. The polarization was calculated from Eqs. (9)–(12) under the assumption that the analyzing power of the analyzers remains unchanged after rotation of the device; $A_1=\tilde{A}_1$, $A_2=\tilde{A}_2$. Points — results of measurements in geometry I (○) and geometry II (■). Solid lines — Monte Carlo calculation for these geometries.

the depolarization accompanying the interaction with an analyzing device is the inhomogeneity of the magnetic field. If the magnetic field is not parallel to the supermirror surface, an abrupt change in the direction of the force lines occurs at the vacuum–material boundary because of magnetic induction in the material. For example, if the angle between the magnetic field and the surface is 4°, the beam depolarization effect equals 0.5%.

In measurements of the mean polarization, angular scans were carried out both in the vertical plane and with the analyzer slit inclined to the vertical. The horizontal dimensions of the neutron guide system of the analyzing device were sufficient to completely cover the angular distribution of the beam in the horizontal plane.

The spectral dependence of the polarization P on neutron wavelength λ_n was measured at each point using a time-of-flight technique. This dependence was then averaged with the spectrum $J(\lambda_n)$. When the neutron spectrum $J(\lambda_n)$ was measured, the analyzing system was oriented in the vertical position, and the beam passed through the gap between the neutron guides in the region of the rotation axis of the analyzing system without interacting with the latter. The spectrum was measured by a detector with an efficiency which depends on the velocity according to a $1/v$ law. The changes in the form of the spectrum as a result of the spectrally dependent absorption and scattering in the air, in the aluminum exit window of the chamber, and in the aluminum entrance window of the detector, were measured experimentally. For this purpose, aluminum foils were placed in the beam, and relative measurements of the spectrum were made when the chamber of the apparatus was filled with air and when it was evacuated. The measured spectral corrections were used to correct the result obtained by measuring the mean polarization. This correction was insignificant because of the weak spectral dependence of the polarization. The correction for the counting errors of the electronic instrumentation was also measured experimentally. The results of the measurements of the polarization and the corrections are presented in Table I.

TABLE I.

1.	Results of polarization measurements with consideration of the spectral, angular, and spatial distributions of the beam	97.5±0.01%
2.	Correction for systematic measurement errors	0.10±0.25%
3.	Spectral correction for absorption and scattering in foils and in air	−0.13±0.01%
4.	Correction for counting errors in the electronic instrumentation	0.05±0.02%
	Final result	97.52±0.25%

The most important elements in the setup for measuring the neutron beam polarization are flipper F_1, which is intended to reverse the sign of the polarization, and flipper F_2, which is located between the analyzers.

The operation of the flippers was based on the occurrence of neutron spin flip in a variable magnetic field when the neutron spin precession frequency coincides with the frequency of the variable magnetic field (neutron magnetic resonance). However, the probability of transitions between the Zeeman sublevels depends on the dwell time of the neutron in the field, i.e., such a spin-flip method is spectrally dependent. This problem has been eliminated in our flipper. The magnetic field gradient created in the region of the radio-frequency coil ensures a spin flip for all neutron velocities in the beam when the amplitude of the variable field is large and the following adiabaticity condition holds:

$$V_z \frac{dH}{dz} \ll 2\pi\gamma H_1^2, \tag{10}$$

where V_z is the velocity of neutrons along the beam axis, H is a constant magnetic field with gradient dH/dz, H_1 is the amplitude of the variable magnetic field, and γ is the gyromagnetic ratio of the neutron. If this condition is satisfied for the fastest neutrons in the beam, the flipper efficiency can reach 99%. Such a flipper was first employed for ultracold

TABLE II.

Geometry	f_1, %	f_1', %	f_2, %	f_2', %
Forward (A_1,A_2)	98.01±0.05	96.82±0.05	98.30±0.05	99.51±0.05
Backward (A_2,A_1)	97.96±0.05	96.75±0.05	98.43±0.05	99.39±0.05

neutrons in Ref. 14, and was subsequently employed for heavy and cold neutrons. The PF1 beam was recently outfitted with such flippers. The important advantages include the absence of matter in the beam path and the simplicity of control (by turning the variable field on and off).

A polarization analyzer with a double flipper system makes it possible to easily measure the efficiency of a flipper. Therefore, besides flippers F_1 and F_2 there are auxiliary flippers F_1' and F_2' for measuring the efficiency of the flippers. For each pair of flippers it is possible to perform four different measurements (four combinations of the states of two flippers), which suffice to determine the efficiency of both flippers. The efficiency of the flippers is determined from the relations

$$f=\frac{N_0^0+N_1^1-3N_1^0+N_0^1}{3N_0^0-N_1^1-N_1^0-N_0^1}, \quad f=\frac{N_0^0+N_1^1-3N_0^1+N_1^0}{3N_0^0-N_1^1-N_0^1-N_1^0}. \tag{11}$$

Here N_0^0, N_0^1, N_1^1, and N_1^0 are the detector counting rates, where the subscripts correspond to the main flipper, and the superscripts correspond to the auxiliary flipper. The results of the measurements of the flipper efficiencies are presented in Table II. The measurements of the efficiencies f_2 and f_2' were performed for two positions of the analyzing device (the forward and backward geometries), with no changes noted in efficiency. The values of f_1 and f_2 were obtained automatically. As is seen from Table II, there is fairly good reproducibility of the measurement results.

The stability of the operation of the flippers was regularly monitored during the measurements, and no variations in their efficiency were noted to within 0.05%.

It is noteworthy that a radio-frequency flipper is asymmetric, since the beam polarization is $+P$ when the variable magnetic field is turned off and $-Pf$ when the variable magnetic field is turned on.

4. MEASUREMENT OF THE EXPERIMENTAL ASYMMETRY

The escape direction of the undetected antineutrino can be determined from the measured proton momentum and electron energy. A diagram of the momenta of the neutron decay products is presented in Fig. 3. For given electron momentum, all possible antineutrino momenta lie on a sphere of radius $P_\nu=(E_0-E_e)/c$, where P_ν is the antineutrino momentum, E_0 the kinetic energy of the decay, and E_e the electron energy. Thus, knowing the electron energy and measuring the projection of the proton momentum onto the x axis by the time-of-flight method, we can determine the antineutrino escape angle and reconstruct the kinematics of the decay process. Since the ratios between P_e and P_ν for different electron momenta are different, the events detected in an experiment should be recorded in the form of

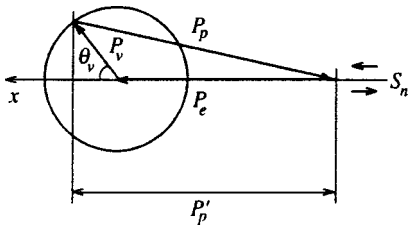


FIG. 3. Diagram of the momenta of the neutron decay products: P_e, P_ν, and P_p are the momenta of the electron, antineutrino, and recoil proton, and P_p' is the projection of the proton momentum onto the x axis.

coincidence matrices with electron energy E_e and proton time of flight t_p as coordinates. Complete information on the neutron decay process can be represented by two matrices corresponding to two directions of the neutron beam polarization relative to the x axis of the apparatus, i.e., the axis joining the two detectors.

The number of events in each cell of such a matrix can be written in the form

$$N_{ik}^{\pm}=f(E_i)[1+a(v_i/c)(\cos\theta_{e\nu})_{ik} \pm PA(v_i/c)(\cos\theta_{\sigma e})_{ik}\pm PB(\cos\theta_{\sigma\nu})_{ik}], \tag{12}$$

where a, A, and B are the electron–antineutrino, electron–spin, and antineutrino–spin angular correlation coefficients, P is the neutron beam polarization, v_i/c is the electron velocity expressed in units of the velocity of light, $f(E_i)$ is the Fermi function with allowance for radiative corrections, and the $\cos\theta_{ik}$ are the cosines of the angles between the respective vectors. The labels i and k specify finite intervals of electron energies (E_i) and proton times of flight (t_k). Accordingly, all quantities with these labels should be averaged over the respective interval.

In the absence of correlations the density of events on the neutrino sphere (see Fig. 3) is constant. Owing to the coefficient a, asymmetry appears between the event counts in the left- and right-hand hemispheres, even for an unpolarized beam. The coefficient A specifies the asymmetry in the total number of events on the sphere when the direction of the neutron spin is reversed. Finally, the coefficient B leads to asymmetry between event counts in the right- and left-hand hemispheres, which changes sign when the sign of the polarization changes. Thus, it is difficult to directly measure the B asymmetry. However, the experimental asymmetry X, which depends on all three coefficients a, A, and B simultaneously, can be measured. It is defined as the asymmetry in the number of events (coincidence of the signals in the electron and proton detectors) upon reversal of the sign of the polarization:

$$X_{ik}=\frac{N_{ik}^+-N_{ik}^-}{N_{ik}^++N_{ik}^-}=\frac{PB(\cos\theta_{\sigma\nu})_{ik}+PA(v_i/c)(\cos\theta_{\sigma e})_{ik}}{1+a(v_i/c)(\cos\theta_{\sigma\nu})_{ik}}. \tag{13}$$

Thus, if we measure the experimental asymmetry X_{ik}, we can calculate $(PB)_{ik}$ using the relation

$$(PB)_{ik}=\frac{X_{ik}[1+a(v_i/c)(\cos\theta_{e\nu})_{ik}]-PA(v_i/c)(\cos\theta_{\sigma e})_{ik}}{(\cos\theta_{\sigma\nu})_{ik}}. \quad (14)$$

Equations (13) and (14) are valid if the absolute values of the polarization during measurements of N^+ and N^- are identical. Consideration of flipper asymmetry would require making these formulas more complicated. On the other hand, the sum of two measurements with guiding magnetic fields of opposite sign permits symmetrization of the values of the polarization for N^+ and N^- and the use of Eqs. (13) and (14). The effective value of the polarization then equals $P_{\text{eff}}=P(1+f_1)/2$.

Determination of B requires knowledge of the neutron beam polarization and calculation of the values of $(v_i/c)\times(\cos\theta_{e\nu})_{ik}$, $(v_i/c)(\cos\theta_{\sigma e})_{ik}$, and $(\cos\theta_{\sigma\nu})_{ik}$. The values of a and A are known from the preceding experiments,[15,6–8] and since they are small compared with B, their error does not make a significant contribution to the total error in the measurement of B.

The equipment diagram is shown in Fig. 1a. The apparatus consists of an electron detector, a system of electrodes, which creates the required configuration of electrostatic fields, and a proton detector. The method based on the coincidence of the electron and proton signals was employed to measure the proton momentum by the time-of-flight technique. The recoil proton passed through the free-flight pedestal *3*, where there is no electric field, and was then accelerated in the field of the spherical capacitor *4*. A proton accelerated to an energy of 25 keV was detected by the detector. The acceleration time, which is appreciably shorter than the time of flight in region *3*, was taken into account in simulating the process of recording decay events.

The signals from the electron detector acted as ''start'' commands for the temporal encoder and permitted measurement of the electron energy. Pulses from the proton signal served as the ''stop'' command for the temporal encoder, providing for time measurements. The background of accidental coincidences was measured simultaneously by the shifted coincidence technique using the same electronics. The information obtained was accumulated in the form of valid and random (background) coincidence matrices for the two spin directions in computer memory for subsequent offline processing.

The electron detector consisted of a 75-mm diameter plastic scintillator with a photomultiplier. The energy resolution and response function of the electron detector were determined in a separate experiment. A magnetic β spectrometer with a working detector placed at its exit was employed for this purpose. The data from these measurements were used to determine the width of the response function of the electron detector in different parts of the energy range. For example, the full width of the line at half-maximum is equal to 81 keV at an energy of 357 keV and to 108 keV at 616 keV. This is in good agreement with the known dependence for scintillation detectors: $\Delta E\propto\sqrt{E}$, i.e., the energy resolution is proportional to the square root of the energy. As a result, the value $\Delta E=0.227\sqrt{E}$ keV was taken for the energy resolution of the detector in the calculations. The ''tail'' of electrons backscattered from the detector then amounts to 6%.

A diaphragm was positioned in front of the electron detector to determine the solid angle for detecting electrons and to isolate the decay region detected, as was a collimator of special design, which suppressed the scattering of electrons that did not enter the detector.

The proton detector was a striped assembly of two microchannel plates. The diameter of the working region of the detector was 60 mm. The temporal resolution of the proton detector and the electronics was determined directly during the experiment from the form of the instantaneous coincidence peak. The width of the instantaneous coincidence peak was 15 ns. The main contribution to the instantaneous coincidence peak was made by the background processes from cascade γ quanta and rescattered electrons.

The focusing of the protons onto the sensitive region of the detector was verified by a calculation and experimentally. Calculations of a map of the electric field followed by calculations of the proton trajectories and the acceleration dynamics were performed on a computer. It was shown that alignment of the detector in the working position relative to the electrodes to within 1 cm ensures the collection of all protons to within 0.1%. This was confirmed in direct experiments on decay protons. For this purpose the sensitive region of the detector was covered by a screen with a diameter of 50 mm so that only the peripheral region with a width of 5 mm would remain open. The count of decay protons in such a detector amounted to 0.1% of the count in a completely open detector.

The entire chamber was surrounded by three pairs of current-carrying loops to create the guiding magnetic field and to cancel the earth's magnetic field. The amplitude of the guiding magnetic field was 5 Oe, and the earth's magnetic field was cancelled to better than 0.02 Oe. The guiding magnetic field provided for polarization parallel (antiparallel) to the axis of the apparatus. The initial magnetic field near the polarizer was vertical, but in front of the apparatus it was rotated by $+90°$ $(-90°)$, i.e., parallel (antiparallel) to the axis of the apparatus. The sign of the polarization was varied during the asymmetry measurements by flipper F_1 without altering the configuration of the magnetic fields. However, about once a day the direction of the guiding magnetic field of the apparatus was reversed to eliminate the effect of the asymmetry in the operation of the flipper.

5. MEASUREMENT RESULTS

The counting rate of the decay events was 0.6 events/s, and 719 456 decay events were detected during the 13-day measurement period. The signal-to-background ratio under the time-of-flight coincidence peak was 2.7. The same ratio at the maximum was equal to 15.

Figure 4 presents experimental proton time-of-flight spectra and electron energy spectra for one of the series of measurements.

In order to calculate the mean values of $(v_i/c)\times(\cos\theta_{e\nu})_{ik}$, $(v_i/c)(\cos\theta_{\sigma e})_{ik}$, and $(\cos\theta_{\sigma\nu})_{ik}$ appearing in (5), we wrote a program to simulate the β decay process

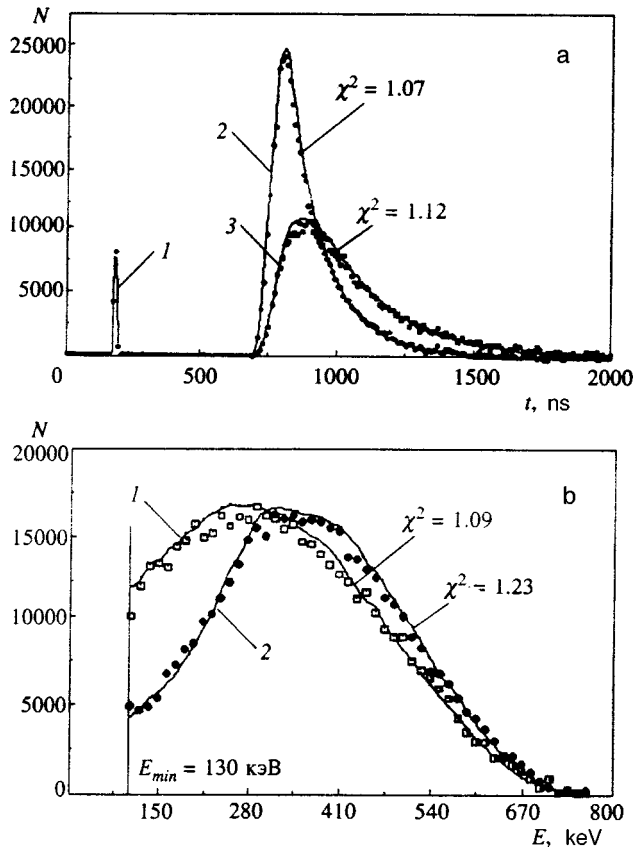


FIG. 4. Experimental and calculated spectra for two beam polarization directions. a) Proton time-of-flight spectra (the channel width is 10 ns): *1* — correlated peak (the intensity is diminished by a factor of 10), *2* — polarization directed toward the electron detector, *3* — polarization directed toward the proton detector. b) Electron energy spectra (the channel width is 13 keV): *1* — polarization directed toward the electron detector, *2* — polarization directed toward the proton detector. Solid lines — Monte Carlo calculation, points — experimental data.

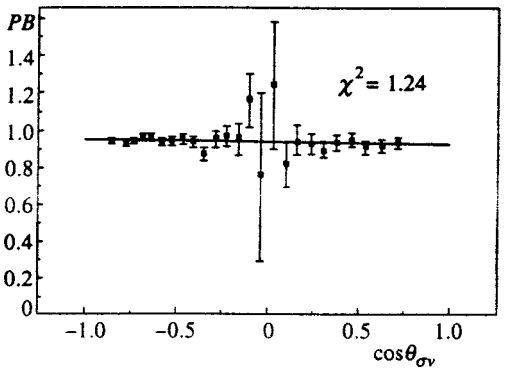


FIG. 5. Dependence of *PB* on $\cos\theta_{\sigma\nu}$ for one of the series of measurements.

under the real conditions of the experimental apparatus by the Monte Carlo method. The model took into account all the necessary geometric parameters, the spectrum of decay electrons in the form of a Fermi function, the response functions of the electron and proton detectors, the characteristics of the analog-to-digital converter (ADC) and the time-to-code converter, the neutron density distribution in the beam, and the calculated map of the electric field between the electrodes.

The program yields four matrices with the same coordinates i and k as the experimental coincidence matrices. Three of these matrices are matrices of corresponding cosines, and the fourth matrix is the calculated two-dimensional spectrum for an unpolarized beam N_{ik}. The subtlest procedure for processing the results is to match the matrices obtained from the experiment and the cosine matrix calculated by the Monte Carlo method. A shift in the time and energy scales can lead to appreciable errors in the calculations, especially near zeroes of $\cos\theta_{\sigma\nu}$. The inaccuracy in the calibration of the energy and time scales and their possible nonlinearities also cause difficulties in accurately matching the matrices. Therefore, we employed the following simple and effective procedure. The experimental time-of-flight spectrum summed over both polarizations and the spectrum calculated by the Monte Carlo method for the unpolarized beam were divided into a certain number of parts with equal intensities. Each such part corresponded uniquely to certain mean values of the calculated cosines. Matching the fits of the corresponding parts of the spectra from the experiment and the calculation permitted us to solve for the accuracy of the calibration and the relative displacement of the time scales.

The energy scale was calibrated using ^{113}Sn and ^{137}Cs conversion electron sources. The calibration procedure was performed during the measurements every 24 h. However, the calibration accuracy was not very high because of the diffuseness of the peaks, which is associated with the resolution of the detector. This required a series of calculations with varying energy parameters, i.e., the channel width, and adjustment of the zero within the calibration uncertainty range. Agreement between the experimental and calculated energy and time spectra served as a preliminary criterion for selecting the best variant.

The final criterion of the accuracy of the matching of the energy and time scales was the lack of a dependence of *PB* on any of the variables. After removing the results for identical values of $\cos\theta_{\sigma\nu}$ from the event matrix, we constructed plots of the dependence of *PB* on $\cos\theta_{\sigma\nu}$. One example of such an analysis for the constancy of *PB* using the χ^2 criterion is presented in Fig. 5. It was found that the criterion for the constancy of *PB* near zeroes of $\cos\theta_{\sigma\nu}$ is fairly sensitive to the calibration accuracy of the energy scale and essentially replaces the direct calibration relative to the energy of the conversion sources. The results of a comparison of the measurements and the Monte Carlo calculations are presented in Fig. 4, which shows projections of the density of the event matrices onto the time and energy axes. The proton time-of-flight spectra most graphically demonstrate how the experimental asymmetry appears when the sign of the beam polarization is reversed. The values of χ^2 indicate the degree of agreement between the calculations and experimental results.

For final processing on the basis of the energy calibrations, all the statistics compiled were divided into three groups. In each of them the corresponding experimental matrices, which were compiled for opposite directions of the guiding magnetic field, were added. The calculation of *PB* was performed on the basis of these summed matrices for each group. The results of the calculations are presented in

TABLE III.

Group No.	Number of events	*PB*	Error
1	219650	0.9485	0.0046
2	347785	0.9442	0.0037
3	152021	0.9479	0.0061

Table III. The weighted mean of *PB* for the three groups is 0.9463 ± 0.0025.

Possible systematic experimental errors were analyzed by varying the model parameters in the Monte Carlo calculations within their uncertainty ranges. Table IV lists the possible errors.

The largest errors are the statistical uncertainty of 0.0025, the 0.0020 error due to the poor energy resolution of the detector, and the 0.0025 error due to the systematic errors in the measurements of the neutron beam polarization. The total uncertainty in the measurements of the antineutrino spin asymmetry was 0.0044, i.e., the result is approximately two times better than the result of the first run of measurements performed in 1994. The results obtained for $P_{\text{eff}}B = 0.9463 \pm 0.0025$, a neutron beam polarization $P = 97.52 \pm 0.25\%$, and a flipper efficiency equal to $97.99 \pm 0.05\%$ were used to calculate the antineutrino spin asymmetry coefficient with consideration of the fact that the effective polarization $P_{\text{eff}} = P(1+f_1)/2$.

Thus, the value obtained in the present experiment for the antineutrino escape asymmetry coefficient with respect to the spin of the decaying neutron is $B = 0.9801 \pm 0.0046$. The weighted mean with consideration of the result obtained in 1994 ($B = 0.9894 \pm 0.0083$) is $B = 0.9821 \pm 0.0040$.

The difference between the value obtained and the value expected according to the standard model of the weak interaction is 0.0059 ± 0.0040, i.e., it is 1.5 times the error. It is noteworthy that such accuracy, 0.4%, in measurements of the angular correlation coefficients in neutron decay was obtained for the first time and approaches the accuracy of the neutron lifetime determination.

Within the left-right-symmetric models, the bound on the parameters of the right-handed currents can be established in accordance with Eq. (1):

$$2\delta^2 + 1.21\zeta^2 + 2.42\delta\zeta \leqslant 0.0127 \ (90\%\text{C.L.}). \qquad (15)$$

TABLE IV.

Source of error	Error	Error in *B*
Accuracy of the calculation of mean cosines (accuracy of the Monte Carlo model)	–	0.0010
Energy resolution of the electron detector, keV	1.7	0.0020
Fraction of backscattered electrons	0.02	0.0012
Radius of the proton diaphragm, mm	0.25	0.0004
Radius of the electron detector, mm	0.25	0.0001
Radius of the electron diaphragm, mm	0.25	0.0006
The coefficient *a*	0.0051	0.0010
The coefficient *A*	0.0011	0.0005
Total systematic error	–	0.0029
Polarization measurement error	–	0.0025
Statistical error	–	0.0025
Total absolute error	–	0.0046

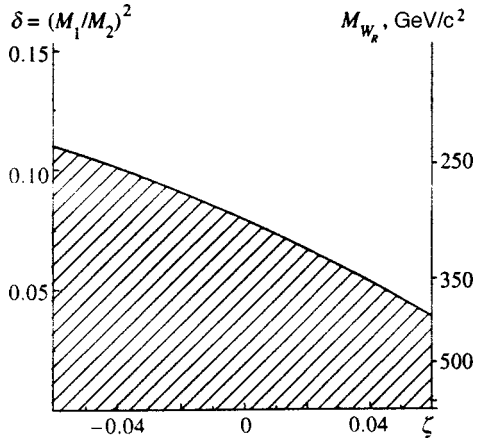


FIG. 6. Bounds on the parameters of the left-right-symmetric model from the new data on the *B* asymmetry ($B = 0.9821 \pm 0.0040$). The region of allowed model parameters at the 90% confidence level is hatched.

These bounds are graphically displayed in Fig. 6. When the mixing angle equals zero, it is seen that the mass of the right-handed *W* boson is

$$M_{W_R} \geqslant 284.3 \text{ GeV}/c^2.$$

On the other hand, when $\delta = 0$, the resulting value of *B* makes it possible to obtain a bound on the helicity of the antineutrino. The helicity of the antineutrino h_ν can be expressed as $h_\nu = 1 - \zeta^2$. It then follows from relation (15) that the strength of the antineutrino–spin correlation in neutron decay permits estimation of the deviation of the helicity from unity:

$$\Delta(h_\nu) \leqslant 0.0105 \ (90\%\text{C.L.}).$$

6. PROSPECTS FOR INCREASING THE ACCURACY OF THE MEASUREMENTS

As follows from an analysis of Table IV, progress in improving the accuracy of the measurements is needed in three main areas: 1) increasing the statistical accuracy; 2) improving the energy resolution of the electron detector; and 3) eliminating systematic effects in the polarization measurements.

Concrete steps can be proposed in each of these areas.

1. To increase the statistical accuracy there is a plan to increase the total flux of polarized neutrons in the PF1 beam. A focusing multislit supermirror polarizer with a cross-sectional area equal to 6×12 cm^2, i.e., 4.8 times the existing value, is being developed for this purpose. Another way to significantly increase the size of the data set of the events is to increase the free-flight pedestal from which the decay products are collected, and to increase the number of detecting systems by a factor of 5–10. However, this procedure should be regarded as a separate approach requiring considerable capital investments.

2. The energy resolution of the electron detector can be improved by replacing the existing scintillation electron detector by a semiconductor Si(Li) detector with a diameter of 70 mm, which has already been developed jointly with the Tokyo Institute of Technology and whose use in future measurements is planned.

3. The systematic effects in the polarization measurements can be eliminated by going over to rectilinear analyzing systems, as well as by diminishing the depolarizing effects as a result of improvements in the uniformity of the field of the magnets.

On the whole, the accuracy of the measurement of antineutrino spin asymmetry can be brought up to the 2×10^{-3} level, and this is fairly important for the left-right-symmetric models of the weak interaction.

We thank O. V. Rozhnov for his assistance in adjusting the experimental equipment; V. V. Ivanov, I. I. Marchenkov, and V. A. Solov'ev for the electronic instrumentation used in the experiment; and N. G. Kolyvanova and V. A. Tyukavin for their assistance in configuring the electronic instrumentation and software for the polarization measurements.

This work was performed with support from the Russian Fund for Fundamental Research (Grant 96-02-18672) and INTAS (Grant 96-537).

*)E-mail: ikuz@lnpi.spb.su

[1] A. P. Serebrov and N. V. Romanenko, JETP Lett. **55**, 503 (1992).
[2] W. Mampe, P. Aregon, J. C. Bates *et al.*, Nucl. Instrum. Methods Phys. Res. A **284**, 111 (1989).
[3] J. Byrne, P. G. Dawber, J. A. Spain *et al.*, Phys. Rev. Lett. **65**, 289 (1990).
[4] V. P. Alfimenkov, V. E. Varlamov, A. V. Vasil'ev *et al.*, JETP Lett. **52**, 373 (1990).
[5] W. Mampe, L. N. Bondarenko, V. I. Morozov *et al.*, JETP Lett. **57**, 82 (1993).
[6] E. Klemt, P. Bopp, L. Hornig *et al.*, Z. Phys. C **37**, 179 (1988).
[7] B. G. Erozolimskiĭ, I. A. Kuznetsov, I. V. Stepanenko *et al.*, Yad. Fiz. **52**, 1583 (1990) [Sov. J. Nucl. Phys. **52**, 999 (1990)].
[8] K. Schreckenbach, P. Liaud, R. Kossakowski *et al.*, Phys. Lett. B **349**, 427 (1995).
[9] A. Jodido, B. Balke, J. Carr *et al.*, Phys. Rev. D **34**, 1967 (1986).
[10] F. Abe, M. G. Albrow, D. Amidel *et al.* (CDF collaboration), Phys. Rev. Lett. **74**, 2900 (1995).
[11] I. A. Kuznetsov, A. P. Serebrov, I. V. Stepanenko, *et al.* Phys. Rev. Lett. **75**, 794 (1995).
[12] B. G. Erozolimskiĭ, Yu. A. Mostovoĭ, and B. A. Obinyakov, Prib. Tekh. Eksp. (**2**), 39 (1964).
[13] A. P. Serebrov, M. S. Lasakov, A. V. Aldushchenkov *et al.*, Nucl. Instrum. Methods Phys. Res. A **357**, 503 (1995).
[14] A. I. Egorov, V. M. Lobatev, V. A. Nazarenko *et al.*, Yad. Fiz. **19**, 300 (1974) [Sov. J. Nucl. Phys. **19**, 147 (1974)].
[15] C. Stratowa, R. Dobrozemsky, and P. Weinzierl, Phys. Rev. D **18**, 3970 (1978).

Translated by P. Shelnitz

Associated production of Higgs bosons with *Z* bosons by charged leptons in strong external fields

P. A. Éminov, K. V. Zhukovskiĭ,*) and K. G. Levchenko

Moscow State Institute of Electronics and Mathematics, 109028 Moscow, Russia
(Submitted 4 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1979–1990 (June 1998)

The probabilities of the associated production of a Higgs boson with a *Z* boson by a charged lepton in the field of a plane electromagnetic wave of arbitrary intensity and in a constant crossed field are obtained. The behavior of the cross section of the process as a function of the particle energies and the external field intensity is investigated for various values of the Higgs boson mass. It is shown that there is a logarithmic increase in the photoproduction cross section at superhigh energies up to a value significantly exceeding the cross section of the reaction $e^+ + e^- \to Z + H$, which is presently regarded as the most probable channel for the production of Higgs bosons. © *1998 American Institute of Physics.* [S1063-7761(98)00406-5]

1. INTRODUCTION

Along with gauge invariance, one of the key elements of the standard model of electroweak interactions is the Higgs mechanism for the appearance of gauge boson masses and for the cancellation of divergences on the basis of spontaneous symmetry breaking.

The fundamental massive Higgs boson, whose existence underlies the Weinberg–Salam–Glashow (WSG) theory, has yet to be found. The experimental detection of scalar Higgs bosons would be a decisive verification of the standard model and, in a broader sense, of the very idea that the Higgs mechanism of spontaneous symmetry breaking operates in the physics of elementary particles.

In the WSG theory the masses of $W^{\pm}$ and Z bosons, as well as the vacuum mean v of the Higgs field, can be expressed in terms of the fine-structure constant α, the Fermi constant G_F, and the Weinberg angle θ_W [1,2]:

$$M_W = \left(\frac{\pi \alpha}{\sqrt{2} G_F} \right)^{1/2} \frac{1}{\sin \theta_W} = 80.37 \pm 0.19 \text{ GeV},$$

$$M_Z = \frac{M_W}{\cos \theta_W} = 91186.3 \pm 1.9 \text{ MeV},$$

$$v = (\sqrt{2} G_F)^{-1/2} \simeq 246 \text{ GeV},$$

while the Higgs particle mass M_H is a free parameter of the model and is related to the unknown dimensionless parameter λ, which characterizes the self-action of Higgs scalars:

$$M_H = \lambda v.$$

Since the interaction constants of Higgs bosons with other particles in the WSG theory are determined by the masses of these particles, the coupling of Higgs bosons to gauge bosons and heavy quarks is far stronger than their coupling to electrons and other light particles. Therefore, the combined production of a Higgs boson with $W^{\pm}$ and Z gauge bosons in e^+e^- and hadronic collisions is believed to be most promising when various mechanisms for production of the Higgs boson are selected, and it has been a subject of ongoing theoretical and experimental research.

In electron–positron collisions, Higgs bosons should be produced mainly as a result of the processes

$$e^+ + e^- \to W^+ W^- \nu_e \tilde{\nu}_e \to H + \nu_e + \tilde{\nu}_e,$$

$$e^+ + e^- \to Z \to Z + H, \quad e^+ + e^- \to e^+ e^- ZZ \to e^+ + e^- + H$$

(see Refs. 3–6), among which $e^+ + e^- \to Z + H$ is the most probable process when $\sqrt{s} \lesssim 500$ GeV,[7] where $\sqrt{s}$ is the energy of the colliding particles in the center-of-mass system. Experiments devised to detect the process $e^+ + e^- \to Z + H$ at LEP2 place a lower bound on the Higgs boson mass of $M_H > 75$ GeV.[1] The reaction cross section is at most 0.3 picobarn when $M_H \in (50, 350)$ GeV, and it decreases as the Higgs boson mass increases.[4,7]

Taking into account the results of Tevatron experiments that directly measure the *W* boson and *t* quark masses, we assume that the Higgs particle mass is 127^{+127}_{-72} GeV, and that the Higgs boson mass is at most 465 GeV with 95% probability.[1]

As for the upper bound on the *H* boson mass, it is not given, in principle, by the standard model and can be estimated only on the basis of ''common sense'' assumptions. For example, if it is taken into account that the constant λ describes both the self-action of Higgs bosons and the interaction of $W^{\pm}$ and Z bosons with one another and with *H* bosons, a strong interaction between the particles appears, which cannot be described by perturbation theory, if $M_H \gg M_Z, M_W$. A detailed analysis shows that the *H* boson mass should be at most 700 GeV in this case.[7–10]

Another possible channel for the production of Higgs particles is provided by electron–photon collisions. For example, Hagiroara *et al.*[11] studied the dependence of the cross section of the process $e + \gamma \to W + H + \nu_e$ on the Higgs boson mass in the energy range $\sqrt{s} = 200 - 2000$ GeV, and Eboli *et al.*[12] examined the reaction $e + \gamma \to e \gamma \gamma \to e + H$, in which the production of Higgs bosons with a mass $M_H > 140$ GeV becomes possible provided $\sqrt{s} > 500$ GeV. In this case it is

proposed that the hard photons be obtained by utilizing the inverse Compton effect, under which the spectrum of scattered radiation for $\chi=2\omega E/m^2\gg 1$ becomes nearly monochromatic and has a sharp maximum at $\omega'\approx E$ (ω, ω', and E are the energies of the incident and scattered photons and the relativistic electron, respectively).

Among the possible mechanisms for the production and decay of H bosons, the processes which take place in external electromagnetic fields should be singled out.

Such investigations are important, because, on the one hand, the probabilities of processes that are forbidden in the free case by the 4-momentum conservation law achieve appreciable values in an intense external field, and, on the other hand, an external field can be a powerful catalyst and can dramatically increase the amount of information provided by processes that are capable of occurring without a field.[13–15]

The possibility of the associated production of a Higgs boson and a Z boson by a charged lepton in external electromagnetic fields of various configuration is studied in the present work. The probabilities of the processes are calculated by finding exact solutions of the relativistic wave equations with exact consideration of the interaction of the charged particles with the external electromagnetic field.[13–16]

An expression for the probability of the process $e\to e+Z+H$ in the field of a plane electromagnetic field of arbitrary intensity is obtained in the second section.

The third section is devoted to the calculation of the probability of the process $e\to e+Z+H$ in a constant crossed field. At ultrarelativistic electron energies and relatively weak fields ($E,H\ll H_0=m^2/e\simeq 4.41\times 10^{13}$ G) the results of this section are also applicable to the description of the process in an arbitrary constant field.

In the fourth section asymptotic formulas for the probability of the process in a constant crossed field and for the cross section of the process in the field of a plane electromagnetic wave in the applicability region of perturbation theory with respect to a wave intensity parameter are found in several limiting cases.

It is shown that at high energies the cross section of the photoprocess $e+\gamma\to e+Z+H$, which we have investigated, can significantly surpass the cross section of the reaction $e^-+e^+\to Z+H$, with which definite prospects for detecting the Higgs boson have been tied.[1,3,4]

2. THE PROCESS $e\to e+Z+H$ IN THE FIELD OF AN ELECTROMAGNETIC PLANE WAVE

In the standard model of electroweak interactions the matrix element of the process under study has the following form:[17]

$$\langle f|S^{(2)}|i\rangle=\frac{ig^2M_Z}{\cos^2\theta_W\sqrt{4k_0k_0'}}\times J^\mu\left[g_{\mu\nu}-\frac{p_\mu p_\nu}{M_Z^2}\right]\frac{e_\nu^{(\lambda)*}(k')}{p^2-M_Z^2+i\Gamma_ZM_Z}.$$

Here $\Gamma_Z\simeq 2494.7\pm 2.6$ MeV is the decay width of the Z boson, $k=(k_0,\mathbf{k})$ and $k'=(k_0',\mathbf{k}')$ are the 4-momenta of the final-state Higgs and Z bosons, p is the 4-momentum of the intermediate Z boson, and J^μ is the electroweak current:

$$J^\mu=\int d^4x\,\overline{\psi}_{q'}(x)\gamma^\mu(g_V+g_A\gamma^5)\psi_q(x)\ \exp(ipx),$$

where $g_A=-1/4$, $g_V=-1/4+\sin^2\theta_W$, and $\psi_q(x)$ is the exact solution of the Dirac equation for an electron moving in the given external field. The wave function of an electron in an arbitrary plane-wave field specified by the 4-potential $A^\mu=A^\mu(\varphi)$, which depends only on the phase $\varphi=nx$ (n is the wave vector and $n^2=0$), can be represented in the form[14,18]

$$\psi_q(x)=(2q_0V)^{-1/2}\times\left[1+\frac{e}{2(nq)}(\gamma n)(\gamma A)\right]u(q)\ \exp(iS_q(x)),\qquad(1)$$

where V is the normalization volume, $u(q)$ is the bispinor amplitude of the free plane wave, which is the solution of the free Dirac equation

$$(\gamma q-m)u(q)=0,\quad q^2=m^2,$$

and $S_q(x)$ coincides with the classical action function for a particle moving in the field of a wave:

$$S_q(x)=-qx-\int_0^\varphi d\varphi\left[\frac{e}{(nq)}(qA)-\frac{e^2A^2}{2(nq)}\right].\qquad(2)$$

In the case of interest to us here of a circularly polarized wave specified by the vector potential

$$A^\mu(x)=a_1^\mu\cos\varphi+a_2^\mu\sin\varphi,$$

$$a_1^2=a_2^2=a^2,\quad a_1a_2=0,\quad a_1n=a_2n=0,\quad \varphi=nx,$$

it follows from Eqs. (1) and (2) that

$$\psi_q(x)=\left[1+\frac{e}{2(nq)}(\hat{n}\hat{a}_1\cos\varphi+\hat{n}\hat{a}_2\sin\varphi)\right]\times\frac{u(q)}{\sqrt{2Q_0}}\exp\left\{-ie\frac{a_1q}{(nq)}\sin\varphi+ie\frac{a_2q}{(nq)}\cos\varphi-iQx\right\}.$$

Here we have introduced the quasimomentum of an electron in a wave field:

$$Q^\mu=q^\mu-e^2\frac{a^2}{2(nq)}n^\mu.$$

Its square plays the role of the electron effective mass in the field:

$$Q^2=m_*^2=m^2(1+\xi^2),$$

where $\xi=\sqrt{-e^2a^2/m^2}$ is the classical wave intensity parameter, which is equal to the ratio of the work performed by the field to the wavelength to the electron rest energy.

We average the square of the modulus of the matrix element of the process over the spin states of the initial electron and sum over the polarizations of the final electron ac-

cording to the conventional rules, and we perform the summation over the polarizations of the Z boson using the formula

$$\sum_{\lambda=1,2,3} e_\nu^{(\lambda)}(k')e_\mu^{(\lambda)}(k')=-\left(g_{\mu\nu}-\frac{k'_\mu k'_\nu}{M_Z^2}\right),$$

where $e_\nu^{(\lambda)}(k')$ is the polarization 4-vector of the Z boson.

Next, performing the integration over the phase volume of the Higgs and final Z bosons in tensor form, for the total probability of the process per unit time and unit volume we obtain

$$W=\frac{G_F^2}{(2\pi)^3}M_Z^6\frac{m^2}{Q_0}\sum_{s>s_0}\int_{u_1}^{u_2}\frac{du}{(1+u)^2}\times\int_{M^2}^{\tau(u)}\frac{d\tau\sqrt{(\tau-M_Z^2-M_H^2)^2-4M_Z^2M_H^2}}{\tau((\tau-M_Z^2)^2+(\Gamma_ZM_Z)^2)}\times\left\{AE-4g_A^2\frac{m^2}{M_Z^2}F\left(B\frac{(\tau-M_Z^2)^2}{M_Z^2}+A\left(2-\frac{\tau}{M_Z^2}\right)\right)\right\}, \tag{3}$$

$$E=(g_A^2+g_V^2)\left[-2\xi^2\frac{u^2+2u+2}{u+1}(J_{s+1}^2+J_{s-1}^2-2J_s^2)-8J_s^2\left(1-\frac{\tau}{2m^2}\right)\right]+16(g_V^2-g_A^2)J_s^2+2g_Ag_V\frac{u+2}{u+1}J_s(J_{s-1}-J_{s+1})\frac{(np)}{m^2}\times4z\left(1-\frac{2\xi^2}{1+\xi^2}\frac{u}{u_sz^2}\right),\quad u_s=\frac{2s(np)}{m_*^2}, \tag{4}$$

$$F=-2\frac{\tau}{m^2}J_s^2+\frac{u^2}{u+1}\xi^2(J_{s+1}^2+J_{s-1}^2-2J_s^2),$$

$$M=M_Z+M_H,\quad A=\frac{8\tau M_Z^2+(\tau+M_Z^2-M_H^2)^2}{12\tau M_Z^2},$$

$$B=\frac{(\tau+M_Z^2-M_H^2)^2-\tau M_Z^2}{3\tau^2}.$$

Each term in (3) corresponds to the production of a Higgs boson and a Z boson as a result of the absorption of s photons from the field, whose minimum number equals

$$s_0=\frac{(M+m_*)^2-m_*^2}{2(nq)}.$$

In formulas (3) and (4) we have introduced the invariant integration variables $u=-1+(nq)/(nq')$ and $\tau=(sn+q-q')^2$. In addition,

$$u_{1,2}=\frac{(sn+q)^2-M^2-m_*^2\pm\sqrt{((sn+q)^2-M^2-m_*^2)^2-4m_*^2M^2}}{m_*^2},$$

$$\tau(u)+\frac{(sn+q)^2u}{1+u}-m_*^2u.$$

The argument of the Bessel functions $J(z)$ in (4) is defined by the formula

$$z=2s\frac{\xi}{\sqrt{1+\xi^2}}\sqrt{\frac{u}{u_s}\left(1-\frac{u}{u_s}-\frac{\tau(1+u)}{uu_sm_*^2}\right)}.$$

We note that the result in the form of (3) and (4) which we have obtained is exact. It is valid for any value of the classical wave nonlinearity parameter, including $\xi\gtrsim1$, at which the interaction of the electron with the field of the intense electromagnetic wave leads to effects which are nonlinearly dependent on the energy density of the wave.

Further integration, however, cannot be performed analytically, but under the condition $\xi\ll1$, which corresponds to the condition for the applicability of perturbation theory with respect to the external field, where the processes with the absorption of the minimum possible number of quanta from the field are most probable, it is possible to perform the expansion of (3) and (4) in powers of ξ^2 (for further details, see Ref. 14).

Along with the condition $\xi^2\ll1$, we also require that

$$2(nq)>(M+m_*)^2-m_*^2, \tag{5}$$

whereupon the process resulting from the absorption of one photon from the field becomes possible.

As a result, after dividing the probability (3) by the incident flux density $j=m^2\kappa/2\omega EV$ (ω is the photon energy, E is the electron energy, and $\kappa=2(nq)/m^2$) and setting $\xi^2=4\pi\alpha/m^2\omega V$ (α is the fine-structure constant), we obtain the cross section of the process $e+\gamma\to e+Z+H$ in the form

$$\sigma=\left(\frac{eG_Fm}{\pi}\right)^2\left(\frac{M_Z^2}{\kappa m^2}\right)^2\times\int_b^{1-a}\frac{d\lambda(1-M^2/\kappa\lambda m^2)^{1/2}(1-M_1^2/\kappa\lambda m^2)^{1/2}}{(\lambda-M_Z^2/\kappa M^2)^2}\times\left\{2AC-4g_A^2\frac{m^2}{M_Z^2}D\left[B\frac{m^4\kappa^2}{M_Z^4}\left(\lambda-\frac{M_Z^2}{\kappa m^2}\right)^2+A\left(2-\frac{\kappa\lambda m^2}{M_Z^2}\right)\right]\right\}, \tag{6}$$

where $\lambda=\tau/m^2\kappa$,

$$C=(g_V^2+g_A^2)[2\lambda(1-\lambda)-1]\ln\frac{1-\lambda}{a}-2\lambda(g_V^2+g_A^2)(1-\lambda-a)+4g_Vg_A\left(\frac{1}{2}-\lambda\right)\times\ln\frac{1-\lambda}{a}-4g_Vg_A(1-\lambda-a),$$

$$M_1=M_Z-M_H,$$

$$D=[1-2\lambda(1-\lambda)]\ln\frac{1-\lambda}{a}+2\lambda(1-\lambda-a),$$

$$b=\kappa a=\frac{1}{k}\frac{M^2}{m^2},$$

and A and B are defined by (4).

3. THE PROCESS $e\to e+Z+H$ IN A CONSTANT CROSSED FIELD

In this section we examine the process $e\to e+Z+H$ in a constant crossed field (the intensities of the magnetic and electric fields are equal in magnitude and orthogonal to one another: $|\mathbf{E}|=|\mathbf{H}|$, $\mathbf{E}\perp\mathbf{H}$, and both field invariants are equal to zero).

A crossed field is a special case of an electromagnetic field of the plane-wave type and can be specified by the potential

$$A^\mu=a^\mu\varphi,\quad an=0, \tag{7}$$

and the complete wave function of an electron in a crossed field is obtained from (1) with consideration of (7):

$$\psi_q(x)=\left[1+\frac{e(\gamma n)(\gamma a)}{2(nq)}\varphi\right]\frac{u(q)}{\sqrt{2q_0V}}\times\exp\left[-ie\frac{(aq)}{2(nq)}\varphi^2+ie^2a^2\frac{\varphi^3}{6(nq)}-i(qx)\right].$$

The probability of the process in a crossed field can be obtained from the general formula (1) using the initial- and final-state wave functions of the electron in the crossed field. However, here we shall utilize another method to calculate the quantity of interest to us, which is based on the exact result (3) for the case of a circularly polarized wave. In fact, in a circularly polarized wave the total probability of the process depends on the two invariant parameters ξ and χ:

$$\xi=\sqrt{-\frac{e^2a^2}{m^2}}=\frac{eF}{m\omega},\quad \chi=\frac{e}{m^3}[-(F^{\alpha\beta}q_\beta)]^{1/2}=\xi\frac{(nq)}{m^2}.$$

In this case the electric and magnetic field intensity vectors rotate in a plane perpendicular to the wave propagation direction with a frequency equal to the frequency of the wave.

Therefore, when $\omega\to0$ ($\xi\to\infty$), the total probability of the process in the field of a circularly polarized wave should coincide exactly with the probability of the process in a constant crossed field:[14,15,18]

$$\lim_{\xi\to\infty}W(\xi,\chi)\equiv W(\infty,\chi)\equiv W(\chi). \tag{8}$$

We again note that the result obtained on the basis of the limiting transition (8) is exact for a crossed field at any energy and that in the ultrarelativistic case (when the electron energy $\varepsilon\gg m$), as we have noted above, it describes the probability of the process in a constant external electromagnetic field of arbitrary form with an intensity $F\ll H_0$ (for more exact applicability conditions, see Ref. 14).

Exchanging the order of summation and integration in Eq (3), we obtain the following expression for the probability:

$$W=\sum_{s>s_0}\int_0^{2\pi}d\varphi\int_0^\infty du\int_{M^2}^\infty d\tau W(u,\tau,s,\varphi)=\int_0^{2\pi}d\varphi\int_0^\infty du\int_{(M/m)^2}^\infty d\alpha\sum_{s>s_{\min}}m^2W(u,\alpha,s,\varphi),$$

where

$$\alpha=\frac{\tau}{m^2},\quad s_{\min}=\frac{\xi^3u}{2\chi}\left[1+\frac{1}{\xi^2}\left(1+\alpha\frac{u+1}{u^2}\right)\right],\quad \chi=\frac{(nq)}{m^2}\xi.$$

Since the variables take the values $z\sim s\sim\xi^3\gg1$ over the range of z and s that makes a significant contribution to the total probability when $\xi^2\gg1$, the sum over s can be replaced by integration over the new variable τ. For this purpose we must use the relation

$$s=\frac{\xi^3u}{2\chi}\left(1+\frac{2\tau}{\xi}\right)+s_{\min}.$$

As a result, we obtain

$$W=\int_0^{2\pi}d\varphi\int_0^\infty du\int_{(M/m)^2}^\infty d\alpha\times\int_{-\xi/2}^\infty d\tau\frac{\xi^2u}{\chi}m^2W(\varphi,u,\tau,\alpha). \tag{9}$$

We next apply the asymptotic behavior of the Bessel functions, in terms of which $W(\varphi,u,\tau,\alpha)$ can be expressed when their argument and the index tend to infinity with increasing ξ^2, the ratio between the latter tending to unity[19]:

$$J_s(z)\simeq\frac{1}{\pi}\left(\frac{2}{s}\right)^{1/3}\Phi(y),$$

where $\Phi(y)$ is the Airy function, whose argument is

$$y=\left(\frac{s}{2}\right)^{2/3}\left(1-\frac{z^2}{s^2}\right)=\left(\frac{u}{2\chi}\right)^{2/3}\left[1+\alpha\frac{u+1}{u^2}+\tau^2\right].$$

In the limit $\xi\to\infty$, the integral over the angular variable τ in Eq. (9) can be calculated using relations known from the theory of Airy functions.[14] For the total probability of the process $e\to e+Z+H$ in a constant crossed field we ultimately obtain

$$W=-\frac{1}{\sqrt{\pi}}\frac{G_F^2M_Z^6}{(2\pi)^3Q_0}\left(\frac{m}{M}\right)^2\int_0^1\frac{dx}{(1-xM_Z^2/M^2)^2}\times(1-x)^{1/2}\left(1-x\frac{M_1^2}{M^2}\right)^{1/2}\int_0^\infty\frac{du}{(u+1)^2}G(u,x), \tag{10}$$

where

$$G(u,x)=4F_1G_1-8g_A^2\frac{m^2}{M_Z^2}G_2\left[\frac{F_2}{x^2}\left(1-x\frac{M_Z^2}{M^2}\right)^2\left(\frac{M}{M_Z}\right)^4\right.$$

$$+\frac{F_1}{x}\left(\frac{M}{M_Z}\right)^2\left(2x\frac{M_Z^2}{M^2}-1\right)\Bigg],$$

$$G_1=4(g_A^2-g_V^2)\Phi_1+(g_A^2+g_V^2)\Bigg[\left(2-\left(\frac{M}{m}\right)^2 x\right)\Phi_1 - 2\,\frac{u^2+2u+2}{u+1}\Phi'\left(\frac{\chi}{u}\right)^{2/3}\Bigg],$$

$$G_2=\left(\frac{M}{m}\right)^2 x\Phi_1+2\frac{u^2}{u+1}\Phi'\left(\frac{\chi}{u}\right)^{2/3},$$

$$F_1=\frac{2}{3}+\frac{1}{6}\left(1-\frac{M_H^2}{M_Z^2}\right)+\frac{1}{12}\left(\frac{M}{M_Z}\right)^2\frac{1}{x}+\frac{1}{12}\times\left(1-\frac{M_H^2}{M_Z^2}\right)^2\left(\frac{M_Z}{M}\right)^2 x, \tag{11}$$

$$F_2=\frac{1}{3}+\frac{2}{3}\,\frac{M_Z^2-M_H^2}{M^2}x+\frac{(M_Z^2-M_H^2)^2}{3M^4}x^2-\frac{1}{3}\left(\frac{M_Z}{M}\right)^2 x.$$

Equation (11) employs not only the Airy function $\Phi(z)$, but also the related function

$$\Phi_1(z)=\int_0^\infty \Phi(t)dt,$$

with argument

$$z=\left(\frac{u}{\chi}\right)^{2/3}\left[1+\left(\frac{M}{m}\right)^2\frac{1}{x}\,\frac{u+1}{u^2}\right]. \tag{12}$$

We note that for a constant magnetic field $\mathbf{H}\uparrow\uparrow z$ with intensity $H\ll H_0=m^2/e=4.41\times10^{13}$ G and an ultrarelativistic electron with zero longitudinal momentum (the energy $E\gg m$, and $p_z=0$), the spectral variable and the dynamical parameter in Eqs. (10)–(12) are defined to be

$$u=\frac{p_\perp}{p'_\perp}-1=\sqrt{\frac{n}{n'}}-1,\quad \chi=\frac{H}{H_0}\,\frac{p_\perp}{m},$$

where $p_\perp=\sqrt{2eHn}$ is the transverse momentum of the electron in the magnetic field and n is the principal quantum number. The energy levels of the electron in the magnetic field are given by[16]

$$E=\sqrt{2eHn+m^2+p_z^2}.$$

4. LIMITING CASES AND DISCUSSION OF RESULTS

We first discuss several results following from the process of the associated production of a Higgs boson and a Z boson by an electron in a constant crossed field investigated in Sec. 3.

In the region $\chi\ll(M/m)^2$ the main contribution to (10) is made by the region $z\gg1$, where the Airy function has the asymptotic behavior

$$\Phi(z)\simeq\frac{1}{2}z^{-1/4}\exp\left(-\frac{2}{3}z^{3/2}\right). \tag{13}$$

Using (13), we calculate the integral over the spectral variable in (10) by the saddle-point method, the saddle point being the solution of the equation

$$2-\frac{\lambda}{u_0}-4\frac{\lambda}{u_0^2}=0,\quad \lambda=\left(\frac{M}{m}\right)^2\frac{1}{x},$$

whence $u_0\simeq\lambda/2\gg1$.

As a result, we obtain

$$W=\frac{G_F^2M_Z^6}{(2\pi)^3Q_0}\,\frac{16}{\sqrt{3}}\chi\int_{\left(\frac{M}{m}\right)^2}^\infty\frac{d\lambda}{\lambda}G(\lambda)\left[1-\frac{M^2}{m^2\lambda}\right]^{1/2}\times\exp\left[-\sqrt{3}\,\frac{\lambda}{\chi}\right], \tag{14}$$

where

$$G(\lambda)=\left[1-\frac{M^2}{m^2\lambda}\right]^{1/2}\frac{1}{(\lambda-M_Z^2/m^2)^2}\times\left\{(g_V^2+g_A^2)F_1+2g_A^2\left(\frac{m}{M_Z}\right)^2\left[F_2\frac{m^4}{M_Z^4}\times\left(\lambda-\frac{M_Z^2}{m^2}\right)^2+F_1\left(2-\lambda\frac{m^2}{M_Z^2}\right)\right]\right\}.$$

The integral in (14) over the variable λ is calculated anew by the saddle-point method.

The result for the total probability of the process in the case of relatively small values of the dynamical parameter, where $\chi\ll(M/m)^2$, has the form

$$W\simeq\frac{8G_F^2M_Z^6}{(2\pi)^3Q_0}\,\frac{\sqrt{2\pi}}{\psi^{5/2}}\left(\frac{M}{m}\right)^2 G\left(\frac{M^2}{m^2}\right)\exp(-\psi),$$

$$\psi=\sqrt{3}\left(\frac{M}{m}\right)^2\frac{1}{\chi}.$$

In accordance with the saddle-point method, the function $G(\lambda)$ in the last formula is taken at the saddle point $\lambda=(M/m)^2$.

It is noteworthy that the exponential dependence of the probability of the process at relatively small values of the dynamical parameter χ is characteristic of processes that are forbidden in the absence of an external field.

We next find the probability of the process in the highly interesting case of large values of the dynamical parameter, where $\chi\gg(M/m)^2$.

In the limiting case under consideration, the argument of the Airy function (12) in the real region can be set equal to

$$z\simeq\left(\frac{M}{m}\right)^2\frac{1}{\chi^{2/3}u^{1/3}x}. \tag{15}$$

We note that while the main contribution to the integral over the spectral variable is made by the vicinity of the saddle point $u_0\simeq\lambda/2\geqslant(1/2)(M/m)^2\gg1$ when $\chi\ll(M/m)^2$, the dominant contribution comes from the relatively broad range $1\ll u\ll(M/m)^2$ when $\chi\gg(M/m)^2$.

When (15) is taken into account, the integration over the variable u is performed using the integrals

$$\int_0^\infty t\Phi'(t)dt=-\Phi_1(0)=-\frac{\sqrt{\pi}}{3},$$

$$\int_0^\infty t^2\Phi_1(t)dt=\frac{2}{3}\Phi_1(0).$$

As a result, we find

$$W=\frac{16G_F^2M_Z^6}{3(2\pi)^3Q_0}\left(\frac{m}{M}\right)^6\chi^2\int_0^1\frac{x^2dx}{(1-xM_Z^2/M^2)^2}\times(1-x)^{1/2}\left(1-x\frac{M_1^2}{M^2}\right)^{1/2}\left\{(g_V^2+g_A^2)F_1+2g_A^2\left(\frac{m}{M_Z}\right)^2\left[\frac{F_2}{x^2}\left(1-x\frac{M_Z^2}{M^2}\right)^2\left(\frac{M}{M_Z}\right)^4+\frac{F_1}{x}\left(\frac{M}{M_Z}\right)^2\left(2x\frac{M_Z^2}{M^2}-1\right)\right]\right\}. \tag{16}$$

The integral over x in (16) has been tabulated, but because of the cumbersome nature of the result, here we present only the asymptote of the probability (16) for $M_H\gg M_Z$ and its value at $M_H=M_Z$:

$$W=C(g_V^2+g_A^2)\begin{cases}\dfrac{1}{240}\left(\dfrac{M_H}{M_Z}\right)^2, & M_H\gg M_Z,\\ \dfrac{16(23\sqrt{3}\pi-125)}{9}, & M_H=M_Z,\end{cases} \tag{17}$$

where

$$C=\frac{16}{3}\frac{G_F^2M_Z^6}{(2\pi)^3Q_0}\left(\frac{m}{M}\right)^6\chi^2.$$

The result (17) for $M_H\gg M_Z$ agrees to within a numerical factor of order unity with the result in Ref. 20, where the probability of the process $e\to e+Z+H$ in an ultrastrong magnetic field was calculated and it was shown that the associated production of a Higgs boson with a Z gauge boson in an ultrastrong magnetic field can be a fairly likely process.

We now move on to a study of the limiting cases of Eq. (6), which describes the cross section of the photoproduction process $e+\gamma\to e+Z+H$.

Figure 1 shows the dependence of the cross section of the process $e+\gamma\to e+Z+H$ on κ, which was constructed from Eq. (6) for various values of the Higgs boson mass M_H. Near the reaction threshold, where, according to (5), $\kappa\gtrsim M^2/m^2\approx10^{11}-10^{12}$ GeV, the reaction cross section is small compared to the cross section of the reaction $e^-+e^+\to Z+H$.

When $\sqrt{s}\gg M_H$ ($\sqrt{s}$ is the energy of the colliding particles in the center-of-inertia system), the cross section of the reaction $e^-+e^+\to Z+H$ decreases proportionally to s^{-1} and is given by[17]

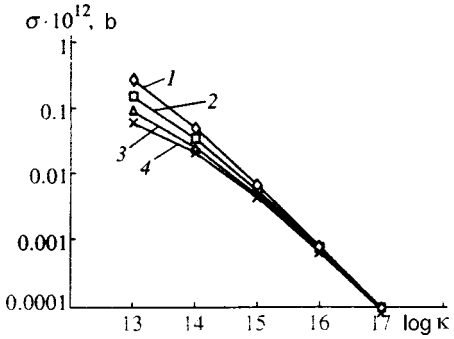


FIG. 1. Dependence of the cross section of the process $e+\gamma\to e+Z+H$ on κ for various values of the Higgs boson mass: $M_H=100$ (*1*), 200 (*2*), 300 (*3*), and 400 (*4*) GeV.

$$\sigma(e^++e^-\to Z+H)=\frac{G_f^2M_Z^4}{48\pi s}\times(1-4\sin^2\theta_W+8\sin^4\theta_W), \tag{18}$$

while in the logarithmic approximation ($\ln(\kappa m^2/M^2)\gg1$) from (6) we obtain

$$\sigma(e^-+\gamma\to e^-+Z+H)=\begin{cases}\sigma_1, & k\gg1,\\ \sigma_2, & k\ll1,\end{cases}$$

$$k=\frac{1}{\kappa}\left(\frac{M_Z}{m}\right)^4\ln\frac{\kappa m^2}{M^2},$$

$$\sigma_1=\frac{1}{3}(g_V-g_A)^2\left(\frac{eG_fm}{\pi}\right)^2\frac{1}{\kappa}\left(\frac{M_Z}{m}\right)^4\ln\left(\frac{\kappa m}{M}\right)\ln\left(\frac{\kappa m^2}{M^2}\right), \tag{19}$$

$$\sigma_2=\frac{2}{3}g_A^2\left(\frac{eG_fm}{\pi}\right)^2\ln\frac{\kappa m^2}{M^2}.$$

In the case of a head-on collision of a photon with an energy equal to the energy of the electron, from formulas (18) and (19) we find

$$\frac{\sigma(e+\gamma\to e+Z+H)}{\sigma(e^++e^-\to Z+H)}\simeq\begin{cases}C_1, & k\gg1,\ \text{i.e.,}\quad \kappa\ll10^{22},\\ C_2, & k\ll1,\ \text{i.e.,}\quad \kappa\gg10^{23},\end{cases}$$

$$C_1=5\alpha\ln\left(\frac{2E}{m}\right)\ln\left(\frac{4E^2}{Mm}\right),\quad C_2=\alpha\ln\left(\frac{2E}{M}\right)\kappa\left(\frac{m}{M_Z}\right), \tag{20}$$

where α is the fine-structure constant.

The conditions for the applicability of Eqs. (19) and (20) are satisfied over a broad range of energy and field intensity values. For example, when $E>1000$ GeV, we obtain $C_1>10$. Thus, as follows from (20), at high energies the cross section of the process that we investigated can significantly exceed the cross section of the reaction $e^++e^-\to Z+H$, which is presently regarded as the most probable channel for the production of Higgs bosons.

We thank A. V. Borisov and A. S. Vshivtsev for discussing the results of this work.

*)E-mail: zhukovsk@th180.phys.msu.su

[1] J.-P. Grivaz, LAL Preprint No. 97-60 (1997), p. 19.
[2] L. B. Okun, *Leptons and Quarks*, 2nd ed., North-Holland, Amsterdam (1984).
[3] V. Barger, K. Cheung, R. J. N. Phillips *et al.*, Phys. Rev. D **46**, 3725 (1992).
[4] V. Barger, K. Cheung, A. Djouadi *et al.*, Phys. Rev. D **49**, 79 (1994).
[5] E. Boos, M. Sachnitz, H. J. Schreiber *et al.*, Int. J. Mod. Phys. A **10**, 2067 (1995).
[6] W. Kilian, M. Kramer, and P. M. Zerwas, Phys. Lett. B **373**, 135 (1996).
[7] P. M. Zerwas, DESY Preprint No. 94-001 (updated May 1996), p. 46.
[8] J. Kuti and Lee Lin Yue Shen, Phys. Rev. Lett. **61**, 678 (1988).
[9] G. Altrelli and G. Isidori, Phys. Lett. B **337**, 141 (1994).
[10] J. A. Casas, J. R. Espinosa, and M. Quirós, Phys. Rev. Lett. **342**, 171 (1995).
[11] K. Hagiroara, I. Watanabe, and P. M. Zerwas, Phys. Lett. B **278**, 187 (1992).
[12] O. J. P. Eboli, M. C. Gonzales-Garsia, and S. F. Novaes, Phys. Rev. D **49**, 91 (1994).
[13] I. M. Ternov, V. Ch. Zhukovskiĭ, and A. V. Borisov, *Quantum Processes in Strong External Fields* [in Russian], Mosk. Gos. Univ., Moscow (1989).
[14] V. I. Ritus, Tr. Fiz. Inst. im. P. N. Lebedev, Akad. Nauk SSSR **111**, 5 (1979);A. I. Nikishov, ibid. **111**, 152 (1979).
[15] A. V. Borisov, A. S. Vshivtsev, V. Ch. Zhukovskiĭ, and P. A. Éminov, Usp. Fiz. Nauk **167**, 241 (1997) [Phys. Usp. Phys. Usp. **40**, 229 (1997)].
[16] A. A. Sokolov and I. M. Ternov, *Radiation from Relativistic Electrons*, AIP, New York (1986).
[17] S. I. Glashow, D. V. Nanopoulos, and A. Yildiz, Phys. Rev. D **18**, 1724 (1978).
[18] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **46**, 776 (1964) [Sov. Phys. JETP **19**, 529 (1964)].
[19] *Bateman Manuscript Project. Higher Transcendental Functions, Vol. 2*, A. Erdélyi (ed.), McGraw-Hill, New York (1953).
[20] P. A. Éminov, Zh. Éksp. Teor. Fiz. **98**, 34 (1990) [Sov. Phys. JETP **71**, 17 (1990)].

Translated by P. Shelnitz

Three-photon interference: spectroscopy of linear and nonlinear media

A. V. Burlakov,*) S. P. Kulik,*) A. N. Penin, and M. V. Chekhova

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 1 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1991–2004 (June 1998)

The nonlinear interference accompanying three-photon spontaneous parametric light scattering is considered. The frequency-angle line shape of the scattering is calculated for nonlinear-crystal/linear-dispersive-medium/nonlinear-crystal and nonlinear-crystal/nonlinear-crystal systems. The question of the possible use of nonlinear interference in spectroscopy is discussed.
 [S1063-7761(98)00506-X]

1. INTRODUCTION

Spontaneous parametric light scattering (SPS) spectroscopy and the measurement of scattering by polaritons have become traditional methods for studying the dispersive properties of crystals. The advantages of this type of spectroscopy include, first of all, its simplicity of implementation, as well as the possibility of measuring a wide set of quantities: the real (associated with the refractive index) and imaginary (associated with absorption) parts of the dielectric constant, and the real and imaginary parts of components of the second-order ($\chi^{(2)}$) and third-order ($\chi^{(3)}$) susceptibilities. These characteristics are measured over a wide spectral range—from the visible to the far-IR range, including the regions of fundamental crystal-lattice vibrations.[1] Among the drawbacks of the method we should mention the limitations on the class of investigatable materials: spontaneous parametric scattering is observed only in noncentrosymmetric media, where $\chi^{(2)} \neq 0$.

Phenomenologically, spontaneous paramagnetic scattering is explained by the spontaneous decay of a laser pump photon with frequency ω_p into a pair of photons—a signal photon (ω_s) and an idler photon (ω_i)—due to the quadratic susceptibility $\chi^{(2)}$ (Ref. 2). By virtue of the energy conservation law, the sum of the frequencies of the signal and idler radiation is exactly equal to the pump frequency, i.e.,

$$\omega_p = \omega_s + \omega_i, \tag{1}$$

and the momentum conservation law

$$\mathbf{k}_p = \mathbf{k}_s + \mathbf{k}_i \tag{2}$$

ensures coupling of the frequencies and angles of the scattered light $\omega_s(\theta_s)$.

The condition of spatial synchronism $\Sigma \mathbf{k}_i = 0$ (2), which holds for optical parametric processes, can be interpreted as a result of the nonlinear interference of the interacting waves. Physically, this condition is equivalent to stationarity of the total phase of all the modes, which ensures growth of the intensity in space in analogy to what occurs in ordinary linear interference. Parametric processes are characterized by a specific line shape,[3] i.e., by a dependence of the conversion efficiency on the wave-vector mismatch $\Delta = \sum_{i=1}^{m} \mathbf{k}_i$, where m is the number of interacting modes. For a nonlinear layer of thickness L in the plane wave approximation this dependence has the form

$$g(\Delta) = \sin^2(\Delta L/2)/(\Delta L/2)^2 \equiv \mathrm{sinc}^2(\Delta L/2). \tag{3}$$

For parametric light scattering, if the absorption at one of the frequencies[1)] is large

$$\alpha_i \gg 1/L, \tag{4}$$

then the scattering line shape becomes Lorentzian. In this case spontaneous parametric scattering transforms into scattering by polaritons and, in the limit $\omega_i \to \Omega$ (where Ω is the optical phonon frequency), into Raman scattering.[4] Condition (4) means that the signal fields created at points in space separated by an interval $x \gg 1/\alpha_i$ are now out of phase, i.e., their intensities, rather than their amplitudes, are added.

The interference nature of the formation of the frequency-angle spectrum for spontaneous paramagnetic scattering raises two obvious questions.

1) How do the phase delays added to any of the modes $\mathbf{k}_p$, $\mathbf{k}_s$, and $\mathbf{k}_i$ influence the spontaneous parametric scattering line shape $g(\Delta)$?

2) Is it possible to draw any conclusions about the dispersive properties of the medium used to introduce optical delays from the line shape?

The answer to the first question (for different parametric processes) has been known for a long time. For example, a method for measuring the relative phase between the first and second harmonics was proposed back in 1965.[5] The interference of second-harmonic signals from two nonlinear crystals separated by a dispersive medium underlied the technique for determining the signs of the components of the quadratic susceptibility $\chi^{(2)}_{ijk}$ in Ref. 6. Similarly, nonlinear interference has been used in coherent anti-Stokes Raman scattering (CARS),[7] as well as for determining the components of the third-order nonresonant susceptibility $\chi^{(3)}_{ijkl}$ in gases.[8] Nonlinear interference also describes quasisynchronous processes: the additional phase shift that interacting waves undergo in periodically nonuniform nonlinear

media—in the generation of the second harmonic,[9] as well as sum and difference frequencies,[10] and in CARS.[11]

The situation regarding the interpretation of nonlinear interference takes a dramatic turn in the case of spontaneous parametric scattering. Despite the fact that the nonlinear-optical aspects of the problem have not been investigated in detail either theoretically[12,13] or experimentally,[14,15] an unexpected surge of interest in the interference accompanying spontaneous parametric scattering took place in quantum optics in the middle of the 1980's. The paradoxicality of the problem is partly caused by the fact that spontaneous emission, by its very nature is noisy and contributions from macroscopically separated regions would not be expected to interfere. Even today different experimental schemes are being considered, in which a modulation structure is manifested either in the intensity of the scattered light or in coincidences of photon counts.[16,17] The observed features are associated with the quasimystical power of the photons to have a nonlocal effect on each other during spontaneous parametric scattering.[18] However, it seems that the most consistent description of nonlinear interference accompanying such scattering was given in Refs. 19 and 20. We note that the problem of the formation of an assigned frequency-angle spectrum of a biphoton field has been considered within the context of devising principles for quantum cryptography and a quantum computer.[21]

The resolution of the second question, which is closely associated with the first, is the subject of nonlinear interferometry—a spectroscopic method that has so far not found wide application. Nevertheless, different types of nonlinear interferometers have been used, for example, to measure the phases of nonlinear susceptibilities[22] (in various materials, including thin films[23]) and to measure the propagation times of photons through dispersive media in quantum optics.[24]

In the present work we have undertaken an attempt to answer the second question, i.e., to estimate the sensitivity of nonlinear interferometry to variation of the medium parameters[2)] which influence the phase of any of the three modes participating in the process of spontaneous parametric scattering (three-photon interferometry). We limit the discussion to the case in which the spatial inhomogeneity causing the phase shifts is in the direction of the pump wave vector. The scattering geometry with transverse inhomogeneity—nonlinear interference in Young's scheme[19]—is not considered here.

The work consists of two parts. In the first part, we consider a three-photon Mach–Zehnder interferometer on the basis of general relations for the spontaneous parametric scattering line shape. The frequency-angle line shape is calculated for a nonlinear-crystal/linear-dispersive-medium/nonlinear-crystal system. The case where a thin film with an isolated dielectric-constant resonance serves as the interlayer is treated separately. In the second part we obtain an expression for the spontaneous parametric scattering line shape of a nonlinear-crystal/nonlinear-crystal system, where the two nonlinear crystals differ only with respect to the dispersion of their refractive indices. We analyze the possibility of using three-photon interferometry to monitor the linear optical properties of nonlinear crystals. Obviously, a definitive solution for the problems treated would, on the one hand, substantially broaden the class of materials accessible to conventional SPS spectroscopy and, on the other, introduce a new (and often more convenient) technique for determining the dispersion law of noncentrosymmetric crystals.

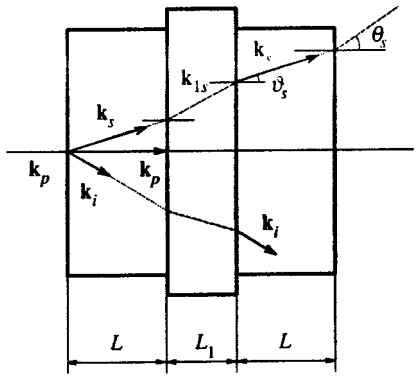


FIG. 1. Diagram for observation of three-photon interference: two nonlinear crystals (L) separated by a linear medium (L_1).

2. NONLINEAR-CRYSTAL/LINEAR-DISPERSIVE-MEDIUM/NONLINEAR-CRYSTAL SYSTEMS

We consider two planar nonlinear crystals of thickness L separated by a linear medium of thickness L_1 (Fig. 1). Such a three-layer system has been termed a nonlinear Mach–Zehnder interferometer[25] in analogy to the corresponding scheme in linear optics.[26] Here the linear medium introduces phase delays proportional to L_1 at all three frequencies:

$$\Phi_{1p} \propto n_{1p}(\omega_p) L_1/\lambda_p, \quad \Phi_{1s} \propto n_{1s}(\omega_s) L_1 \cos\vartheta_s/\lambda_s,$$

$$\Phi_{1i} \propto n_{1i}(\omega_i) L_1 \cos\vartheta_i/\lambda_i.$$

An expression for the spontaneous parametric scattering line shape for this case was obtained in Ref. 13:

$$g(\omega_s) = \left\{ \frac{\sin(\delta/2)}{\delta/2} \cos\left[\frac{\delta+\delta_1}{2} \right] \right\}^2, \tag{5}$$

where $\delta(\omega_s, \vartheta_s) = \Delta L = (k_p - k_s - k_i)L$ and $\delta_1(\omega_s, \vartheta_s) = \Delta_1 L_1 = (k_{1p} - k_{1s} - k_{1i})L_1$ are the wave-vector mismatches in the nonlinear and linear media, respectively, the k are the projections of the wave vectors onto the direction perpendicular to the layers, and ϑ_s is the scattering angle inside the crystals. We are interested below in the case where the orientations of the polar axes in the nonlinear crystals are opposed, which is equivalent to a phase shift of $\pi/2$ in the argument of the cosine in Eq. (5):

$$g(\omega_s, \vartheta_s) = \left\{ \operatorname{sinc} \frac{\delta}{2} \sin\left[\frac{\delta+\delta_1}{2} \right] \right\}^2, \tag{6}$$

under the assumption that the signal photons do not leave the interaction region (defined by the thickness of the crystal L and the diameter of the pump beam d, Ref. 19): $(2L + L_1)\tan\vartheta_s \ll d$. This condition imposes a practical limit on the angular range for observing the effect: when the distance between the crystals is increased, the contribution to the interference is due to small-angle scattering. Therefore, below

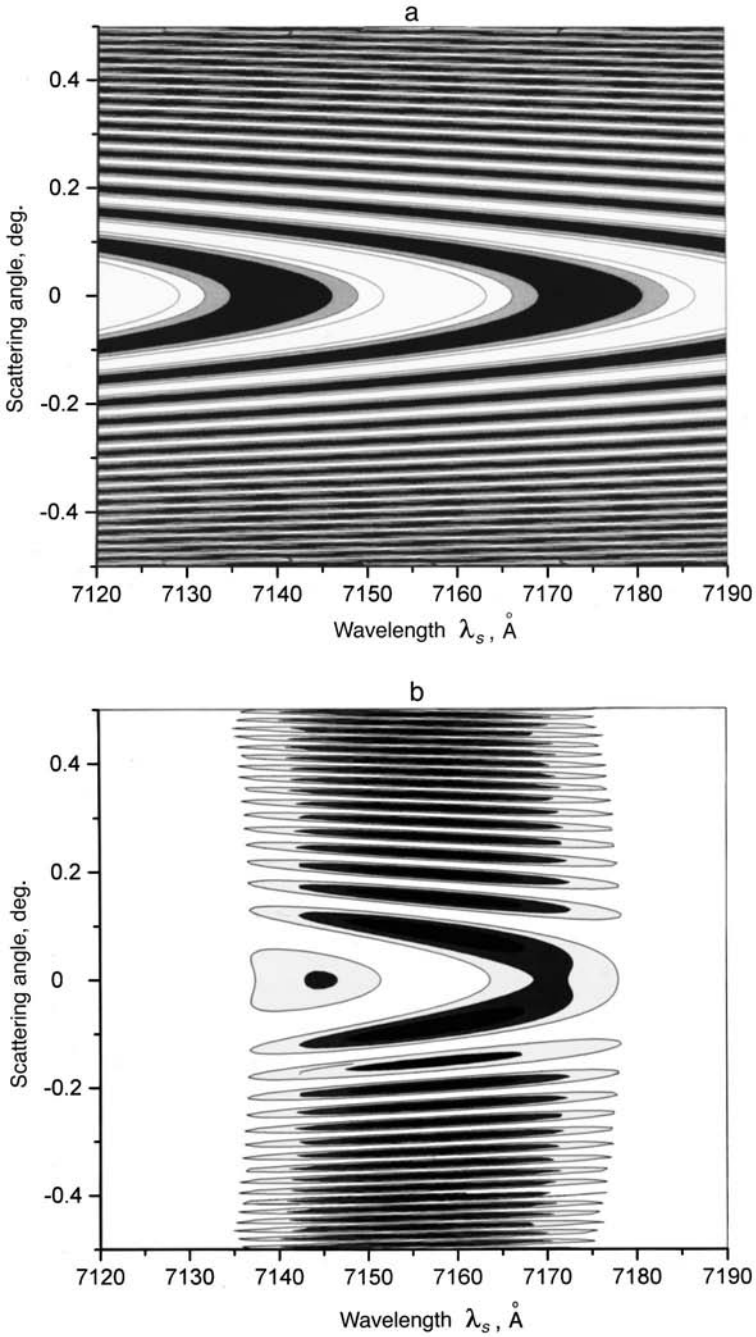


FIG. 2. Frequency-angle intensity distribution for three-photon interference: a—modulation function (7), b—consideration of the envelope (spatial synchronism).

we shall analyze the line shape for the case of collinear synchronism. The normalization in Eq. (6) is chosen such that the scattering intensity at the maximum would be equal to unity for $\delta_1=0$ and polar axes of identical direction. Expression (6) has a simple physical meaning: the first factor describes ordinary parametric scattering in the layer of thickness L. The second factor is due to the interference of the spontaneous fields arising in the two crystals, and the phase of the interference pattern depends on the dispersion of the linear medium. We note two peculiarities, which follow directly from the form of (6). First, this expression is reminiscent of the intensity distribution for the ordinary (linear) diffraction of a plane wave by a screen with two slits (Ref. 26).[3)] Second, the optical properties of the linear medium influence only the second factor in (6): the role of the first factor reduces to filtration of the frequency-angle distribution

$$g_1=\left\{\sin\left[\frac{\delta+\delta_1}{2}\right]\right\}^2 \tag{7}$$

by the envelope $g_0=\{\mathrm{sinc}(\delta/2)\}^2$ in accordance with the condition of spatial synchronism (2) in the nonlinear crystal of thickness L. Therefore, it would be reasonable to analyze the perceivable frequency-angle intensity distributions (7), which depend on the dielectric constants of the linear and nonlinear media, and to take into account the actual synchronism width in the final stage together with the comparison with experiment. Here the thickness of the nonlinear crystals,

as it were, assigns the size of the ''window'' through which we view the modulation structure: the smaller is L, the wider is the frequency-angle range of the observed interference pattern.

The explicit dependences of the mismatches appearing in (6) on the observable parameters ω_s and θ_s have the form

$$\Delta(\omega_s,\theta_s)=2\pi\left(n_p\omega_p-n_s\omega_s\sqrt{1-\left[\frac{\sin\theta_s}{n_s}\right]^2}-n_i\omega_i\sqrt{1-\left[\frac{\omega_s}{\omega_i n_s}\sin\theta_s\right]^2}\right), \tag{8}$$

$$\Delta_1(\omega_s,\theta_s)=2\pi\left(n_{1p}\omega_p-n_{1s}\omega_s\sqrt{1-\left[\frac{\sin\theta_s}{n_{1s}}\right]^2}-n_{1i}\omega_i\sqrt{1-\left[\frac{\omega_s n_i}{\omega_i n_{1i} n_s}\sin\theta_s\right]^2}\right), \tag{9}$$

where θ_s is the external scattering angle. The expressions (8) and (9) together with (6) are inputs to the calculation of the frequency-angle distributions of the spontaneous paramagnetic scattering intensity.

Figure 2 corresponds to the case in which two $LiNbO_3$ crystals of identical thickness $L=1$ mm are separated by a vacuum gap of thickness $L_1=10$ cm, the polar axes of the crystals are antiparallel, and absorption of the idler waves in this range can be neglected: $\alpha_i L\ll 1$ (Ref. 27). In the direction of collinear synchronism ($\theta_s=0$ and $\delta=\delta_1=0$) an intensity minimum is observed according to Eq. (6) regardless of the value of L_1 since the vacuum does not have dispersion:

$$\Phi_{1p}=\Phi_{1s}=\Phi_{1i}.$$

As the scattering angle θ_s increases the wave-vector mismatch in the gap δ_1 grows and a two-dimensional interference pattern appears. Figure 2a illustrates the behavior of the modulation function (7) for the above parameters, and Fig. 2b illustrates the result of taking the spatial synchronism into account. Let us turn our attention to the outward similarity between the frequency-angle distribution in Fig. 2a and spectra known already at the beginning of this century that were obtained by Rozhdestvenskiĭ[28] at the exit of a similar interferometer with illumination by white light.[4)] However, the ''illumination'' of a nonlinear interferometer is provided by the wideband radiation due to spontaneous parametric scattering, while that of a linear interferometer is provided by an external source (an incandescent lamp). Besides, in the linear schemes (the Mach–Zehnder interferometer or the similar Rozhdestvenskiĭ interferometer) the phase difference between the coherent waves shifts due to the difference between the optical paths in the arms of the interferometer. In the nonlinear case the arms are spatially degenerate, but the wave-vector mismatch ($\delta+\delta_1$) appearing in (7) is always a function of the frequency and angle. Only for exact collinear synchronism ($\delta=0$, $\theta_s=0$) and a vacuum gap does the total phase trajectory $\Phi_\Sigma=\Phi_{1p}-\Phi_{1s}-\Phi_{1i}$ vanish, regardless of the thickness of the gap L_1.

Figure 3 illustrates the frequency line shape of the scattering for a $LiNbO_3$—air—$LiNbO_3$ configuration with variation of the refractive index of the air layer at the idler frequency.

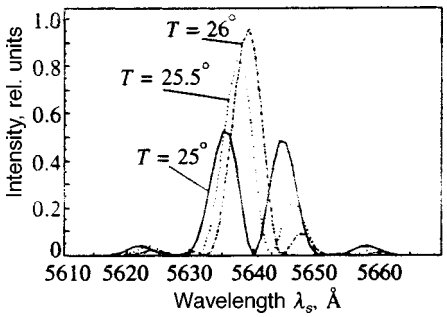


FIG. 3. Frequency line shape of the parametric scattering for different temperatures of the air between the crystals (collinear synchronism).

Variations of the refractive index of the intervening medium strongly influence the scattering line shape. For example, if the thickness of both crystals is $L=1$ mm and the thickness of the interlayer is $L_1=10$ cm in the $LiNbO_3$—air—$LiNbO_3$ configuration, then the interference phase will vary by π when the refractive index varies by 0.0001. Without dwelling on the reasons for deviations of the refractive index, we note that they can be caused not only by fluctuations in the temperature, pressure, and humidity, but also by the presence of all kinds of impurities—gases, small particles, etc. A quantitative analysis of the influence of these factors on the dispersion of the dielectric constant $\varepsilon_1(\omega)$ is beyond the scope of the present work. We would simply like to underscore the fact that there is an abrupt dependence of the scattering line shape on the parameters just mentioned. As a unique illustration we chose the sensitivity of the line shape to variation of the air temperature in the same configuration ($L=1$ mm, $L_1=10$ cm). Besides the dispersion of the refractive index of air at all three frequencies, we also took into account its dependence on the atmospheric pressure, the temperature, and the partial pressure of water vapor.[26] Figure 3 displays frequency scans of the spontaneous paramagnetic scattering spectrum at zero scattering angle. The three curves correspond to different air temperatures in the gap: 25 °C, 25.5 °C, and 26 °C.

Increasing the base of the interferometer to 1 m raises the sensitivity of the line shape to variation of the refractive index to 0.000 001 (and to variation of the temperature to a tenth of a degree). Here the traditional question arises: how stable is the scattering line shape toward errors in the determination of the interlayer thickness L_1? In other words, how strongly is the interference structure of the line shape smeared by fluctuations in the distance between the crystals? For linear interferometric devices the addition of a half-wavelength to the optical length of an arm shifts the phase of the interference pattern by π. The situation is completely different for three-photon interference. Displacements of one crystal relative to the other (without a change in the orientation of the optical axis) have much less of an effect on the intensity distribution in the interference pattern. The reason for this is the spatial degeneracy of the interferometer arms. Since the argument in (7) is a function of the three frequen-

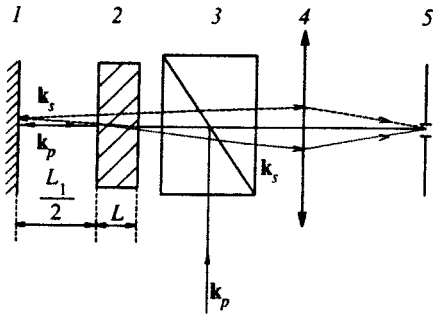


FIG. 4. Diagram of three-photon interference: *1*—wideband mirror, *2*—nonlinear crystal, *3*—Glan–Thompson prism, *4*—converging lens, *5*—entrance slit of the spectrograph.

cies ω_p, ω_s, and ω_i, the phase trajectory contains contributions at all three frequencies:

$$\Phi_\Sigma = 2\pi\{n_{1p}(\omega_p)/\lambda_p - n_{1s}(\omega_s)\cos\,\vartheta_s/\lambda_s - n_{1i}(\omega_i)\cos\,\vartheta_i/\lambda_i\}L_1 \tag{10}$$

(we assume that the wave-vector mismatch δ in the nonlinear medium is constant). For small scattering angles and weak dispersion of the material the expression in curly brackets in formula (10) differs insignificantly from zero; therefore, large variations in L_1 are necessary to produce appreciable variations in the phase. The derivatives defining the sensitivity and stability of a three-photon interferometer are equal, respectively, to

$$\partial\Phi_\Sigma/\partial n_{1i} = -2\pi L_1/\lambda_i, \tag{11}$$

$$\partial\Phi_\Sigma/\partial L_1 = 2\pi\{n_{1p}(\omega_p)/\lambda_p - n_{1s}(\omega_s)/\lambda_s - n_{1i}(\omega_i)/\lambda_i\} \tag{12}$$

(here $\cos\vartheta_s \approx \cos\vartheta_i \approx 1$). Expression (11) reflects the fact that increasing the base L_1 increases the accuracy of the measurement of the refractive index; Eq. (12) implies a weak dependence of the line shape on fluctuations of L_1 (e.g., caused by vibrations).

Figure 4 depicts one variant of the scheme of a three-photon interferometer. Here instead of two nonlinear crystals separated by a linear medium, one crystal is used with a wideband metallic mirror located a distance $L_1/2$ from it. Such a scheme is equivalent to the one discussed above (Fig. 1) in a regime similar to collinear synchronism. Upon reflection from the mirror a phase shift of π is added to all three modes, and thus the polar axis is rotated by 180° in the image of the crystal. A Glan–Thompson prism is used to spatially separate the orthogonally polarized pump (p) and signal (s). The scheme shown in Fig. 4 is convenient in that it eliminates the need to exactly align the orientations of the optical axes of the two crystals and significantly reduces the parasitic illumination characteristic of the recording of small-angle scattering. Frequency tuning is achieved by rotating the crystal in the plane containing the polar axis.

Figure 5a displays an experimental spontaneous parametric scattering spectrum recorded at the exit of a three-photon interferometer assembled according to the scheme in Fig. 4.[5)]

A lithium iodate crystal of thickness $L = 1$ mm is placed in front of an aluminum mirror. The crystal and the mirror are separated by an air gap of thickness $L_1/2 = 5$ mm. Pumping is provided by a 3-W argon laser with a wavelength $\lambda_p = 4880$ Å. The frequency of the idler photons coupled with the observed photons is 3770 cm^{-1}, and the synchronism width is 20 Å, which substantially exceeds the frequency width of the laser line. Figure 5b displays the calculated frequency-angle distribution of the intensity for the parameters indicated. The good agreement between the calculated and experimental spectra justifies an optimistic assessment of the prospects of using a three-photon interferometer.

So far we have assumed that all three frequencies ω_p, ω_s, and ω_i fall within the transparency range of the linear medium, where the dispersion of the dielectric constant is small. The wide spectrum of spontaneous parametric scattering, however, makes it possible to closely approach the re-

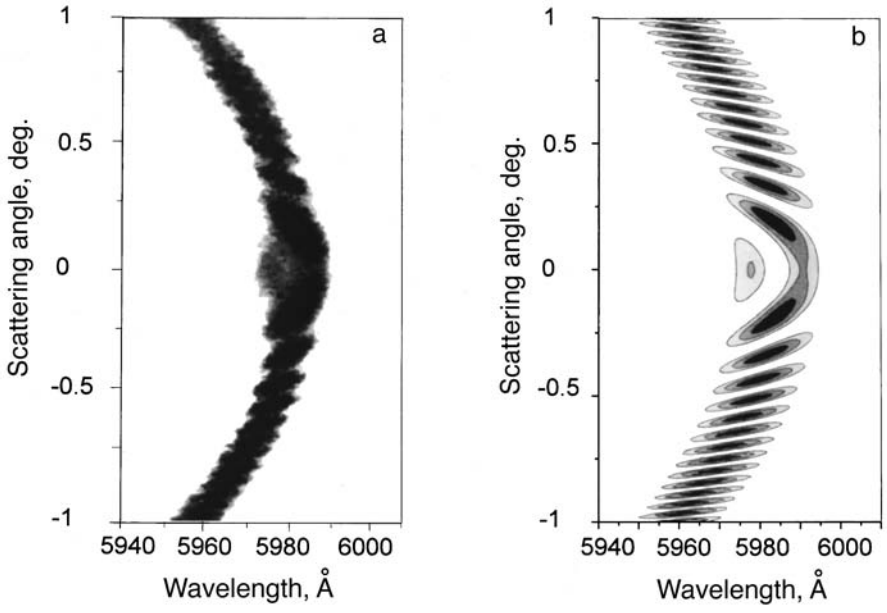


FIG. 5. a—Photograph of a spontaneous parametric scattering spectrum taken at the exit of a three-photon interferometer (Fig. 4). $LiIO_3$ crystal. b—Calculated frequency-angle distribution of the parametric scattering intensity at the exit of the interferometer (Fig. 4).

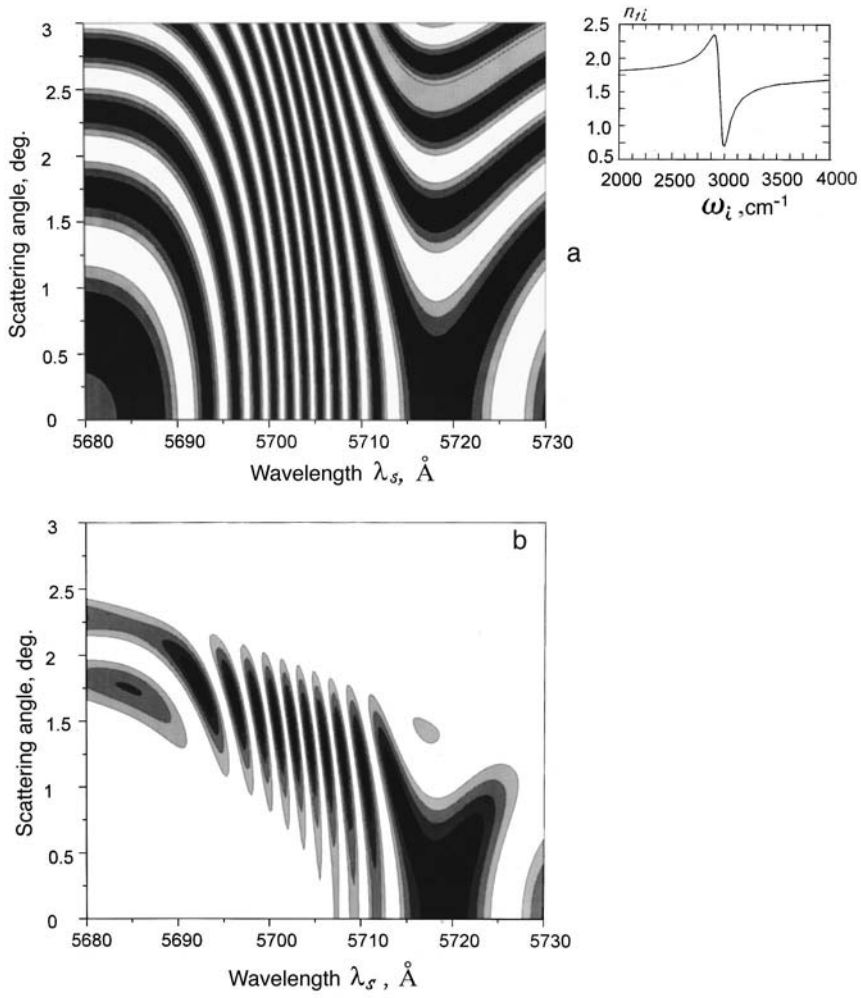


FIG. 6. Behavior of the modulation function (a) and the scattering line shape $g(\omega_s, \theta_s)$ (b) in the vicinity of a dielectric constant resonance with the parameters $\varepsilon_0 = 3.5$, $S = 0.17$, $\Gamma = 150$ cm^{-1}, and $\omega_0 = 2950$ cm^{-1}. The thickness of the linear medium $L_1 = 10$ μm.

gion of strong dispersion. Let us consider the behavior of the scattering line shape in the case of an isolated resonance of the intervening medium.

Let the frequency of the pump and signal photons lie, as before, in the transparency range, and let the dispersion of the medium at the idler frequencies be described in the single-oscillator approximation:

$$\varepsilon_1(\omega_i) = \varepsilon_0 + S\omega_0^2 \frac{\omega_0^2 - \omega_i^2}{(\omega_0^2 - \omega_i^2)^2 + \omega_0^2\Gamma^2}. \tag{13}$$

The values of the parameters ε_0, S (the oscillator strength), ω_0 (the eigenfrequency), and Γ (the damping constant) in (13) were taken from Ref. 29 for the high-frequency vibration of the O–H group: $\varepsilon_0 = 3.5$, $S = 0.17$, $\Gamma = 150$ cm^{-1}, and $\omega_0 = 2950$ cm^{-1} ($\lambda_s = 5701$ Å). Figures 6a and 6b show the frequency-angle distributions of the modulation function (7) and the scattering intensity (6), respectively. The inset in the upper right-hand corner of Fig. 6a illustrates the frequency dependence $\varepsilon(\omega)$ for the above parameters. In the vicinity of the resonance the behavior of the interference bands (Fig. 6a) is analogous to Rozhdestvenskiĭ's ''hooks,'' [28] which are observed near regions with anomalous dispersion. As follows from Eq. (10), the frequency position of the ''hook'' for the exact synchronism direction ($\partial\Delta_1/\partial\omega_s = 0$) corresponds to equality between the group velocities of the signal and idler photons in the linear medium:

$$\frac{\partial}{\partial\omega_s}(n_{1s}\omega_s) = \frac{\partial}{\partial\omega_i}(n_{1i}\omega_i).$$

It is obvious that the repetition period of the bands in the spectrum along the wavelength axis decreases as the resonance is approached: $\omega_i \to \omega_0$ since $\partial\Phi_\Sigma/\partial\omega_i \approx \partial n_{1i}/\partial\omega_i$. We emphasize that, according to Eq. (12), a decrease in the thickness of the linear layer L_1 can be compensated by strong dispersion of the dielectric constant of the linear medium. Hence we have the important result that it is possible to use three-photon interference to investigate thin dispersive films. Thus, the value of L_1 used in the calculation (Fig. 6) is 10 μm.

One of the conditions for observing three-photon interference in such a scheme is weak absorption of the idler photons in the nonlinear crystals (4). In other words, the idler photons must leave the first crystal and enter the second. Absorption lowers the contrast of the interference pattern but has essentially no effect on the positions of the interference maxima. A large number of nonlinear crystals having low absorption in the near-IR range are presently known,[30] and

the question of which one to choose does not pose serious difficulties for the practical implementation of a nonlinear interferometer.

Another contrast-lowering factor arises when the characteristic angular and frequency scales of the modulation line shape are smaller than the angular divergence and frequency width of the pump line, respectively. For example, consideration of the finite width of the pump line leads to the following restriction:

$$l_{\mathrm{coh}} \geqslant 2\frac{c}{n_{1p}} L_1 \left(\frac{1}{u_s \cos \vartheta_{1s}} - \frac{1}{u_i \cos \vartheta_{1i}} \right), \tag{14}$$

where l_{coh} is the pump coherence length, c/n_{1p} is the speed of light in the linear interval at the pump frequency, u_s and u_i are the group velocities of the signal and idler radiation, and ϑ_{1s} and ϑ_{1i} are the corresponding scattering angles inside the linear layer.

3. NONLINEAR-CRYSTAL/NONLINEAR-CRYSTAL SYSTEMS

The frequency-angle line shape for the spontaneous parametric scattering of a nonlinear-crystal/nonlinear-crystal system was calculated using the approach developed in Ref. 31. The quadratic susceptibilities of the crystals were assumed to be the same while the refractive indices differed. Such a situation arises, for example, when samples of $LiNbO_3$ are doped with magnesium atoms to prevent optical damage.[32] The Mg concentration strongly influences the value of the refractive index, especially in the IR region of the spectrum.[33] If the thicknesses of the crystals are equal to L and L_2, then the scattering line shape has the form

$$g(\omega_s, \theta_s) = \frac{1}{(L+L_2)^2} \left[L^2 \operatorname{sinc}^2 \frac{L\Delta}{2} + L_2^2 \operatorname{sinc}^2 \frac{L_2\Delta_2}{2} + 2LL_2 \operatorname{sinc} \frac{L\Delta}{2} \operatorname{sinc} \frac{L_2\Delta_2}{2} \cos \frac{L\Delta + L_2\Delta_2}{2} \right], \tag{15}$$

where $\Delta = k_p - k_s - k_i$ and $\Delta_2 = k_{2p} - k_{2s} - k_{2i}$ are the wave-vector mismatches in the first and second crystal, respectively, which coincide with formula (8). The normalization is chosen such that the scattering intensity at the maximum would be equal to 1 for $\Delta = \Delta_2$. The first two terms in the square brackets in (15) take into account the additive reinforcement of the scattering spectra of the two crystals, and the third term is the interference term. The interference phase is given by the ''three-photon sum'' of the optical paths $(L\Delta + L_2\Delta_2)/2$. Thus, nonlinear interference is manifested in the configuration under consideration only in the region of intersection of the frequency-angle spectra for the spontaneous parametric scattering of the two crystals. We note that the scattering line shape is especially sensitive to variations in the refractive index at any of the three frequencies when L and L_2 are large. In this case, however, the frequency width of the scattering line decreases. Thus, for example, in order to detect a change in the refractive index $\Delta n \sim 10^{-3}$ at the idler frequency $\omega_i = 5000\ \mathrm{cm}^{-1}$, it is sufficient to use a sample of thickness 1 mm, and in this case the expected frequency width of the scattering line turns out to be of the order of 50 cm^{-1}. If, on the other hand, the change in the refractive index at this same frequency is on the order of 10^{-4}, it becomes necessary to work with a crystal 10 mm in thickness, which corresponds to an order-of-magnitude decrease in the frequency width of the spectrum. In this case the absorption at the idler frequency can be substantial. In the presence of absorption the contrast of the interference pattern is reduced; strictly speaking, this description is applicable for $L, L_2 \ll 1/\alpha$, where α is the absorption coefficient at the idler frequency. All this imposes a limit on the accuracy of measurements of the refractive index

$$\Delta n_i \sim c\alpha/\omega_i,$$

i.e., the real part of a wave vector can be measured only to within the magnitude of its imaginary part. A similar problem arises in ordinary polariton scattering spectroscopy, where the real part of the dielectric constant at a given frequency is determined from the position of the maximum of the angular spectra (to within an order of the line width).[34]

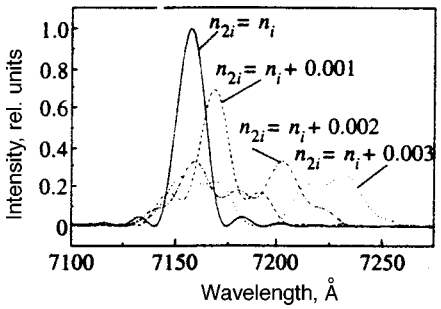


FIG. 7. Frequency line shape for scattering in two $LiNbO_3$ crystals with different values of the refractive index at the idler frequency. Here $L = L_2 = 1$ mm, and the pump wavelength is 4880 Å.

Figure 7 displays a family of calculated frequency distributions of the spontaneous parametric scattering intensity for collinear synchronism. Both crystals are magnesium-doped lithium niobate ($L = L_2 = 1$ mm).[35] The value of the refractive index at the idler frequency in the second crystal, which is more sensitive to variations in the magnesium concentration, serves as a parameter.[33] It can be seen that a deviation of the refractive index by 0.002 markedly distorts the scattering line shape. For $\Delta n_i \geqslant 0.005$ the influence of the interference term is considerably less due to the weak overlap of the spectra of the two crystals.

In conclusion, we note that the three-photon interference, by virtue of its high sensitivity and stability, might be useful in monitoring the presence of impurities in optical materials (especially if the impurity has resonances in the transparency band of the crystals employed), in the investigation of thin dispersive films, etc.

Also, we hope that the use of three-photon interferometry will raise the accuracy of standard SPS spectroscopy in determining the frequency dependence of the refractive index in nonlinear crystals and, in particular, in estimating their degree of spatial homogeneity and monitoring their twinning structure. Finally, interferometers of this kind can be used to study induced variations in the refractive index in photorefractive crystals. In one of the possible experimental

setups the crystal investigated is placed in the gap between two crystals of a different nonlinear material, which is insensitive to the radiation, and the interference pattern is observed as a function of the pump intensity. Typical values of the change in the refractive index in weakly photorefractive crystals (such as, for example, magnesium-doped lithium niobate) in the transparency region can amount to $10^{-4}-10^{-5}$ (Ref. 36), and crystals of thickness 1–10 cm are needed to detect such changes.

We are grateful to D. N. Klyshko for helpful advice. This work was carried out with the financial support from the Russian Fund for Fundamental Research (Grant No. 97-02-17498).

*)E-mail: postmast@qopt.ilc.msu.su

1)Usually the pump and signal frequencies lie in the transparency band of the crystal, while the idler frequencies fall in the IR range, including the regions of strong lattice absorption.

2)We are actually referring here to the set of parameters to which the frequency dependence of the refractive index $n_1(\omega)$ of the material under investigation (including linear materials) is sensitive: temperature, pressure, humidity, doping level, presence of impurities, etc.

3)For linear diffraction the first factor, which also has the functional form $\mathrm{sinc}(\pi b/\lambda)$, describes diffraction of a plane wave by one slit b, and the second describes interference of the fields of the two sources: $\cos(\pi d/\lambda)$. The phase of the interference pattern is determined by the distance between the slits d.

4)Both families of spectra were obtained using crossed dispersion.[28,4]

5)The first photograph of a frequency-angle spectrum for spontaneous parametric scattering under the conditions of nonlinear Mach–Zehnder interference in two crystals is shown in Ref. 20. The signal and idler photons belong to the visible range.

[1] Yu. N. Polivanov, Usp. Fiz. Nauk **126**, 185 (1978) [Sov. Phys. Usp. **21**, 86 (1978)].
[2] D. N. Klyshko, *Photons and Nonlinear Optics*, Gordon and Breach, New York (1988).
[3] G. D. Boyd and D. A. Kleinmann, J. Appl. Phys. **39**, 3597 (1968).
[4] D. N. Klyshko, A. N. Penin, and B. V. Polkovnikov, JETP Lett. **11**, 5 (1970).
[5] J. Chang, J. Ducuing, and N. Blombergen, Phys. Rev. Lett. **15**, 6 (1965).
[6] J. J. Wynne and N. Blombergen, Phys. Rev. **188**, 1211 (1969).
[7] Y. Yacoby, R. Fitzgibbon, and B. Lax, J. Appl. Phys. **51**, 3072 (1980).
[8] Jae Won Hahn and Eun Seony Lee, J. Opt. Soc. Am. B **12**, 1021 (1995).
[9] N. Blombergen and A. J. Sievers, Appl. Phys. Lett. **17**, 483 (1970); I. Freund, Phys. Rev. Lett. **21**, 1404 (1968).
[10] A. Szilagyi, A. Hordvik, and H. G. Schlossberg, Appl. Phys. **47**, 2025 (1975); M. Okada, T. Tarizana, and S. Ieiri, Opt. Commun. **18**, 331 (1976); Feng Duan, Ming Nai-Ben, Hong Jing-Fen, *et al.*, Appl. Phys. Lett. **37**, 607 (1980).
[11] G. Kh. Kitaeva, A. A. Mikhaĭlovskiĭ, and A. N. Penin, Zh. Éksp. Teor. Fiz. **112**, 2001 (1997) [JETP **85**, 1094 (1997)].
[12] G. Kh. Kitaeva, D. N. Klyshko, and I. V. Taubin, Kvantovaya Elektron. **9**, 561 (1982) [Sov. J. Quantum Electron. **12**, 333 (1982)].
[13] D. N. Klyshko, Zh. Éksp. Teor Fiz. **104**, 2676 (1993) [JETP **77**, 222 (1993)]; Zh. Éksp. Teor. Fiz **105**, 1574 (1994) [JETP **78**, 848 (1994)].
[14] O. A. Aktsipetrov, G. Kh. Kitaeva, E. D. Mishina, and A. N. Penin, Fiz. Tverd. Tela (Leningrad) **21**, 1834 (1979) [Sov. Phys. Solid State **21**, 1051 (1979)].
[15] A. L. Aleksandrovskiĭ, G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Zh. Éksp. Teor. Fiz. **90**, 1051 (1986) [Sov. Phys. JETP **63**, 613 (1986)]; G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Ferroelectrics **172**, 469 (1995).
[16] X. Y. Zou, L. J. Wang, and L. Mandel, Phys. Rev. Lett. **67**, 318 (1991); L. J. Wang, X. Y. Zou, and L. Mandel, Phys. Rev. A **44**, 4615 (1991).
[17] T. J. Herzog, J. G. Rarity, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **72**, 629 (1994).
[18] Z. Y. Ou, X. Y. Zou, L. J. Wang, and L. Mandel, Phys. Rev. Lett. **65**, 321 (1990); M. A. Horne, A. Shimony, and A. Zeilinger, Phys. Rev. Lett. **62**, 2209 (1989).
[19] A. V. Burlakov, M. V. Chekhova, D. N. Klyshko, S. P. Kulik *et al.*, Phys. Rev. A **56**, 3214 (1997).
[20] A. V. Burlakov, D. N. Klyshko, S. P. Kulik, A. N. Penin, and M. V. Chekhova, JETP Lett. **65**, 19 (1997).
[21] J. Mod. Opt. (Special issue on quantum communication) **41**, 12 (1994).
[22] R. Stolle, G. Marowsky, E. Schwarzberg, and G. Berkovic, Appl. Phys. B **63**, 491 (1996).
[23] O. A. Aktsipetrov, E. D. Mishina, T. V. Mitsuryaev, A. A. Nikulin *et al.*, submitted to Surf. Sci. (1998).
[24] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **68**, 2421 (1992); A. M. Steinberg and R. Y. Chiao, Phys. Rev. A **51**, 3525 (1995).
[25] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986).
[26] M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, Oxford, 1969).
[27] Yu. S. Kuz'minov, *The Electro-Optical and Nonlinear-Optical Crystal Lithium Niobate* [in Russian] (Nauka, Moscow, 1987).
[28] D. S. Rozhdestvenskiĭ, *Works on Anomalous Dispersion in Metal Vapors* [in Russian] (Izdat. Akad. Nauk SSSR, Moscow, 1951).
[29] M. V. Chekhova, A. V. Zhdanov, T. V. Laptinskaya, and A. N. Penin, Ferroelectr. Lett. Sect. **9**, 131 (1988).
[30] D. N. Nikogosyan, Kvantovaya Elektron. **4**, 5 (1977) [Sov. J. Quantum Electron. **7**, 1 (1977)].
[31] D. N. Klyshko, Usp. Fiz. Nauk **154**, 133 (1988) [Sov. Phys. Usp. **31**, 74 (1988)].
[32] U. Schlarb and K. Betzler, Phys. Rev. B **50**, 751 (1994).
[33] G. Kh. Kitaeva, I. I. Naumova, A. A. Mikhailovsky, P. S. Losevsky, and A. N. Penin, Appl. Phys. B **66**, 201 (1998).
[34] M. V. Chekhova and A. N. Penin, J. Raman Spectrosc. **24**, 581 (1993).
[35] A. L. Aleksandrovskiĭ, G. I. Ershova, G. Kh. Kitaeva, S. P. Kulik, I. I. Naumova, and V. V. Tarasenko, Kvantovaya Elektron. **18**, 254 (1991) [Sov. J. Quantum Electron. **21**, 225 (1991)].
[36] T. R. Volk and N. M. Rubinina, Fiz. Tverd. Tela (Leningrad) **33**, 1192 (1991) [Sov. Phys. Solid State **33**, 674 (1991)].

Translated by Paul F. Schippnick
Edited by P. Shelnitz

Influence of Coster–Kronig transitions on the polarization of *L*-shell x rays induced by proton impact

S. Gelfort and H. Kerkow

Institut für Physik der Humboldt-Universität Berlin, D-10115 Berlin, Germany

V. P. Petukhov and E. A. Romanovsky

Institute of Nuclear Physics, Moscow State University, 119899 Moscow, Russia

(Submitted 29 April 1997)
Zh. Éksp. Teor. Fiz. **113**, 2005–2010 (June 1998)

The alignment parameters for L_3-subshells of Cd and Sb atoms are obtained by measuring the degree of polarization of the L_1-lines excited by proton impact in the range from 0.1 to 0.5 MeV. To compare the experimental alignment parameter with theory, either experimental or numerical results must be corrected for Coster–Kronig (CK) transitions. The uncertainty in CK transition yields makes this comparison difficult. In this work, semiempirical values of the corrections have been derived from measured *L*-line intensities. The semiempirical correction factors exceed the theoretical ones. For constant reduced velocity, the semiempirical correction factor depends on the atomic number of the target. The semiempirical correction factor obtained in the same experiment improves the agreement between the theoretical alignment parameter and the experimental data. © *1998 American Institute of Physics.* [S1063-7761(98)00606-4]

1. INTRODUCTION

The alignment of atomic inner shells induced by ion impact has been the object of extensive theoretical and experimental investigations, because it provides a more sensitive test of theoretical models and atomic wave functions than the total excitation and ionization cross sections. Alignment leads to emission anisotropy and x-ray polarization, which can influence measurements of the x-ray production cross sections in ion–atom collisions. We have studied proton-induced cadmium and antimony L_3 subshell alignment by measuring L_1 x-ray line polarization.

To better test theory, an effort should be made to eliminate effects that hinder comparison of theory and experiment. To compare the experimental alignment parameter with theory, either experimental or numerical results must be corrected for Coster–Kronig (CK) transitions. Present knowledge of the CK rates f_{12}, f_{13}, and f_{23}, which describe the probability of vacancy transfer from L_1 to L_2 and L_3 subshells and from L_2 to L_3 subshells, respectively, is unsatisfactory, given the paucity and low accuracy of the experimental and numerical data. This is particularly true in the region of the periodic table around atomic number $Z=50$, where the onset of the L_1-L_2-$M_{4,5}$ and cutoff of the L_1-L_3-$M_{4,5}$ CK transitions occur.[1] The onset and cutoff of CK transitions at certain atomic numbers cause sharp discontinuities in the initial-state lifetimes as functions of atomic number. This results in a dramatic change in the associated yields.[1–3] Rosato's results[4] show no sharp discontinuities in the range $47 \leqslant Z \leqslant 53$. The exact location of these cutoffs is also somewhat uncertain.

The uncertainty in the Coster–Kronig yields makes an accurate comparison of experiment and theory difficult. To avoid these uncertainties, we derive the correction factor from measurements of the L x-ray line intensities.[5] In this work we also study the effect of the abrupt change of the CK transition probability around $Z=50$ on the measured alignment of L_3 subshell vacancies.

2. EXPERIMENT

The experimental work was carried out at the 0.5 MeV Cockcroft–Walton generator at the Institute of Nuclear Physics of Moscow State University. A Soller type (flat crystal) x-ray spectrometer–polarimeter was used to measure the intensity and polarization of the L x-ray lines. The experimental setup is described in detail elsewhere.[6] Protons of 0.1 MeV to 0.5 MeV energy were used. The collimated proton beam was stopped by a thick target. The target was oriented at 45° to the incoming beam. The incident beam intensity was monitored by a current integrator. The x rays emitted perpendicular to the beam axis were analyzed by the spectrometer, which was equipped with a graphite crystal ($2d = 6.71$ Å). The energy resolution of the spectrometer was $E/\Delta E = 600$.

A polarization experiment with the crystal spectrometer can be carried out, due to the linear polarization dependence of the crystal diffraction. The polarization P is defined by

$$P = \frac{1}{Q}\frac{(J_{\parallel} - J_{\perp})}{(J_{\parallel} + J_{\perp})}, \tag{1}$$

where Q is the polarization sensitivity of the crystal. The alignment parameter A_{20} can then be deduced from the polarization P,

$$A_{20}=\frac{2P}{\alpha(P-3)}, \tag{2}$$

where α is a constant determined by the angular momentum of the initial and final states.[7] In our experiment we used a thick target, so the experimental alignment parameter is corrected for the energy loss of protons and for the absorption of x rays in the target.[8] Errors in the alignment parameter are mainly due to errors in the determination of the L_1 and L_β yields, which are caused by statistical fluctuations, background substraction, and fitting procedures. The EWA code[9] was used to evaluate the spectra. As a rule, the statistical error in the measurements of the degree of polarization does not exceed 1%.

To determine the L_α x-ray production cross sections, we measured the L_α x-ray line yields as a function of incident particle energy. The x-ray production cross section can be determined from these measurements by the method and formulas described by Merzbacher and Lewis.[10] The proton stopping power was taken from Andersen and Ziegler.[11] The absorption coefficient was calculated by averaging the data of Storm and Israel.[12] The x-ray production cross sections were determined to 15% accuracy.

3. RESULTS AND DISCUSSION

The alignment parameter can be inferred from measurements of the proton-induced L_1 line polarization. To compare experimental and theoretical alignment parameters, we should take into account possible changes in the L subshell populations due to Coster–Kronig transitions (indirect formation of a vacancy in the L_3 subshell). In proton–atom collisions, vacancies are also produced in the L_1 and L_2 subshells, which then decay via Coster–Kronig transitions, so the number of L_3 vacancies increases. The alignment of L_1 and L_2 vacancies is zero, so that this two-step process leads to a decrease in L_3 vacancy alignment. The alignment parameter correction factor F is

$$A_{20}=FA_{20}^{\exp}, \tag{3}$$

$$F=1+f_{23}\frac{\sigma_2}{\sigma_3}+(f_{13}+f_{12}f_{23})\frac{\sigma_1}{\sigma_3}, \tag{4}$$

where σ_i is the L_i subshell ionization cross section, and f_{ij} are the Coster–Kronig yields.

The uncertainty in the Coster–Kronig yields makes it difficult to compare experiment and theory accurately. In order to eliminate this uncertainty, we determine the correction factor F from our measured L_α x-ray production cross sections. This cross section is defined as

$$\sigma_\alpha=\omega_3\frac{\Gamma_{\alpha1}+\Gamma_{\alpha2}}{\Gamma_{R_3}}\left(1+f_{23}\frac{\sigma_2}{\sigma_3}+(f_{13}+f_{12}f_{23})\frac{\sigma_1}{\sigma_3}\right)\sigma_3. \tag{5}$$

From this equation, the correction factor is

$$F_{\exp}=\frac{\Gamma_{R_3}}{\Gamma_{\alpha1}+\Gamma_{\alpha2}}\frac{\sigma_\alpha}{\omega_3\sigma_3}. \tag{6}$$

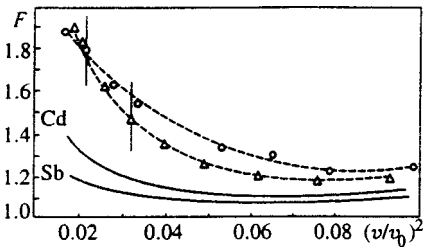


FIG. 1. Correction factor for CK transitions vs. reduced velocity. △—Cd, ○—Sb. Dashed lines are drawn through the experimental data to guide the eye. The solid lines are the result of the calculation.

As experimental data on ω_3 for elements around $Z=50$ are incomplete, we have taken the fluorescence yield ω_3 from Ref. 3 and the radiative transition rates $\Gamma_{\alpha1},\Gamma_{\alpha2},\Gamma_{R_3}$ from Ref. 13. Here we have not used the experimental values of σ_3, because these data are found with Eq. (5) based on measurements of L_α line intensities, using the CK yields, and the basic parameters (ω_i, f_{ij} and Γ) can strongly affect the final results. The ionization cross section σ_3 was calculated in terms of the so-called ECPSSR theory of Brandt and Lapicky,[14,15] which is a derivative of the plane wave Born approximation with corrections for ion energy loss effects (E), Coulomb repulsion (C), polarization and binding-energy changes via perturbed stationary states (PSS), and relativistic effects (R). Obviously, the uncertainty in the ω_3, which can change in the course of the collision, increases the uncertainty in the semiempirical correction factor. References 16 and 17 examine the influence of line shape and the satellite contribution on the interpretation of data in the measurements of x-ray spectra, and show that ignoring line shape effects can lead to systematic errors. In the present work it was assumed that satellite structure (in particular, the CK satellites) can be ignored in analyzing the spectra. It is clear that disregard of line shape effects and the fact that satellites contribute to polarization measurements can lead to systematic errors in measurements of the alignment parameter and in the semiempirical correction factor.

Figure 1 presents the semiempirical correction factors $F_{\exp}$ for Cd and Sb atoms obtained using Eq. (6) with the σ_α values measured in this experiment. The same figure shows the correction factors calculated with Eq. (4), using σ_i calculated within the ECPSSR theory and the CK yields taken from Ref. 3. It can be seen that the semiempirical correction $F_{\exp}$ exceeds its theoretical value. It is also seen that the correction factors depend on collision velocity $(v/v_0)^2$ (v is the proton velocity and v_0 is the Bohr velocity of L_3 electrons), and reach their minimum when the reduced velocity is about $(v/v_0)^2=0.1$. This minimum results from a minimum in the ratio σ_1/σ_3 of cross sections for ionization of L_1 and L_3 subshells of atoms.[18] The minimum in the σ_1/σ_3 ratio is due to the node of the $2s$-electron wave function. The discrepancy between $F_{\exp}$ and the theoretical value is partly due to the incorrect calculated ionization cross sections. Nevertheless, we believe that the semiempirical correction factor is more accurate, as it contains more accurate basic parameters.

In Fig. 2 we compare the experimental alignment param-

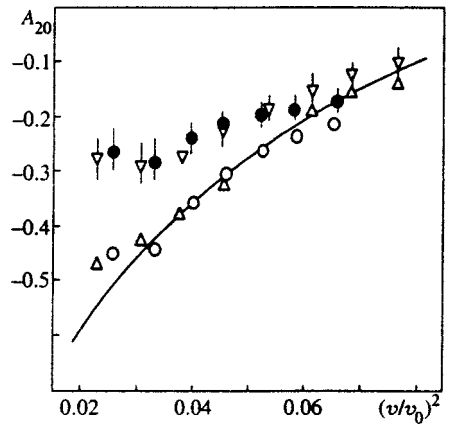


FIG. 2. L_3 subshell alignment parameter vs. reduced velocity. Experimental data: ▽—Cd, ●—Sb without correction; △—Cd, ○—Sb with correction for CK transitions. The curve presents the calculation in the Born approximation.

eter $A_{20}^{\exp}$ for Cd and Sb atoms with the alignment parameters corrected as described above, and A_{20} calculated in the Born approximation by the method and equations of Ref. 19. It can be seen that when the semiempirical correction factors $F_{\exp}$ are used, the agreement with the experimental data improves.

Based on measurements of L_α x-ray production cross sections, we determined the semiempirical correction factors for elements ranging from $Z=45$ to $Z=51$. The semiempirical correction factors obtained for reduced velocity $(v/v_0)^2=0.02$, together with the calculated correction factors, are presented in Fig. 3a as functions of the atomic number of the target. The decrease in the theoretical values of F around $Z=50$ is evident, and correlates with the decrease in f_{13} CK transition yield (see Fig. 3b). The f_{13} reduction is due to the fact that the L_1-L_3-$M_{4,5}$ CK transition becomes energetically forbidden. From this figure it is evident that there is no agreement of theoretical correction factors with experimental ones. The experimental results do not indicate any visible change in the correction factor around $Z=50$.

Summarizing, we conclude that accurately corrected values of the alignment parameter have to be introduced in order to reach good agreement between theoretical and experimental alignments. As the Coster–Kronig yield can change in the course of the collision (post-collision interactions, multiple ionization, chemical effects), the semiempirical correction factor obtained in the same experiment that measured the alignment parameter is a more realistic one. But even so, the uncertainty due to errors in the L_3 subshell fluorescence yield and ionization cross sections remains. The Z dependence of the correction factor is smooth, and the experimental results do not show sharp discontinuities anywhere within the range $47 \leqslant Z \leqslant 51$.

It is a pleasure for us to thank Dr. N. M. Kabachnik for calculations and Dr. T. Papp for helpful discussions and comments.

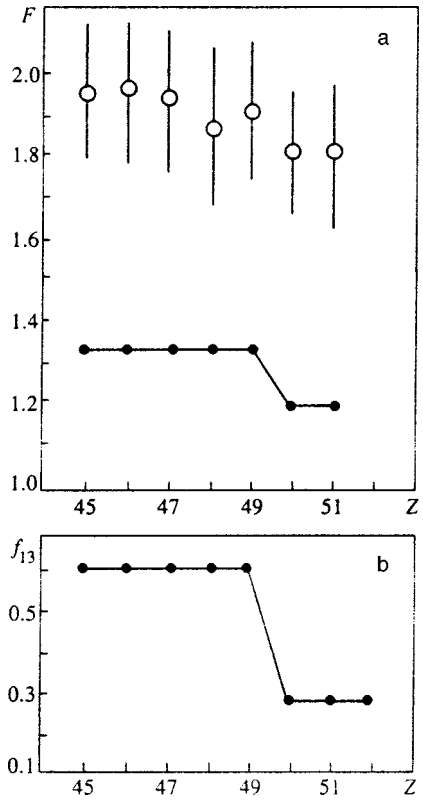


FIG. 3. a) Correction factor for CK transitions as a function of the atomic number of the target; upper part—experiment, lower part—result of the calculations. b) CK probabilities f_{13} (from Ref. 3) as a function of the atomic number.

[1] W. Bambynek, B. Crasemann, R. W. Fink, H. V. Freud, C. D. Swift, R. E. Price, and P. V. Rao, Rev. Mod. Phys. **44**, 716 (1972).
[2] E. J. McGuire, Phys. Rev. A **3**, 587 (1971).
[3] M. O. Krause, J. Phys. Chem. Ref. Data **8**, 307 (1979).
[4] E. Rosato, Nucl. Instr. Meth. B **15**, 591 (1986).
[5] V. P. Petukhov, E. A. Romanovsky, and H. Kerkow, Nucl. Instr. Meth. B **109/110**, 19 (1996).
[6] V. P. Petukhov, Prib. Tekh. Éksp. **1**, 195 (1990).
[7] E. G. Berezhko and N. M. Kabachnik, J. Phys. B **10**, 2467 (1977).
[8] G. Richter, Doctoral Thesis, Universität Munster (1982).
[9] J. Vegh, Thesis, ATOMKI, Debrecen (1990).
[10] E. Merzbacher and H. W. Lewis, Handbuch der Physik, Berlin (1958), vol. 34, p. 166.
[11] H. H. Andersen and J. F. Ziegler, *The Stopping and Ranges of Ions in Matter*, Pergamon, New York (1977), vol. 3.
[12] E. Storm and H. I. Israel, Nucl. Data Tables A **7**, 565 (1970).
[13] J. H. Scofield, At. Data Nucl. Data Tables **14**, 121 (1974).
[14] W. Brandt and G. Lapicki, Phys. Rev. A **20**, 465 (1979).
[15] W. Brandt and G. Lapicki, Phys. Rev. A **23**, 1717 (1981).
[16] T. Papp, J. L. Campbell, and S. Raman, Phys. Rev. A **49**, 729 (1994).
[17] J. L. Campbell and T. Papp, X-Ray Spectrom. **24**, 307 (1995).
[18] V. P. Petukhov, V. S. Nikolaev, E. A. Romanovsky, and V. A. Sergeev, Zh. Éksp. Teor. Fiz. **71**, 968 (1976) [Sov. Phys. JETP **44**, 508 (1976)].
[19] V. V. Sizov and N. M. Kabachnik, J. Phys. B **13**, 1601 (1980).

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

Interaction of two spatially separated light beams in a nonlinear Kerr medium

A. S. Desyatnikov and A. I. Maĭmistov*)

Moscow Engineering Physics Institute, 115409 Moscow, Russia
(Submitted 27 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 2011–2021 (June 1998)

The evolution of two spatially separated light beams in a nonlinear Kerr medium described by a system of coupled nonlinear Schrödinger equations is studied. An analytic solution is found for the variational problem. It is shown that when two crossed beams interact, a bound state can develop in which the distance between the centers of the beams and their radii vary periodically. Here the mutual curvature of the trajectories of the centers of the beams causes the beams to bend into a helical structure whose parameters (pitch and diameter) are also periodic functions. The threshold power for mutual trapping is determined and the period of the oscillations is found. © *1998 American Institute of Physics.* [S1063-7761(98)00706-9]

1. INTRODUCTION

The evolution of a light beam in a nonlinear dispersive medium is a typical problem in nonlinear optics and has been studied now for two or three decades. But interest in this problem has not flagged. An example of this continued interest is research on the formation of small-scale transverse structures in the electromagnetic field (diffraction autosolitons), dislocations in the wave front, and the interaction of these dislocations.[1–4] When a broadband interferometer is excited by radiation of this type, the autosoliton is scattered on the dislocation and it is trapped, with the buildup of rotation about the dislocation. In a semi-infinite nonlinear medium, the same effect leads to a rotation of the electromagnetic field structures in space as the radiation propagates into the medium.[3]

In the initial stage of these studies attention was focused on studies of a single beam, but recently, systems of two beams have been studied.[5–9] There are a number of papers[10–14] that examine bound states in which pulses with different polarization states propagate in the form of two-component (vector) solitons. The spacetime analogy makes it possible to transfer the results obtained for pulses to a description of a noncollinear interaction of wave beams.[15] It is natural to expect that, besides the well known effects of self-focusing and beam bending, when the self- and cross-modulation effects are combined the interaction of two spatially separated light beams can lead to a mutual bending of the ray trajectories and mutual trapping. An effect of this sort, referred to as the ''entrainment'' of light beams, was recently predicted theoretically and studied experimentally using second harmonic generation in a quadratic medium.[7,8]

It should be noted that the interaction of noncollinear light beams in a Kerr (focusing and defocusing) medium has been studied in a two-dimensional geometry,[16,17] where it was assumed that the beams propagate either parallel or antiparallel to one another. A case in which a beam is reflected from a mirror and Snell's law is violated was also examined.

The many studies of spatial solitons in photorefractive media[18–24] are another example of the active research in progress on the interaction between beams of electromagnetic radiation.

In this paper we examine the interaction (owing to cross-modulation of the refractive index) of two spatially separated light beams in a nonlinear Kerr medium. To simplify the problem it is assumed that although the radii of the beams vary, they remain identical all the time. As in orthogonally polarized coaxial beams,[9] monotonic and oscillatory propagation regimes are found. In the case of crossed (noncoplanar) beams, the oscillatory regime (mutual trapping) is characterized by a spatial rotation of the beams, with the angular rotation velocity which determines the pitch and diameter of the helices being a periodic function of the propagation variable.

In Kerr media, beams are unstable: they either experience collapse or they spread out. The analysis of beam propagation considered here assumes that their power differs slightly, on the low side, from the threshold power required for self-focusing of an isolated beam. The estimates given in the Conclusion show that it is possible to choose beam parameters such that the diffraction spread of the beams is a slow process compared to helix formation.

2. DERIVATION OF THE BASIC EQUATIONS

In an isotropic Kerr medium with an instantaneous response, the refractive index can be assumed to depend linearly on the intensity in many cases. For an incoherent superposition of the two waves

$$\mathbf{E}_1=\mathbf{e}_1 u(x,y,z)\exp\{ik\cos(\varphi_1 z)+i\mathbf{k}_{\perp 1}\cdot\mathbf{r}_{\perp 1}-i\omega t\}+\text{c.c.},$$

$$\mathbf{E}_2=\mathbf{e}_2 v(x,y,z)\exp\{ik\cos(\varphi_2 z)+i\mathbf{k}_{\perp 2}\cdot\mathbf{r}_{\perp 2}-i\omega t\}+\text{c.c.},$$

the permittivity for wave i can be written in the form[15]

$$\varepsilon_i=\varepsilon^{(0)}+\varepsilon^{(2)}|\mathbf{E}_i|^2+\varepsilon_c^{(2)}|\mathbf{E}_j|^2,\quad i,j=1,2,\quad i\neq j.$$

Here $k=(\omega/c)\sqrt{\varepsilon^{(0)}}$, $\mathbf{e}_i$ are the unit vectors of the waves, φ_i are the angles of incidence of the beams at the boundary $z=0$, and $\varepsilon^{(2)}$ and $\varepsilon_c^{(2)}$ are nonlinear coefficients corresponding to self- and cross-modulation, respectively. In the following we assume that the directions of beam propagation deviate little from the z axis—so little that the factors $\exp(i\mathbf{k}_{\perp i}\cdot\mathbf{r})$ can be included in the slowly varying envelopes $u(x,y,z)$ and $v(x,y,z)$. We assume that the region within which the electric field varies in the transverse direction is much smaller than the region within which it varies in the longitudinal direction. Following the standard procedure for constructing an evolution equation for the slowly varying envelope, one can obtain a system of coupled nonlinear Schrödinger equations:

$$i\,\frac{\partial u}{\partial z}+\sigma\nabla_\perp^2 u+\mu(|u|^2+\varepsilon|v|^2)u=0,$$

$$i\,\frac{\partial v}{\partial z}+\sigma\nabla_\perp^2 v+\mu(|v|^2+\varepsilon|u|^2)v=0, \tag{1}$$

where

$$\sigma=\frac{1}{2k\,\cos\,\varphi_1}\approx\frac{1}{2k\,\cos\,\varphi_2},\quad \mu=\frac{k\varepsilon^{(2)}}{2\varepsilon^{(0)}},$$

$$\varepsilon=\frac{\varepsilon_c^{(2)}}{\varepsilon^{(2)}},\quad \nabla_\perp^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}.$$

It can be shown[14,25–27] that the system of Eqs. (1) has the same form as the Euler equations generated by the variational problem $\delta S=0$, where the action functional is defined as

$$S=\int_{-\infty}^{\infty}\langle L\rangle dz \tag{2}$$

with the lagrangian

$$\langle L\rangle=\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}L\,dy,\quad L=L_u+L_v+L_{uv},$$

where

$$L_u=\frac{i}{2}\left(u^*\,\frac{\partial u}{\partial z}-u\,\frac{\partial u^*}{\partial z}\right)-\sigma\left|\frac{\partial u}{\partial x}\right|^2-\sigma\left|\frac{\partial u}{\partial y}\right|^2+\frac{\mu}{2}\,|u|^4,$$

$$L_v=L_u(u\to v),\quad L_{uv}=\mu\varepsilon|u|^2|v|^2.$$

We seek a solution of the variational problem in the class of trial functions of the form

$$u(x,y,z)=A_1(z)\exp\left(-\frac{|\mathbf{r}-\mathbf{r}_1(z)|^2}{2a^2(z)}+i\phi_1(x,y,z)\right),$$

$$v(x,y,z)=A_2(z)\exp\left(-\frac{|\mathbf{r}-\mathbf{r}_2(z)|^2}{2a^2(z)}+i\phi_2(x,y,z)\right), \tag{3}$$

$$\phi_i(x,y,z)=B(z)|\mathbf{r}-\mathbf{r}_i(z)|^2+C_i(z)[x-x_i(z)]+D_i(z)[y-y_i(z)]+\alpha_i(z),$$

where $\mathbf{r}=\mathbf{e}_x x+\mathbf{e}_y y$, and $\mathbf{r}_i(z)=\mathbf{e}_x x_i(z)+\mathbf{e}_y y_i(z)$ are the radius vectors of the beam centers. The choice of Gaussian trial functions was justified previously[26]; furthermore, it is easier to calculate $\langle L_{uv}\rangle$ for these trial functions. We note also that the phases $\phi_i(x,y,z)$ account for the transverse components of the wave vectors of the carrier wave in the initial representation of the electric field strengths of the two waves.

Substituting Eq. (3) in Eq. (2) and calculating the variational derivatives yields equations for the extremum of the action (2) ($i,j=1,2;\ i\neq j$):

$$a^2(z)A_i^2(z)=a^2(0)A_i^2(0)=E_i, \tag{4a}$$

$$\frac{da(z)}{dz}=4\alpha a(z)B(z). \tag{4b}$$

$$2\sum_{k=1}^{2}\left[-2a^2\left(\frac{dB}{dz}+4\sigma B^2\right)+C_k\left(\frac{dx_k}{dz}-\sigma C_k\right)+D_k\left(\frac{dy_k}{dz}-\sigma D_k\right)-\frac{d\alpha_k}{dz}+\frac{\mu}{4}\,A_k^2\right]A_k^2+\mu\varepsilon A_1^2A_2^2\left(1+\frac{R^2}{2a^2}\right)\exp\left(-\frac{R^2}{2a^2}\right)=0, \tag{4c}$$

$$2\left[-a^2\left(\frac{dB}{dz}+4\sigma B^2\right)-\frac{\sigma}{a^2}+C_i\left(\frac{dx_i}{dz}-\sigma C_i\right)\right]+2\left[D_i\left(\frac{dy_i}{dz}-\sigma D_i\right)-\frac{d\alpha_i}{dz}+\frac{\mu}{2}\,A_i^2\right]+\mu\varepsilon A_j^2\exp\left(-\frac{R^2}{2a^2}\right)=0; \tag{4d}$$

$$\frac{dx_i}{dz}=2\sigma C_i,\quad \frac{dy_i}{dz}=2\sigma D_i, \tag{5a}$$

$$\frac{dC_i}{dz}=(-1)^iA_j^2(x_1-x_2)\,\frac{\mu\varepsilon}{2a^2}\,\exp\left(-\frac{R^2}{2a^2}\right),$$

$$\frac{dD_i}{dz}=(-1)^iA_j^2(y_1-y_2)\,\frac{\mu\varepsilon}{2a^2}\,\exp\left(-\frac{R^2}{2a^2}\right), \tag{5b}$$

Here $R(z)\equiv|\mathbf{R}(z)|=|\mathbf{r}_1(z)-\mathbf{r}_2(z)|$.

After some transformations, Eqs. (4) reduce to an equation for the radius $a(z)$ of the beams:

$$\frac{d^2a}{dz^2}=\frac{\sigma}{a^3}\left[4\sigma-\mu W(1-\delta)-\mu\varepsilon W\delta\left(1-\frac{R^2}{2a^2}\right)\exp\left(-\frac{R^2}{2a^2}\right)\right], \tag{6}$$

where $W=E_1+E_2$ is the total power of the beams and $\delta=2E_1E_2(E_1+E_2)^{-2}$ is a dimensionless quantity characterizing the ratio of the beam powers, with $0<\delta\leqslant 0.5$.

The self-interaction of beams in a focusing Kerr medium is determined[15] by the competition between two factors: diffraction (first term on the right-hand side of Eq. (6)) and nonlinear compression (second term). The term (proportional to ε) corresponding to the interaction between beams in Eq. (6) can be regarded as a measure (strength) of the influence of one beam on the other. When the variables change in the region $R<\sqrt{2}a$, this force increases the nonlinear compres-

sion, and when they change in the region $R>\sqrt{2a}$, it weakens it. In particular, as will be shown below, the radii can oscillate (like the oscillations found in Ref. 9), which is a qualitative difference from the case of self-action.

Using the expression $\mathbf{R}(z)=[x_1(z)-x_2(z)]\mathbf{e}_x+[y_1(z)-y_2(z)]\mathbf{e}_y$, we can reduce Eqs. (5) to the form

$$\frac{d^2\mathbf{R}}{dz^2}=-\mathbf{R}\,\frac{\sigma\mu\varepsilon W}{a^4}\exp\left(-\frac{R^2}{2a^2}\right). \tag{7}$$

Equation (7) makes it possible to find an integral of the motion,

$$\mathbf{R}\times\frac{d\mathbf{R}}{dz}=M\mathbf{e}_z, \quad M=\text{const}. \tag{8}$$

With Eq. (8), Eq. (7) yields

$$\frac{d^2R}{dz^2}=\frac{M^2}{R^3}-R\,\frac{\sigma\mu\varepsilon W}{a^4}\exp\left(-\frac{R^2}{2a^2}\right). \tag{9}$$

Thus, the system of first-order equations for the parameters of the trial functions reduces to a system of two second-order equations (6) and (9) for the radius $a(z)$ and the distance $R(z)$ between the beam centers.

Let us introduce polar coordinates in the x,y plane,

$$R_x(z)\equiv x_1(z)-x_2(z)=R(z)\cos\,\psi(z),$$

$$R_y(z)\equiv y_1(z)-y_2(z)=R(z)\sin\,\psi(z),$$

which then yield an equation for $\psi(z)$ from Eq. (8):

$$\frac{d\psi}{dz}=\frac{M}{R^2(z)}. \tag{10}$$

The polar angle $\psi(z)$ amounts to the angle of rotation of the vector $\mathbf{R}(z)$ joining the beam centers. When the beams do not interact and are parallel ($M=0$), $\mathbf{R}(z)$ is a constant vector. When the beams do not interact and are noncoplanar, this vector varies in direction and magnitude, but the overall angle of rotation of $\mathbf{R}(z)$ in the xy plane does not exceed π. The bending of the beams owing to their interaction changes the maximum value of the overall angle of rotation; for example, if this angle should exceed 2π, then we can speak of helical bending of the beams.

Before proceeding to solve the system of Eqs. (6) and (9), we can simplify the problem, without loss of generality, by making the following choice of boundary conditions:

1. Let the coordinate origin in the plane $z=0$ on the segment joining the beam centers be chosen from the condition $E_1\mathbf{r}_1(0)+E_2\mathbf{r}_2(0)=0$ and the direction of the x axis be chosen to be along $\mathbf{R}_0\equiv\mathbf{R}(0)$; then $\psi(0)=0$.

2. Choose the parallel planes in which the axes of the beams incident upon the boundary $z=0$ to be perpendicular to the boundary $z=0$; then these planes become the planes of incidence of the beams.

3. Choose the angles of incidence of the beams, φ_1 and φ_2, to be such that $E_1\tan\varphi_1=E_2\tan\varphi_2$.

The physical significance of the vectors $\dot{\mathbf{r}}_1(0)$ and $\dot{\mathbf{r}}_2(0)=0$ (the initial values of the derivatives of $\mathbf{r}_{1,2}$ with respect to z) is clear from the equations $|\dot{\mathbf{r}}_1(0)|=\tan\varphi_1$ and $|\dot{\mathbf{r}}_2(0)|=\tan\varphi_2$. We choose the direction of incidence of the beams so that when the second and third conditions are satisfied, the equation $E_1\dot{\mathbf{r}}_1(0)+E_2\dot{\mathbf{r}}_2(0)=0$ is satisfied. Then an explicit expression for the constant M can be written in the form

$$M=R_0R_0'\tan\beta=R_0(\tan\varphi_1+\tan\varphi_2)\sin\beta, \tag{11}$$

where $R_0'\equiv dR/dz|_{z=0}=(\tan\varphi_1+\tan\varphi_2)\cos\beta$, with β being the angle between the initial directions of $\mathbf{R}$ and the derivative of $\mathbf{R}$ with respect to z. To enlarge the beam interaction region, it is necessary to put $R_0'\leqslant 0$, i.e., $\beta\geqslant\pi/2$, in the following.

3. SOLUTION OF THE EQUATIONS OF THE VARIATIONAL PROBLEM

The system of Eqs. (6) and (9) has two conservation laws:

$$Wa^2(z)+mR^2(z)=P_2z^2+P_1z+P_0, \tag{12}$$

$$P_2=W\left(\frac{da}{dz}\right)^2+m\left(\frac{dR}{dz}\right)^2+\frac{\sigma W[4\sigma-\mu W(1-\delta)]}{a^2}+\frac{mM^2}{R^2}-\frac{\sigma\mu\varepsilon W^2\delta}{a^2}\exp\left(-\frac{R}{2a^2}\right), \tag{13}$$

where $m=\delta W/2=E_1E_2(E_1+E_2)^{-1}$, $P_1=2Wa_0a_0'+2mR_0R_0'$, $a_0=a(0)$, and $P_0=Wa_0^2+mR_0^2$.

Let us introduce the new variables

$$Wa^2(z)=\rho^2(z)t(z), \quad mR^2(z)=\rho^2(z)(1-t(z)). \tag{14}$$

In terms of the new variables, Eqs. (12) and (13) take the form

$$\rho^2(z)=P_2z^2+P_1z+P_0, \tag{15}$$

$$P_2=\left(\frac{d\rho}{dz}\right)^2+\frac{\rho^2}{4t(1-t)}\left(\frac{dt}{dz}\right)^2+\frac{\sigma W^2[4\sigma-\mu W(1-\delta)]}{\rho^2t}+\frac{m^2M^2}{\rho^2(1-t)}-\frac{\sigma\mu\varepsilon W^3\delta}{\rho^2t}\exp\left(\frac{t-1}{\delta t}\right). \tag{16}$$

By separating variables in Eq. (16), we can obtain a general analytic solution of Eqs. (6) and (9):

$$2\int_0^z\frac{dz}{p_2z^2+p_1z+p_0}=\int\frac{dt}{\sqrt{G(t)}}, \tag{17}$$

$$G(t)=st(t_0-t)+(1-t)\left[\Delta(t_0-t)+\lambda t+\exp\left(\frac{t-1}{\delta t}\right)-\frac{t}{t_0}\exp\left(\frac{t_0-1}{\delta t_0}\right)\right], \tag{18}$$

where, according to Eqs. (11) and (13),

$$s=\frac{a_0^2(R_0')^2\tan^2\beta}{2\sigma\mu\varepsilon Wt_0}, \quad t_0\equiv t(0)=\frac{Wa_0^2}{Wa_0^2+mR_0^2},$$

$$\Delta=\frac{\mu W(1-\delta)-4\sigma}{\mu\varepsilon W\delta t_0}, \quad \lambda=\frac{(a_0R_0'-r_0a_0')^2}{2\sigma\mu\varepsilon W},$$

$$p_j=\frac{P_j}{\sqrt{\sigma\mu\varepsilon W^3\delta}},\quad j=0,1,2. \tag{19}$$

The description of the dynamics of the system becomes complete after integrating Eq. (10):

$$\psi(z)=\frac{M}{4}\sqrt{\frac{\delta}{\sigma\mu\varepsilon W}}\int\frac{dt}{(1-t)\sqrt{G(t)}}, \tag{20}$$

where we have used Eqs. (14) and (17).

Equations (17) and (20) are an exact solution of the variational problem. However, Eqs. (17) and (20) are unsuitable for further analytic study, since the integrals on the right-hand side with the expression (18) under the radical sign cannot be expressed in terms of any known functions.

In the next section, a qualitative analysis of Eqs. (17) and (20) is used to obtain the general features of the monotonic propagation regime for spatially separated light beams, with principal attention devoted to the oscillatory regime. Approximate solutions of Eqs. (17) and (20) are obtained that describe the bound state (oscillatory regime), and the range of parameters over which this state is realized is determined analytically.

4. ANALYSIS OF THE SOLUTION

A study of the integrals in Eqs. (17) and (20) shows that the beam dynamics are qualitatively determined by the behavior of the function $\rho^2(z)$ and the zeroes of the functions $G(t)$ in Eq. (18).

If the total power W of the beams is less than the critical value $W_{cr}=4\sigma[\mu(1-\delta)]^{-1}$, then the parameter $\Delta<0$, the zeroes t_1 and t_2 of $G(t)$ lie within the limits $0<t_2<t_0<t_1<1$ and, since $P_2>0$ in this case, the beams undergo monotonic diffraction spreading. Here the interaction leads to a negligible mutual deflection of the beams.

When the total power exceeds the critical value, $\Delta>0$ and one of the zeroes of $G(t)$ becomes negative. If $t_0\in(0,t_1)$, where t_1 is a zero of $G(t)$, then, according to Eq. (14), upon reaching $t=0$ the light beams collapse to a finite distance z_{coll} determined from the condition $t(z_{coll})=0$.

These results are consistent with the well known self-focusing behavior of light beams, so in the following we concentrate on studying the oscillatory propagation regime and on determining the spatial rotation parameters of the beams. The above discussion implies that in principle, $t(z)$ can vary periodically only when $G(t)$ has three zeroes in the interval (0,1) and the initial value t_0 lies within an interval that does not contain the point $t=0$. If the total power is much higher than the critical value, so that $P_2<0$, then the function $\rho^2(z)$ goes to zero at a finite distance, which corresponds to collapse and ''adhesion'' of the beams. Thus, there is yet another condition for the existence of a periodic solution, $\rho^2=\text{const}$.

The condition $\rho^2=\text{const}$ is equivalent to the three conditions $P_1=0$, $P_2=0$, and $\rho^2=P_0$. Thus,

$$P_1=0\rightarrow a_0'=-\frac{\delta}{2}\frac{R_0}{a_0}R_0', \tag{21}$$

and since it was assumed that $R_0'\leqslant 0$, it is necessary to set $a_0'\geqslant 0$, i.e., at the boundary of the medium the beams must have positive wavefront curvature. This implies external defocusing of the beams.

Using Eq. (21), we rewrite Eq. (19) for s as

$$s=\lambda t_0\tan^2\beta=\frac{a_0^2(R_0')^2\tan^2\beta}{2\sigma\mu\varepsilon W t_0}, \tag{22}$$

so that the condition $P_2=0$ can be written explicitly as

$$W=W_{thr}\equiv W_{cr}\left[1+\frac{\delta a_0^2(\tan\varphi_1+\tan\varphi_2)^2}{8\sigma^2}\times\left(1+\frac{1-t_0}{t_0}\cos^2\beta\right)\right]. \tag{23}$$

Equation (23) determines the threshold power W_{thr} at which mutual trapping of the beams takes place, since $W<W_{thr}\rightarrow P_2>0$ and $W>W_{thr}\rightarrow P_2<0$, which correspond to the cases discussed above.

Now we have to find the relationships among the parameters such that $G(t)$ will have three zeroes $t_{1,2,3}$ with $0<t_3<t_2<t_1<1$ and $t_0\in(t_2,t_1)$. Given Eqs. (21), (22), and (23), Eq. (18) for $G(t)$ transforms to

$$G(t)=s(t_0-t)+(1-t)\left[\lambda t_0+\exp\left(\frac{t-1}{\delta t}\right)-\frac{t}{t_0}\exp\left(\frac{t_0-1}{\delta t_0}\right)\right]. \tag{24}$$

An analysis of this expression shows that the desired situation occurs only for $\delta>0.4$ (which corresponds to E_1/E_2 (or E_2/E_1)<2), i.e., the beam powers must be similar. When this condition is satisfied, Eq. (24) can be well approximated by the polynomial

$$G_a(t)=\lambda t_0(1-t)+(t_0-t)[s-t^2(1-t)], \tag{25}$$

for which, as opposed to the exact expression (24), it is possible to determine the desired range of the parameters s, t_0, and β analytically.

Let the inequalities

$$s<\frac{2}{27},\quad t_0>t^*,\quad \tan^2\beta>\frac{1}{t_0-t'}\left(\frac{t'}{s}-\frac{1}{1-t'}\right)^{-1}, \tag{26}$$

be satisfied, where t^* is the root of the trinomial $s-t^2(1-t)$ closest to the zero from the right, and t' corresponds to the local minimum of the function $(t_0-t)[s-t^2(1-t)]$. Then the fourth-degree polynomial $G_a(t)$ (25) has four real roots $t_{1,2,3,4}$ with $t_4<0<t_3<t_2<t_0<t_1<1$.

Note that the lower bound on t_0 leads to an upper bound on the ratio R_0/a_0:

$$\frac{R_0}{a_0}=\sqrt{\frac{2(1-t_0)}{\delta t_0}}<\sqrt{\frac{2(1-t^*)}{\delta t^*}}.$$

The condition for applicability of the trial function method in problems of this type is that the beams be far from one another. In order for the beams to be regarded as different, it is necessary to set $R_0>2a_0$, which imposes an upper

bound on t_0: $t_0<(1+2\delta)^{-1}$. At the end of this section, we present numerical estimates showing that both conditions can be satisfied. We note additionally that the last inequality in Eq. (26) can always be satisfied by choosing the angle β to be close enough to $\pi/2$. Equations (21), (22), (23), and (26) determine the parameter domain within which the oscillatory regime occurs.

Replacing the exact expression in Eqs. (17) and (20) by the approximation $G_a(t)$, we obtain an approximate solution of the problem,

$$R(z)=R_0\sqrt{\frac{1-t(z)}{1-t_0}},\quad a(z)=a_0\sqrt{\frac{t(z)}{t_0}},$$

$$\psi(z)=\frac{\sqrt{s(1-t_0)}}{1-t_3}\left\{F(\varphi(z),k)-F(\varphi_0,k)\right.$$

$$\left.+\frac{t_2-t_3}{1-t_2}\left[\Pi(\varphi(z),\nu,k)-\Pi(\varphi_0,\nu,k)\right]\right\}. \qquad (27)$$

Here

$$t(z)=\frac{t_2(t_1-t_3)-t_3(t_1-t_2)\mathrm{sn}^2\,w(z)}{(t_1-t_3)-(t_1-t_2)\mathrm{sn}^2\,w(z)},$$

$$w(z)=F(\varphi_0,k)+\frac{\sqrt{(t_1-t_3)(t_2-t_4)}}{p_0}z,$$

$\varphi(z)=\mathrm{am}(w(z))$ is the amplitude and $\mathrm{sn}(w(z))$ is the Jacobi elliptic sine; $F(\varphi,k)$ and $\Pi(\varphi,\nu,k)$ are the incomplete Legendre elliptic integrals of the first and third kinds, respectively; and,

$$k^2=\frac{t_1-t_2}{t_1-t_3}\frac{t_3-t_4}{t_2-t_4},\quad \nu=\frac{t_1-t_2}{t_1-t_3}\frac{1-t_3}{1-t_2},\quad \varphi(t_2)=0.$$

$$\varphi(t_0)=\varphi_0\equiv\sin^{-1}\sqrt{\frac{t_1-t_3}{t_1-t_2}\frac{t_0-t_2}{t_0-t_3}},\quad \varphi(t_1)=\frac{\pi}{2}.$$

The approximate solution (27) determines the dynamics of the system approximately: the distance between the beam centers (helix diameter) varies periodically between $R_{\min}=R_0\sqrt{(1-t_1)/(1-t_0)}$ (here $a_{\max}=a_0\sqrt{t_1/t_0}$) and $R_{\max}=R_0\sqrt{(1-t_2)/(1-t_0)}$ ($a_{\min}=a_0\sqrt{t_2/t_0}$), while the angle of rotation increases nonuniformly over a period z_p by an amount ψ_p, i.e., the pitch of the helix is also a periodic function. For numerical estimates of the parameters z_p and ψ_p of the periodic structure, we can use the exact expressions

$$z_p=p_0\int_{t_{\min}}^{t_{\max}}\frac{dt}{\sqrt{G(t)}},$$

$$\psi_p=\sqrt{s(1-t_0)}\int_{t_{\min}}^{t_{\max}}\frac{dt}{(1-t)\sqrt{G(t)}}. \qquad (28)$$

Here $t_{\max}$ and $t_{\min}$ are the zeroes of $G(t)$ (Eq. (24)) corresponding to the roots t_1 and t_2 of the polynomial $G_a(t)$ (25).

Using the procedure with Eqs. (21)–(23), (26), and (28) described above, we can optimize the given parameters. Taking $s\simeq 0.03$, $t_0\simeq 0.3$, and $\beta\simeq 100°$, we obtain $\psi_p\simeq\pi$, i.e., a complete turn ($\psi=2\pi$) occurs over a distance $z_{2\pi}\simeq 2z_p$. The integration is over limits from $t_{\min}\simeq 0.288$ ($t_2\simeq 0.271$) to $t_{\max}\simeq 0.978$ ($t_1\simeq 0.968$); the variables vary over the limits $1.01\geqslant R/R_0\geqslant 0.18$ and $0.98\leqslant a/a_0\leqslant 1.81$ for an initial $R_0\simeq 3.3a_0$. The magnitude of z_p for given s, t_0, and β is determined by the magnitude of a_0 and is ~ 10 cm for $a_0\sim 10$ μm. The initial divergence of the beams (external defocusing) a_0' is of order 10^{-3}, which corresponds to diffractive divergence for Gaussian beams. For estimates, we set the coefficients ε and $\varepsilon^{(0)}$ equal to unity and the frequency of the carrier wave to $\nu\approx 10^{15}$ s^{-1}.

5. CONCLUSION

In this paper we have used a variational technique to study the mutual effect of two spatially separated light beams owing to cross-modulation of the refractive index in a cubic nonlinear medium. The interaction is attractive in nature and causes the ray trajectories to bend. Under the conditions determined here, crossed beams are trapped into a bound state, which causes them to rotate in space. Since the rate of rotation is an oscillatory function, the spatial parameters of the helical structure (pitch, diameter) vary periodically.

In our model, an oscillatory regime that is not limited by collapse will be realized for beam powers that are equal or close in magnitude. This property is evidently a consequence of our assumption that the beam radii behave in the same way, i.e., it is inherent only to the model chosen here. If we assume that the beam radii are different in the trial functions for the variational method, then instead of the system of two equations, (6) and (9), we obtain a system of three second-order equations for the radii and the distances between the beam centers. No analytic solution can be found for that system of equations.

We have not dwelt here on the change in phase of both of the beams, although these quantities are taken into account in the trial functions (3) used here. The variational problem yields equations for the evolution of the phases ϕ_1 and ϕ_2 along the z axis and in the xy plane, which are not coupled to the equations (4) considered here. However, given the solution of the system of Eqs. (4), it would be possible to calculate ϕ_1 and ϕ_2 by direct integration. This problem might turn out to be extremely attractive if we are interested in the topological properties of the wave front, as for example in a study of dislocations in a wave front.[4,5,28–30] The difference in phase of neighboring beams is extremely important when they propagate in quadratic media,[7,8] where Eqs. (4) acquire additional terms that account for the parametric interaction of the beams.

The interest in spatial optical structures is based on the possibility, in principle, of using them for information processing.[31,32] The rotation of optical beams considered here can obviously be used in simple optical switching schemes, since it does not require the creation of special conditions (e.g., feedback) to operate. Further studies of the model developed here with differing beam radii will be useful from the standpoint of studying the mutual effect of beams on collapse, or from a more general standpoint, of the control of light by light.[33]

*)E-mail: maimistov@pico.mephi.ru

[1] N. N. Rozanov and V. A. Smirnov, Zh. Éksp. Teor. Fiz. **70**, 2060 (1976) [Sov. Phys. JETP **43**, 1075 (1976)].
[2] N. N. Rozanov, Opt. i Spektrosk. **72**, 447 (1992) [Opt. Spectrosc. **72**, 243 (1992)].
[3] N. N. Rozanov, A. V. Fedorov, S. V. Fedorov, and G. V. Khodova, Zh. Éksp. Teor. Fiz. **107**, 376 (1995) [JETP **80**, 199 (1995)].
[4] M. Ya. Darsht, B. Ya. Zel'dovich, I. V. Kataevskaya, and I. D. Kundikova, Zh. Éksp. Teor. Fiz. **107**, 1464 (1995) [JETP **80**, 817 (1995)].
[5] N. N. Rozanov, Opt. i Spektrosk. **82**, 820 (1997) [Opt. Spectrosc. **82**, 761 (1997)].
[6] J. U. Kang, G. I. Stegeman, and J. S. Aitchison, Opt. Lett. **20**, 2069 (1995).
[7] W. E. Torruellas, Z. Wang, L. Torner, and G. I. Stegeman, Opt. Lett. **20**, 1949 (1995).
[8] L. Torner, W. E. Torruellas, G. I. Stegeman, and C. R. Menyuk, Opt. Lett. **20**, 1952 (1995).
[9] A. M. Goncharenko and V. G. Kukushkin, Vestsi AN BSSR, ser. Fiz.-mat., No. 6, 32 (1990).
[10] C. R. Menyuk, IEEE J. Quantum Electron. **QE-25**, 2674 (1989).
[11] Yu. S. Kivshar' and V. V. Konotop, Kvant. Élektron. **17**, 1599 (1990) [Sov. J. Quantum Electron. **20**, 1495 (1990)].
[12] Yu. S. Kivshar', Kvant. Élektron. **17**, 1603 (1990) [Sov. J. Quantum Electron. **20**, 1498 (1990)].
[13] Y. S. Kivshar and B. A. Malomed, Rev. Mod. Phys. **61**, 763 (1989).
[14] A. I. Maimistov and S. O. Elyutin, J. Mod. Opt. **39**, 2193 (1992).
[15] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *The Theory of Waves* [in Russian], Nauka, Moscow (1990).
[16] S. D. Kunitsyn, A. P. Sukhorukov, and V. A. Trofimov, Izv. Ross. Akad. Nauk, Ser. Fiz. **57**, 172 (1993).
[17] S. D. Kunitsyn, A. P. Sukhorukov, and V. A. Trofimov, Pis'ma Zh. Tekh. Fiz. **19**(3), 39 (1993) [Tech. Phys. Lett. **19**, 144 (1993)].
[18] B. Crosignani, M. Segev, and D. Engin, J. Opt. Soc. Am. B **10**, 446 (1993).
[19] M. Segev, B. Crosignani, and A. Yariv, Phys. Rev. Lett. **68**, 923 (1992).
[20] D. N. Christodoulides and M. I. Carvalho, J. Opt. Soc. Am. B **12**, 1628 (1995).
[21] M. Morin, G. Duree, G. Salamo, and M. Segev, Opt. Lett. **20**, 2066 (1995).
[22] M. Segev, G. C. Valley, S. R. Singh, M. I. Carvalho, and D. N. Christodoulidis, Opt. Lett. **20**, 1764 (1995).
[23] D. N. Christodoulides, S. R. Singh, M. I. Carvalho, and M. Segev, Appl. Phys. Lett. **68**, 1763 (1996).
[24] Z. Chen, M. Segev, T. H. Coskun, and D. N. Christodoulidis, Opt. Lett. **21**, 1436 (1996).
[25] D. Anderson and M. Lisak, Phys. Rev. A **27**, 1393 (1983).
[26] A. Anderson, Phys. Rev. A **27**, 3135 (1983).
[27] D. Anderson, M. Karlsson, M. Lisak, and A. Sergeev, Phys. Rev. E **47**, 3617 (1993).
[28] V. I. Kruglov, R. A. Vlasov, and V. M. Volkov, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 1182 (1989).
[29] J. W. Grantham, H. M. Gibbs, G. Khitrova, J. F. Valley, and Xu Jiajin, Phys. Rev. Lett. **66**, 1422 (1991).
[30] L. M. Pismen, Phys. Rev. Lett. **75**, 228 (1995).
[31] S. A. Akhmanov (ed)., *New Physical Principles of Optical Information Processing* [in Russian], Nauka, Moscow (1990).
[32] H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic, New York (1985).
[33] N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems* [in Russian], Nauka, Moscow (1997).

Translated by D. H. McNeill

Influence of the nonlocal nature of fluctuations on coherent effects in multiple scattering

V. L. Kuz'min

St. Petersburg Institute of Trade and Economics, 194021 St. Petersburg, Russia

V. P. Romanov*)

St. Petersburg State University, 198904 St. Petersburg, Russia

(Submitted 11 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2022–2033 (June 1998)

We solve the Bethe–Salpeter equation that describes radiation transfer in highly inhomogeneous media with an anisotropic scattering with allowance for the contributions of Legendre polynomials of the zeroth, first, and second degrees. An analytical expression for the radiation-transfer propagator is derived. We show that as the average value of the second-degree Lagrange polynomial increases, the region where the diffusion approximation is valid shifts toward large distances. Within this approach we calculate the coherent backscattering intensity and study the effect of higher-order moments on the angular dependence of this intensity. Finally, we show that it is possible to experimentally detect the coherent backscattering peak in the critical region. © *1998 American Institute of Physics.* [S1063-7761(98)00806-3]

1. INTRODUCTION

During recent years extensive study, both theoretical and experimental, has been in progress in the field of interference effects in multiple scattering of light (see the review articles in Refs. 1–5). A consistent theory of these effects has been developed for systems of point scatterers or for an inhomogeneous medium in which the spatial dispersion of the structure factor is small compared to the wavelength of the incident radiation. However, most experiments have been conducted with systems in which the spatial dimensions of the inhomogeneities are comparable to the wavelength of the radiation or are larger.

Allowing for the finite size of the scatterers usually amounts to replacing the photon mean free path l by the transport length[6,7] $l^*=l/(1-\overline{\cos\theta})$, where $\overline{\cos\theta}$ is the average cosine of the single-scattering angle. It is assumed that the results obtained in the theory of point scatterers remain valid but that the parameter l^*, which can be much larger than l, becomes the natural scale. Such an approach is justified by the diffusion approximation in radiation-transfer theory.[8,9] It actually amounts to allowing for the anisotropy of the scattering indicatrix via an expansion in Legendre polynomials[10,11] and keeping only first- and second-order terms.

However, in comparing theory and experiment, such an approach is also used for large scatterers, when the anisotropy of the indicatrix is not small. In this case it is natural to allow higher-order expansion terms in the indicatrix. In the research devoted to multiple scattering of light in liquid crystals within the scope of the Bethe–Salpeter equation,[12–15] the problem was solved numerically by using expansions in spherical functions for indicatrices of the Ornstein–Zernike type. There the asymptotic part of the coherence function, or the radiation-transfer propagator, was calculated. The numerical results were found to support the validity of the diffusion approximation at large distances. Gorodnichev and Rogozkin[16] examined highly anisotropic indicatrices of the form $(\cos\theta)^{-\alpha}$, where θ is the scattering angle, for layers whose thickness L was much smaller than the transport length, $L \ll l^*$. This corresponds to the case where the diffusion regime is not realized.

In the present paper we solve the Bethe–Salpeter equation with allowance for Legendre polynomials of the zeroth, first, and second degrees. Allowing for the second and higher-order terms means going outside the scope of the diffusion approximation. We show that here the main diffusion term in the radiation-transfer propagator of the form $(1-\overline{\cos\theta})r^{-1}$ does not change, and allowing for the second-order term has an effect on the form of the terms that rapidly decrease with increasing distance r, terms that retain the information about the anisotropy of the differential single-scattering cross section. These rapidly decreasing terms are nevertheless important when the scattered radiation is formed in a layer of thickness of the order of the extinction length, say, in backscattering.

The results are used to calculate the angular dependence of the coherent backscattering intensity. We show that allowing for second-order terms leads to a sizable contribution to the intensity. We also analyze the possibility of experimentally detecting the backscattering peak near second-order phase transition points. The coherent backscattering intensity is calculated for the Ornstein–Zernike indicatrix in the critical region. The calculated value of the effect is large enough to be detectible by modern experimental techniques both in temperature and scattering angles.

2. RADIATION TRANSFER IN AN INHOMOGENEOUS MEDIUM

We examine the propagation of light in an inhomogeneous medium whose stochastic properties are described by

fluctuations of the dielectric constant, $\Delta\varepsilon=\varepsilon(\mathbf{r})-\varepsilon$, where $\varepsilon=\overline{\varepsilon(\mathbf{r})}$. In such a medium the wave equation is

$$(\text{curl curl}-k^2)\mathbf{E}(\mathbf{r})=k_0^2\Delta\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}), \tag{2.1}$$

where $\mathbf{E}(\mathbf{r})$ is the electric field in the random medium, $k=k_0\sqrt{\varepsilon}$, and $k_0=2\pi/\lambda$ is the wave number, with λ the wavelength.

If we ignore polarization effects, we can replace the wave equation (2.1) by a scalar one, which in integral form is

$$E(\mathbf{r})=\langle E(\mathbf{r})\rangle+\frac{1}{4\pi}\int d\mathbf{r}_1 T(\mathbf{r}-\mathbf{r}_1)\Delta\varepsilon(\mathbf{r}_1)E(\mathbf{r}_1), \tag{2.2}$$

where $T(r)=k_0^2\exp(ikr)/r$ is the Green's function of the scalar wave equation.

Let us define the scattering intensity at large distances r_0 from the medium as

$$I(\mathbf{r})=\overline{\delta E(\mathbf{r})\,\delta E^*(\mathbf{r})}\equiv r_0^{-2}I(\mathbf{k}_s|\mathbf{k}_i),$$

where $\mathbf{k}_i$ and $\mathbf{k}_s$ are the wave vectors of the incident and scattered waves, and $\delta E(\mathbf{r})=E(\mathbf{r})-\overline{E(\mathbf{r})}$ is the scattered field. The function $I(\mathbf{k}_s|\mathbf{k}_i)$ can be written as[1–5]

$$I(\mathbf{k}_s|\mathbf{k}_i)=I_1(\mathbf{k}_s|\mathbf{k}_i)+I_M(\mathbf{k}_s|\mathbf{k}_i), \tag{2.3}$$

where I_1 is the single-scattering intensity, which is not discussed below, and I_M is the total intensity of all multiplicities of scattering beginning with the second.

In the weak-scattering approximation, $\lambda/l\ll 1$, only ladder diagrams contribute to the principal order in the parameter λ/l in the scattering-multiplicity series, i.e., $I_M(\mathbf{k}_s|\mathbf{k}_i)=I_L(\mathbf{k}_s|\mathbf{k}_i)$, where

$$I_L(\mathbf{k}_s|\mathbf{k}_i)=\int d\mathbf{R}_1 d\mathbf{R}_2\,\exp(2\,\text{Im}\,\mathbf{k}_s\mathbf{R}_2-2\,\text{Im}\,\mathbf{k}_i\mathbf{R}_1)\times\Gamma(\mathbf{R}_2,\mathbf{R}_1|\mathbf{k}_s,\mathbf{k}_i)|E|^2, \tag{2.4}$$

$\text{Im}\,\mathbf{k}$ is the imaginary part of the wave vector, $2\,\text{Im}\,k=l^{-1}$, and the function $\Gamma(\mathbf{R}_2,\mathbf{R}_1|\mathbf{k}_s,\mathbf{k}_i)$ the radiation-transfer propagator. Both terms in the exponential in (2.4) describe the damping of plane waves in the medium. For instance, for a medium occupying the half-space $z\geqslant 0$ the exponential factor is

$$\exp\left[\left(-\frac{z_2}{\cos\theta_s}-\frac{z_1}{\cos\theta_i}\right)\frac{1}{l}\right],$$

where θ_i is the angle of incidence, and θ_s is the scattering angle measured from the backscattering direction; for such a geometry we have $\theta_i,\theta_s<\pi/2$. Here we consider nonabsorbing media, so that damping is due solely to scattering.

The scattering-multiplicity series can be summed, and the result is a Bethe–Salpeter equation for the radiation-transfer propagator:

$$\Gamma(\mathbf{R}_2,\mathbf{R}_1|\mathbf{k}_s,\mathbf{k}_i)=k_0^8\tilde{G}(-\mathbf{k}_s+\mathbf{k}_{21})\Lambda(\mathbf{R}_{21})\tilde{G}(\mathbf{k}_i+\mathbf{k}_{12})+k_0^4\int d\mathbf{R}_3\tilde{G}(\mathbf{k}_{23}-\mathbf{k}_s)\Lambda(\mathbf{R}_{23})\times\Gamma(\mathbf{R}_3,\mathbf{R}_1|\mathbf{k}_{23},\mathbf{k}_i). \tag{2.5}$$

Here $\tilde{G}(\mathbf{q})=\int d\mathbf{r}G(\mathbf{r})\exp(-i\mathbf{q}\mathbf{r})$ is the Fourier transform of the correlation function of dielectric-constant fluctuations,

$$G(\mathbf{r})=(4\pi)^{-2}\overline{\Delta\varepsilon(0)\Delta\varepsilon(\mathbf{r})}, \tag{2.6}$$

$\Lambda(\mathbf{r})=r^{-2}\exp(-r/l)$, $\mathbf{k}_{ij}=k\mathbf{R}_{ij}/R_{ij}$ is the wave vector of the scattered wave propagating between the points $\mathbf{R}_i$ and $\mathbf{R}_j$, and $\mathbf{R}_{ij}=\mathbf{R}_i\mathbf{R}_j$. Note that the Bethe–Salpeter equation in the form (2.5) is similar to the mixed Wigner representation.[17]

If only ladder diagrams are considered, interference effects are not present in the picture discussed in Ref. 3. These effects are taken into account in Sec. 4.

3. ALLOWING FOR THE NONLOCAL NATURE OF FLUCTUATIONS

The correlation function $G(\mathbf{r})$ is characterized by a correlation radius r_c. Generally, for an arbitrary value of kr_c, it is impossible to solve Eq. (2.5) even for a homogeneous medium. However, when the correlation length is small, so that $G(\mathbf{r})$ can be replaced by a delta function, which in the wave-vector space corresponds to an isotropic single-scattering indicatrix, $\tilde{G}(\mathbf{q})=\text{const}$ and the Bethe–Salpeter equation can easily be solved by going over to the Fourier spectrum in the variable $\mathbf{R}_2-\mathbf{R}_1$. Ordinarily this case is used as a starting approximation in analyzing coherent and correlation effects in multiple scattering.

To allow for the finiteness of the correlation length, i.e., for the dependence of $\Gamma(\mathbf{R}_2,\mathbf{R}_1|\mathbf{k}_s,\mathbf{k}_i)$ on the orientations of $\mathbf{k}_s$ and $\mathbf{k}_i$, the radiation-transfer propagator can be expanded in a series in Legendre polynomials with only the zeroth- and first-order terms retained. However, for indicatrices with a strong anisotropy, higher-order moments $\overline{\cos^n\theta}$ may play an important role. The approach based on the expansion in spherical functions has been widely used in connection with the equation of radiation transfer.[11,18] In particular, Apresyan and Kravtsov[11] analyzed this equation in detail with first- and second-order moments taken into account.

According to (2.5), in a homogeneous medium the function $\Gamma(\mathbf{R}_2,\mathbf{R}_1|\mathbf{k}_s,\mathbf{k}_i)$ depends on the relative vector $\mathbf{R}_{21}=\mathbf{R}_2-\mathbf{R}_1$ and the mutual orientation of the vectors $\mathbf{k}_i$, $\mathbf{k}_s$, and $\mathbf{R}_{21}$. Generally, this dependence can be expressed in the form of an expansion in spherical functions:

$$\Gamma(\mathbf{R}|\mathbf{k}_s,\mathbf{k}_i)=\frac{1}{4\pi}\sum_{k,l,|m|\leqslant\min\{k,l\}}\gamma_{klm}(R)\times P_k^m(\cos\theta_s)P_l^m(\cos\theta_i)\exp[im(\phi_s-\phi_i)], \tag{3.1}$$

where $P_k^m(\cos\theta)$ is the associated Legendre polynomial,

$$\cos\theta_s=\frac{\mathbf{k}_s\mathbf{R}}{kR},\quad \cos\theta_i=\frac{\mathbf{k}_i\mathbf{R}}{kR},$$

and $\phi_s-\phi_i$ is the angle between the projections of the vectors $\mathbf{k}_s$ and $\mathbf{k}_i$ on a plane orthogonal to $\mathbf{R}$.

To simplify the analysis, we integrate Eq. (2.5) over the orientations of $\mathbf{k}_i$. The equation remains closed, since the

unknown function in its left- and right-hand sides depends on the same incident wave vector $\mathbf{k}_i$. Introducing a new function $\Gamma_0(\mathbf{R}_{21}|\mathbf{k}_s)=\int d\mathbf{\Omega}_i\Gamma(\mathbf{R}_{21}|\mathbf{k}_s,\mathbf{k}_i)$, we get

$$\Gamma_0(\mathbf{R}_{21}|\mathbf{k}_s)=k_0^4 l^{-1}\tilde{G}(-\mathbf{k}_s+\mathbf{k}_{21})\Lambda(\mathbf{R}_{21}) + k_0^4\int d\mathbf{R}_3\tilde{G}(\mathbf{k}_{23}-\mathbf{k}_s)\Lambda(\mathbf{R}_{23})\Gamma_0(\mathbf{R}_{31}|\mathbf{k}_{23}). \tag{3.2}$$

In deriving this Eq. (3.2) we used the optical theorem, $l^{-1}=k_0^4\int d\mathbf{\Omega}_s\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)$.

Although in going over to the equation for Γ_0 we lost the contribution of some terms, the equation makes it possible to calculate exactly the isotropic part of the radiation-transfer propagator outside the scope of the diffusion approximation. We allow for the orientational dependence of both vectors later, when we calculate coherent backscattering.

Expanding the function $\Gamma_0(\mathbf{R}_{21}|\mathbf{k}_s)$ in a series in Legendre polynomials,

$$\Gamma_0(\mathbf{R}|\mathbf{k})=\sum_{n=0}^{\infty}\gamma_n(R)P_n\left(\frac{\mathbf{Rk}}{Rk}\right), \tag{3.3}$$

where $\gamma_n(R)=\gamma_{n,0,0}(R)=\gamma_{0,n,0}(R)$, and using the orthogonality of the polynomials $P_n(\cos\theta)$, we reduce Eq. (3.2) to a system of equations for the expansion coefficients in (3.3):

$$\frac{4\pi}{(2n+1)}\gamma_n(R_{21})=\overline{P_n}\left[l^{-2}\Lambda(R_{21})+l^{-1}\times\int d\mathbf{R}_3P_n\left(\frac{\mathbf{R}_{23}\mathbf{R}_{21}}{R_{23}R_{21}}\right)\Lambda(R_{23})\times\sum_{m=0}^{\infty}\gamma_m(R_{31})P_m\left(\frac{\mathbf{R}_{31}\mathbf{k}_{23}}{R_{31}k}\right)\right]. \tag{3.4}$$

Here the parameters $\overline{P_n}$ are defined as averages over the single-scattering indicatrix:

$$\overline{P_n}=\frac{\int d\mathbf{\Omega}_s\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)P_n\left(\frac{\mathbf{k}_s\mathbf{k}_i}{k^2}\right)}{\int d\mathbf{\Omega}_s\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)}. \tag{3.5}$$

Next we solve the system of equations (3.4), keeping only the first three terms in the expansion (3.3). Instead of the functions $\gamma_1(R_{21})$ and $\gamma_2(R_{21})$ we introduce the vector function $\mathbf{g}_1(\mathbf{R}_{21})=\gamma_1(R_{21})\mathbf{R}_{21}/R_{21}$ and the tensor function

$$\hat{\mathbf{g}}_2(\mathbf{R}_{21})=\gamma_2(R_{21})\left(\frac{3\mathbf{R}_{21}\mathbf{R}_{21}}{R_{21}^2}-\hat{I}\right).$$

Applying the Fourier transformation in the spatial variables $\mathbf{R}_{ij}$, we can write the system of equations (3.4) as follows:

$$4\pi\tilde{\gamma}_0(q)=l^{-2}\tilde{\Lambda}_0(q)+l^{-1}\left\{\tilde{\Lambda}_0(q)\tilde{\gamma}_0(q)+\tilde{\mathbf{\Lambda}}_1(\mathbf{q})\tilde{\mathbf{g}}_1(\mathbf{q}) + \frac{1}{6}[(3\hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q})-\hat{\mathbf{I}}\tilde{\Lambda}_0(q))\hat{\tilde{\mathbf{g}}}_2(q)]\right\},$$

$$\frac{4\pi}{3\overline{P_1}}\tilde{\mathbf{g}}_1(\mathbf{q})=l^{-2}\tilde{\mathbf{\Lambda}}_1(\mathbf{q})+l^{-1}\left\{\tilde{\mathbf{\Lambda}}_1(\mathbf{q})\tilde{\gamma}_0(q) + \hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q})\tilde{\mathbf{g}}_1(\mathbf{q})+\frac{1}{2}\hat{\tilde{\mathbf{\Lambda}}}_3(\mathbf{q})\hat{\tilde{\mathbf{g}}}_2(\mathbf{q})\right\}, \tag{3.6}$$

$$\frac{4\pi}{5\overline{P_2}}\hat{\tilde{\mathbf{g}}}_2(\mathbf{q})=l^{-2}[3\hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q})-\hat{\mathbf{I}}\tilde{\Lambda}_0(q)]+l^{-1}\left\{[3\hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q}) - \hat{\mathbf{I}}\tilde{\Lambda}_0(q)]\tilde{\gamma}_0(q)+3\hat{\tilde{\mathbf{\Lambda}}}_3(\mathbf{q})\tilde{\mathbf{g}}_1(\mathbf{q})-\hat{\mathbf{I}}\tilde{\mathbf{\Lambda}}_1(\mathbf{q})\tilde{\mathbf{g}}_1(\mathbf{q}) + \frac{3}{2}\hat{\tilde{\mathbf{\Lambda}}}_4(\mathbf{q})\hat{\tilde{\mathbf{g}}}_2(\mathbf{q})-\frac{1}{2}\hat{\mathbf{I}}\ \mathrm{Tr}(\hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q})\hat{\tilde{\mathbf{g}}}_2(\mathbf{q}))\right\},$$

where the tilde indicates the Fourier transform of the corresponding function. The function

$$\hat{\tilde{\mathbf{\Lambda}}}_n(\mathbf{q})=\int d\mathbf{r}\ \exp(-i\mathbf{qr})\frac{\mathbf{rr}\ldots\mathbf{r}}{r^n}\Lambda(\mathbf{r}) \tag{3.7}$$

is a tensor of rank n. Integrals of type (3.7) can easily be calculated. In particular, we have

$$\tilde{\Lambda}_0(q)=4\pi lp_0,$$
$$\tilde{\mathbf{\Lambda}}_1(\mathbf{q})=-4\pi i\mathbf{q}l^2p_1, \tag{3.8}$$
$$\hat{\tilde{\mathbf{\Lambda}}}_2(\mathbf{q})=4\pi l\left[\frac{\mathbf{qq}}{q^2}p_1+\frac{p_0-p_1}{2}\left(\hat{I}-\frac{\mathbf{qq}}{q^2}\right)\right],$$

where we have introduced the auxiliary functions

$$p_0=\frac{\arctan w}{w},\quad p_1=\frac{1-p_0}{w^2},\quad p_2=\frac{1}{w^2}\left(\frac{1}{3}-p_1\right),$$

and $w=ql$ is the dimensionless argument.

As a result the solution of the system (3.6) can be written as

$$\tilde{\gamma}_0(q)=\frac{p_0-3p_1\,\overline{\cos\theta}}{lw^2p_1}+\frac{3p_2-p_1}{2Alp_1},$$
$$\tilde{\mathbf{g}}_1(\mathbf{q})=\frac{\mathbf{q}}{q}\tilde{\gamma}_1(q),\quad \tilde{\gamma}_1(q)=-\frac{3i\,\overline{\cos\theta}}{lw}, \tag{3.9}$$
$$\hat{\tilde{\mathbf{g}}}_2(\mathbf{q})=\left(\frac{3\mathbf{qq}}{q^2}-\hat{I}\right)\tilde{\gamma}_2(q),\quad \tilde{\gamma}_2(q)=-\frac{1}{lA},$$

where

$$A=\frac{p_1}{2(3p_2-p_1)}\left[\frac{4}{5\overline{P_2}}-9p_2+6p_1-p_0-\frac{w^2(3p_2-p_1)^2}{p_1}\right]. \tag{3.10}$$

Formula (3.9) determines the propagator $\Gamma_0(\mathbf{R}|\mathbf{k})$ when the scattering indicatrix is described by three Legendre polynomials.

The propagator of radiation transfer over large distances, i.e., the behavior of its Fourier transforms for $ql\leqslant 1$, plays an important role in coherent effects of multiple scattering. In this region the formula for $\tilde{\gamma}_0(q)$ in (3.9) yields

$$\tilde{\gamma}_0(q)\approx\frac{3}{l}\left[\frac{1-\overline{\cos\theta}}{w^2}+\frac{28}{27}\frac{1}{(35/9)(1/\overline{P_2}-1)+w^2}\right]. \tag{3.11}$$

The first term in the square brackets corresponds to the well-known expression in the diffusion approximation $\tilde{\gamma}_0(q)\sim 3/l^*w^2$. Within the equation of radiation transfer the form of this term in the absence of absorption does not change even if one allows for $\overline{P_2}$ (Ref. 11). The second term in (3.11) reflects a more detailed description of radiation transfer by the Bethe–Salpeter equation. This term vanishes as $\overline{P_2}\to 0$. As $\overline{P_1}=\overline{\cos\theta}$ increases, it is natural to expect that for real systems the average of the second Legendre polynomial over the scattering angle, $\overline{P_2}$, also increases, and the second term becomes much more important.

The region of applicability of the diffusion approximation is determined from the condition that the first term in the square brackets in (3.11) provides the main contribution, i.e.,

$$\frac{27}{28}(1-\overline{\cos\theta})>\frac{q^2l^2}{(35/9)(1/\overline{P_2}-1)+q^2l^2}. \tag{3.12}$$

We see from (3.12) that range of wave vectors $\mathbf{q}$ for which the diffusion approximation is valid narrows as the anisotropy of the scattering indicatrix increases, i.e., when $\overline{\cos\theta}$ and $\overline{P_2}$ grow.

Let us establish the limits of the approach based on allowing for a finite number of terms in the expansion in the Legendre polynomials in the scattering indicatrix. With allowance for (3.5) this expansion becomes

$$\frac{\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)}{\int\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)d\Omega}=\sum_0^\infty(2n+1)\overline{P_n}P_n(x), \tag{3.13}$$

where $x=(\mathbf{k}_i\mathbf{k}_s)/k^2$. Here $\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)$ describes the intensity of scattering in the direction of $\mathbf{k}_s$ and must be nonnegative, $\tilde{G}(\mathbf{k}_s-\mathbf{k}_i)\geqslant 0$. In particular, if we limit ourselves to second-degree polynomials, we have

$$1+3\overline{P_1}x+5\overline{P_2}(\tfrac{3}{2}x^2-\tfrac{1}{2})\geqslant 0 \tag{3.14}$$

for values of x in the interval $-1\leqslant x\leqslant 1$.

Equations (3.9)–(3.11) imply that the parameters $\overline{P_n}$ with $n\geqslant 2$ have no effect on the asymptotic behavior of the diffusion, so that for every value of $\overline{\cos\theta}$ there exists a range of distances within which the diffusion approximation is valid. However, in calculating the correction terms, which contain information about the higher-order moments of the indicatrix, the restriction (3.14) and other conditions that make it possible to describe the indicatrix by a finite number of moments are important.

4. COHERENT BACKSCATTERING

To illustrate the above results, we take the example of calculating the shape of the coherent backscattering peak. As is known, the peak appears because for scattering angles $\theta_s\ll\lambda/l$ measured from the backscattering direction, in addition to (2.3) and (2.4) there are cyclic diagrams that contribute to the scattering intensity, and the contribution is of the same order as that of ladder diagrams. Such interference corrections to the intensity of backscattering of light were first mentioned in Refs. 19–23 and discovered in experiments of Van Albada and Lagendijk[6] and Wolf and Maret.[7]

We consider the case of normal incidence, $\theta_i=0$. The contribution of cyclic diagrams to scattering is given by the following expression:[8,9]

$$I_C(\mathbf{k}_s,\mathbf{k}_i)\propto\int d\mathbf{R}_1d\mathbf{R}_2\,\exp[-(z_1+z_2)l^{-1}+ik_0\theta_s(x_1-x_2)]\times[\Gamma(|\mathbf{R}_2-\mathbf{R}_1||\mathbf{k}_s,\mathbf{k}_i)-\Gamma(|\mathbf{R}_2-\mathbf{R}_1^{(m)}||\mathbf{k}_s,\mathbf{k}_i)], \tag{4.1}$$

where $\mathbf{R}_1^{(m)}=(x_1,y_1,-z_1-2l_m)$ is the mirror image of point $\mathbf{R}_1$ with respect to the plane $z=-l_m$. The parameter l_m is usually chosen in the form $l_m=0.71l^*$ (Ref. 23). Equation (4.1) is written for the case of scattering from the half-space $z>0$ for a wave backscattered at an angle θ_s in the x,z plane.

We see that the scattering intensity depends on the complete function $\Gamma(\mathbf{R}|\mathbf{k}_s,\mathbf{k}_i)$. Hence we must refine the analysis of Sec. 3 by including the dependence on the orientation of both vectors, $\mathbf{k}_s$ and $\mathbf{k}_i$. Applying to Eq. (2.5) the operation of integration in the form

$$\int\int d\mathbf{\Omega}_sd\mathbf{\Omega}_iP_1(\cos\theta_s)P_1(\cos\theta_i)$$

and allowing for the expansion (3.1), for a homogeneous medium we have

$$\frac{4\pi}{9}\gamma_{110}(R_{21})=\frac{\overline{\cos\theta^2}\Lambda(R_{21})}{l^2}+\frac{\overline{\cos\theta}}{3l}\int d\mathbf{R}_3\Lambda(R_{23})\times\frac{(\mathbf{R}_{21}\mathbf{R}_{23})}{R_{21}R_{23}}\frac{(\mathbf{R}_{21}\mathbf{R}_{31})}{R_{21}R_{31}}\gamma_1(R_{31})+O(\overline{\cos\theta^3}) \tag{4.2}$$

Inserting the expansion (3.1) into (4.1) and doing a Fourier transformation in the variables $\mathbf{R}_2-\mathbf{R}_1$ and $\mathbf{R}_2-\mathbf{R}_1^{(m)}$, we obtain

$$I_C(\mathbf{k}_s,\mathbf{k}_i)\propto\int_{-\infty}^{\infty}\frac{dq_z}{2\pi}f(q_z)\times\left[\tilde{\gamma}_0(q_s)+2P_2\left(\frac{q_z}{q_s}\right)\tilde{\gamma}_2(q_s)+\frac{\mathbf{k}_i\mathbf{k}_s}{k^2}\hat{\tilde{g}}_{11}(\mathbf{q}_s)\right], \tag{4.3}$$

where $\mathbf{q}_s=(k_0\theta_s,0,q_z)$, $q_s=\sqrt{(k_0\theta_s)^2+q_z^2}$, and

$$\hat{g}_{11}(\mathbf{q}_s)=\int d\mathbf{R}\,\frac{\mathbf{RR}}{R^2}\gamma_{110}(R)\exp(-i\mathbf{q}_s\mathbf{R}). \tag{4.4}$$

The function $f(q)$ depends on the choice of boundary conditions. Taking $l_m=0.71l^*$, we get[8]

$$f(q)=\frac{1-w^2}{(1+w^2)^2}\left(1-\cos\frac{1.42wl^*}{l}\right)+\frac{2w}{(1+w^2)^2}\left(w+\sin\frac{1.42wl^*}{l}\right).$$

The function $\gamma_{110}(R)$ can be immediately calculated by (4.2) if we use the expression for the vector $(\mathbf{R}/R)\gamma_1(R) = \mathbf{g}_1(\mathbf{R})$ found in Sec. 3.

Parametrizing the tensor $\hat{\tilde{\mathbf{g}}}_{11}(\mathbf{q}_s)$ as

$$\hat{\tilde{\mathbf{g}}}_{11}(\mathbf{q}_s) = a_{11}(q_s)\hat{\mathbf{I}} + \frac{1}{2}b_{11}(q_s)\left(3\frac{\mathbf{q}_s\mathbf{q}_s}{q_s^2} - \hat{\mathbf{I}}\right), \tag{4.5}$$

from (4.4) we get

$$a_{11}(q_s) = \frac{1}{3}\int d\mathbf{R}\gamma_{110}(R)\exp(-i\mathbf{q}_s\mathbf{R}), \tag{4.6}$$

$$b_{11}(q_s) = \frac{2}{3}\int d\mathbf{R}P_2\left(\frac{\mathbf{q}_s\mathbf{R}}{q_sR}\right)\gamma_{110}(R)\exp(-i\mathbf{q}_s\mathbf{R}). \tag{4.7}$$

Substituting the right-hand side of Eq. (4.2) for $\gamma_{110}(R)$ in (4.6) and (4.7) and employing (3.8) and (3.9), we obtain

$$a_{11}(q_s) = \frac{3\overline{\cos\theta^2}}{l}\times\left[\frac{1}{w_s}\left(1+\frac{1}{w_s^2}\right)\arctan w_s - \frac{1}{w_s^2} + 2\int_{q_s}^{\infty}\frac{p_1}{q}dq\right], \tag{4.8}$$

$$b_{11}(q_s) = \frac{3\overline{\cos\theta^2}}{l}\times\left[\frac{1}{w_s^2} - \frac{1}{w_s}\left(1+\frac{1}{w_s^2}\right)\arctan w_s + 2q_s^{-3}\int_0^{q_s}q^2p_1dq\right], \tag{4.9}$$

where $w_s = lq_s$.

As a result the angle-dependent part of the coherent backscattering intensity can be represented (to within terms of order $\overline{\cos\theta^2}$) as

$$I_C(\mathbf{k}_s,\mathbf{k}_i) \propto \int_{-\infty}^{\infty}\frac{dq_z}{2\pi}f(q_z)\left[\tilde{\gamma}_0(q_s) + 2P_2\left(\frac{q_z}{q_s}\right)\times\tilde{\gamma}_2(q_s) - a_{11}(q_s) - b_{11}(q_s)P_2\left(\frac{q_z}{q_s}\right)\right]. \tag{4.10}$$

In the case of backscattering the contribution of γ_1 vanishes due to parity considerations.

The diffusion approximation amounts to allowing for the term $\tilde{\gamma}_0(q_s)$ in (4.10). Only this term yields a linear dependence on the scattering angle θ_s. The other contributions, responsible for anisotropy in the radiation-transfer propagator, yield a peak of Lorentzian form. Generally speaking, their presence leads to a deviation of the dependence from linear and at small angles manifests itself as a change in the initial slope. In the diffusion approximation, the shape of the peak at small angles is described by the formula $1 - \gamma kl^*\theta_s$, where the constant γ is the initial slope. Although the presence of correction terms complicates the description of the shape of the peak, these terms make it possible to determine not only l^* but also l, $\overline{P_1}$, and $\overline{P_2}$ separately, i.e., to extract detailed information about the scattering system, since Eq. (4.10) describes not only the initial slope but also the shape of the peak in a broad range of angles.

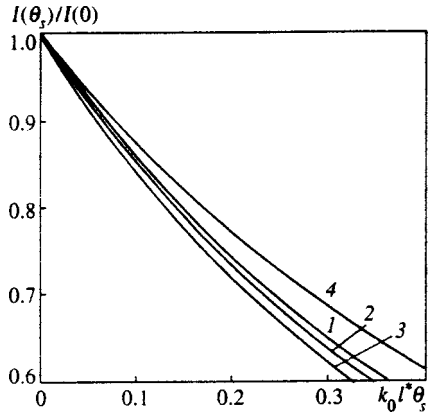


FIG. 1. Backscattering intensity normalized to the peak's height, $I(\theta_s)/I(0)$, as a function of the angular variable $kl^*\theta_s$ at $\overline{\cos\theta}=0.5$ for different values of $\overline{P_2}$: curve *1*, $\overline{P_2}=0.1$; curve *2*, $\overline{P_2}=0.2$; and curve *3*, $\overline{P_2}=0.3$. Curve *4* represents the function with only the first two expansion terms taken into account in (3.3).

Figure 1 shows how the shape of the backscattering peak varies with the parameter

$$\overline{P_2} = \overline{\tfrac{3}{2}\cos^2\theta - \tfrac{1}{2}}$$

for a fixed value $\overline{P_1} = \overline{\cos\theta} = 0.5$. We see that the slope increases with $\overline{P_2}$. For instance, at $\overline{P_2}=0.1$ the slope is $\gamma = 1.5$ and at $\overline{P_2}=0.3$ it is $\gamma = 1.7$. This means that the slope cannot be correctly described by only one parameter $\overline{\cos\theta}$. Curve *4* was obtained by allowing for γ_0 and γ_1 in (3.3) and yields a slope $\gamma = 1.35$. Note that the diffusion approximation also allows for only these two terms, but for $\tilde{\gamma}_0(q)$ it uses not an exact but an approximate expression of the form $(l^*w^2)^{-1}$, which leads to a slope $\gamma = 2.3$ (Ref. 24). Thus, curve *4* can be considered a refinement of the diffusion approximation. We used it as a reference curve to illustrate the role of higher-order terms in the expansion of the propagator.

Figure 2 depicts the backscattering intensity, calculated by (4.10), relative to the value of the intensity calculated

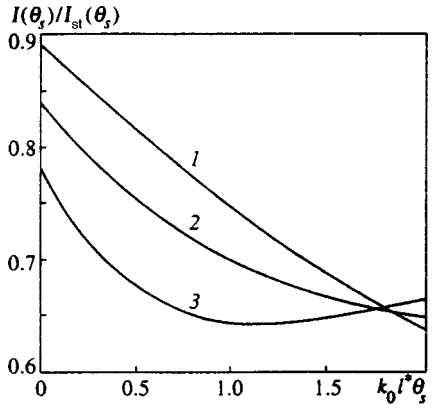


FIG. 2. Ratio of the backscattering intensity $I(\theta_s)$ at $\overline{P_1}=0.5$ to the intensity $I_{st}(\theta_s)$ calculated with $\gamma_n=0$ for $n\geqslant 2$, as a function of the scattering angle for $\overline{P_2}=0.1$ (curve *1*), $\overline{P_2}=0.2$ (curve *2*), and $\overline{P_2}=0.3$ (curve *3*).

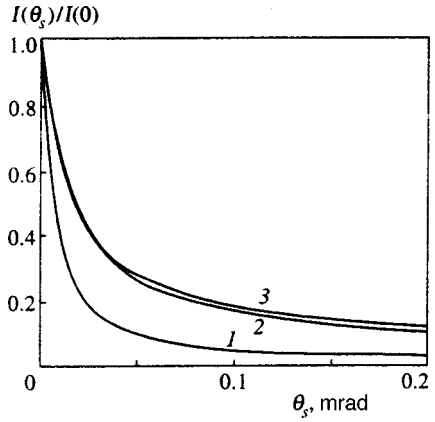


FIG. 3. Angular dependence of the backscattering intensity calculated for the Ornstein–Zernike indicatrix at different values of kr_c: curve *1*, $kr_c=0.775$, ($kl=4.9\times10^4$, $\overline{P_1}=0.2$, and $\overline{P_2}=0.047$); curve *2*, $kr_c=2.45$, ($kl=1.7\times10^4$, $\overline{P_1}=0.46$, and $\overline{P_2}=0.25$); and curve *3*, $kr_c=2.56$, ($kl=1.5\times10^4$, $\overline{P_1}=0.49$, and $\overline{P_2}=0.28$). The values of kl, $\overline{P_1}$, and $\overline{P_2}$ were calculated by formulas (4.11) and (4.12).

with only two terms in the expansion (3.3). We see that allowing for higher-order moments, first, reduces the height of the peak and, second, distorts the peak.

We used the results in an analysis of the possibility of detecting the coherent backscattering peak in the critical region. To simplify matters in describing the angular dependence of single scattering, we used the Ornstein–Zernike formula

$$\tilde{G}(q)\propto(r_c^{-2}+q^2)^{-1},$$

where r_c is the correlation radius. For such an indicatrix,

$$\overline{\cos\,\theta}=\frac{1+2x^2}{2x^2}-\frac{2}{\ln(1+4x^2)},$$

$$\overline{\cos^2\,\theta}=\frac{1+2x^2}{2x^2}\,\overline{\cos\,\theta}, \tag{4.11}$$

and the extinction length[25] is given by the following expression:

$$\frac{1}{l}=\frac{\pi}{2}B\left[\left(2+x^{-2}+\frac{1}{4}\,x^{-4}\right)\ln(1+4x^2)-2-x^{-2}\right], \tag{4.12}$$

where $x=kr_c$ and $B=R_{\rm sc}/x^2$, with $R_{\rm sc}$ the scattering constant.

We calculated the angular dependence of the backscattering peak for the range of kr_c values in which the first two moments of the scattering indicatrix satisfy the condition (3.4). We selected the following values characteristic of highly opalescent systems:[26,27] $B=1\ \text{cm}^{-1}$, $r_c=r_0\tau^{-\nu}$, $\nu=0.63$, $r_0=3.2$ Å, $\lambda=6328$ Å, $k=2n\pi/\lambda$, and $n=1.3$.

Figure 3 depicts the results of calculations for values $kr_c=0.775$, 2.45, and 2.56. The curves are plotted in terms of real angular units, so that it becomes possible to estimate the possibility of experimentally detecting the coherent backscattering peak in the critical region. We see that to an accuracy of 0.05 mrad, the backscattering peak can be detected in the critical region and its temperature dependence can be studied.

5. CONCLUSION

By examining radiation transfer in highly inhomogeneous media we have developed an approach that consistently takes into account the higher-order moments of the single-scattering indicatrix. We have obtained analytical expressions for the radiation-transfer propagator that take into account Legendre polynomials of the first and second degrees.

In solving the problem of radiation transfer, one can always indicate a range of distance $r\gg l^*$ in which the diffusion approximation is valid. There is no need to take into account higher-order terms in the expansion in spherical functions. However, for specific problems of multiple scattering, the boundary region of thickness of order the extinction length l, which is much smaller than l^* as $\overline{\cos\,\theta}\to1$, becomes important. In the intermediate region $l<r<l^*$, the contribution of higher-order spherical harmonics proves to be appreciable. In this connection the problem of explicitly taking into account the boundary conditions outside the scope of the mirror-image method becomes important.

As an example of an application of the new approach, we have calculated the coherent backscattering intensity and have studied the effect of higher-order moments on its angular dependence. We have demonstrated the possibility of experimentally detecting the coherent backscattering peak in the critical region.

This work was partially supported by the Russian Fund for Fundamental Research (Project 92-02-18201).

*)E-mail: vadim.romanov@pobox.spbu.ru

[1] P. J. Pine, D. A. Weitz, G. Maret, P.-E. Wolf, E. Herbolzheimer, and P. M. Chaikin, in *Scattering and Localization of Classical Waves in Random Media*, P. Sheng (ed.), World Scientific, Singapore (1989).
[2] P.-E. Wolf and G. Maret, in *Scattering in Volumes and Surfaces*, M. Nieto-Vesperanos and J. C. Dainty, Elsevier, Amsterdam (1990), p. 37.
[3] Yu. N. Barabanenkov, Yu. A. Kravtsov, V. D. Ozrin, and A. I. Saichev, in *Progress in Optics*, E. Wolf (ed.), Vol. 29, North-Holland, Amsterdam (1991), p. 65.
[4] A. Lagendijk and B. A. van Tiggelen, Phys. Rep. **270**, 145 (1996).
[5] V. L. Kuz'min and V. P. Romanov, Usp. Fiz. Nauk **166**, 247 (1996) [Phys. Usp. **39**, 231 (1996)].
[6] M. P. Van Albada and A. Lagendijk, Phys. Rev. Lett. **55**, 2692 (1985).
[7] P.-E. Wolf and G. Maret, Phys. Rev. Lett. **55**, 2696 (1985).
[8] E. Akkermans, P.-E. Wolf, R. Maynard, and G. J. Maret, J. Phys. (Paris) **49**, 77 (1988).
[9] Yu. N. Barabanenkov and V. D. Ozrin, Zh. Éksp. Teor. Fiz. **94**, No. 6, 56 (1988) [Sov. Phys. JETP **67**, 1117, 2175 (1988)].
[10] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Academic Press, New York (1978).
[11] L. A. Apresyan and Yu. A. Kravtsov, *Radiation Transfer Theory* [in Russian], Nauka, Moscow (1983).
[12] B. A. van Tiggelen, R. Maynard, and A. Heiderich, Phys. Rev. Lett. **77**, 639 (1996).
[13] H. Stark and T. C. Lubensky, Phys. Rev. Lett. **77**, 2229 (1996).
[14] B. A. van Tiggelen, R. Maynard, and T. M. Nieuwenhuizen, Phys. Rev. E **53**, 2881 (1996).
[15] H. Stark and T. C. Lubensky, Phys. Rev. E **55**, 514 (1997).
[16] E. E. Gorodnichev and D. B. Rogozkin, Zh. Éksp. Teor. Fiz. **107**, 209 (1995) [JETP **80**, 112 (1995)].

[17] V. I. Tatarskiĭ, Usp. Fiz. Nauk **139**, 587 (1983) [Sov. Phys. Usp. **26**, 311 (1983)].
[18] K. Case and P. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, Mass. (1967).
[19] K. M. Watson, J. Math. Phys. **10**, 688 (1969).
[20] D. A. de Wolf, IEEE Trans. Antennas Propag. **AP-19**, 2545 (1971).
[21] Yu. N. Barabanenkov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **16**, 88 (1973).
[22] A. G. Vinogradov, Yu. A. Kravtsov, and V. I. Tatarskiĭ, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **16**, 1055, 1064 (1973).
[23] A. A. Golubentsev, Zh. Éksp. Teor. Fiz. **86**, 47 (1984) [Sov. Phys. JETP **59**, 26 (1984)].
[24] F. C. MacKintosh and S. John, Phys. Rev. B **40**, 2383 (1989).
[25] V. G. Puglielli and N. C. Ford, Jr., Phys. Rev. Lett. **25**, 143 (1970).
[26] M. A. Anisimov, *Critical Phenomena in Liquids and Liquid Crystals*, Gordon & Breach, Philadelphia (1991); [Russ. orig., Nauka, Moscow (1987)].
[27] V. L. Kuzmin, V. P. Romanov, and L. A. Zubkov, Phys. Rep. **248**, 71 (1994).

Translated by Eugene Yankovsky

Generation of a continuum and stimulated Raman scattering harmonics during scattering of electromagnetic radiation of relativistic intensity

A. V. Borovskiĭ,*) V. V. Korobkin, O. B. Shiryaev

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

A. L. Galkin

Institute of Applied Mathematics, Russian Academy of Sciences, 125047 Moscow, Russia
(Submitted 14 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2034–2046 (June 1998)

A theory of the propagation instability of plane, monochromatic, circularly polarized electromagnetic waves of relativistic intensity in matter is developed for a spatially three-dimensional geometry including arbitrary polarization of the scattered radiation. Harmonic generation owing to striction and relativistic nonlinearity is examined, as well as scattering owing to electron recoil, the decay instability of the harmonics with formation of scattered electromagnetic waves (Stokes components of the stimulated Raman scattering and plasmons), the interaction of electromagnetic waves in the plasma (antistokes stimulated Raman scattering), and the generation of a radiative continuum. The transition of the three-dimensional theory to a one-dimensional problem in the nonrelativistic limit is discussed.
 [S1063-7761(98)00906-8]

1. INTRODUCTION

Experimental and theoretical studies of the scattering of ultrahigh-intensity laser light in matter have been of great interest in recent years.[1–10] By ultrahigh, we mean intensities $I \geqslant 10^{18}$ W/cm^2 at which relativistic effects show up in the electron motion. These intensities are attained at present in experiments with high-power ultrashort laser pulses.[11–13] When an ultrashort laser pulse is focussed onto a material, the bulk of the pulse interacts with the plasma formed at its leading edge. Polarization of the material in a high-power laser radiation field, which leads to scattering, can be caused by nonlinear free-electron currents,[14] deformation of the electronic shells of atoms and ions,[15] and molecular vibrations and rotations.[16] In experiments with light atomic gases, it is possible to isolate the first component of the polarization from this list on reaching full ionization of the material. We consider laser light scattering in plasmas for this specific reason in the following. The scattering of coherent radiation in plasmas at nonrelativistic intensities has been studied previously.[14,17–22] Of the papers which analyze scattering at relativistic intensities, we note above all Refs. 8–10 (see Refs. 1–7 as well).

The study of laser light scattering in matter can be arbitrarily divided into two essentially unrelated problems. The first involves determining the local characteristics of the medium: the temporal growth rates (spatial gain coefficients) of the scattered radiation in an elementary volume of plasma as functions of the components of the wave vector of the scattered wave and of the parameters of the reference wave. The second, in contrast, is an integro-differential transport problem and reduces to calculating the radiation field far from the scattering volume, including amplification and absorption along the propagation path. In this article we examine the first problem in detail; in particular, the temporal growth rates are calculated. The second problem is touched upon only at a qualitative level.

Recently, there has been a tendency to construct spatially three-dimensional models of the scattering of relativistically intense, plane, monochromatic, circularly polarized electromagnetic waves in plasmas.[9,10] The theory comprises a set of varied wave phenomena: harmonic generation, stimulated Raman scattering on electron plasma oscillations excited by the propagating laser pulse, the hydrodynamic analog of the Compton effect, etc., as well as limiting transitions to previously known cases—primarily to the nonrelativistic approximation.

At the same time, describing the scattering of relativistically intense electromagnetic waves in plasmas turns out to be so complicated that various authors have restricted themselves to examining a number of approximations. These include: 1) the one dimensional approximation[8]; 2) the assumption that a certain polarization of the scattered radiation, such as circular, is maintained (the approximation of a given polarization)[8,10]; 3) searching for the growth rates assuming that one of the transverse components of the wave vector equals zero[9]; 4) resonance approximations, which reduce to using exact conditions for phase synchrony for several wave processes.[20]

In this article, we propose a variant of the theory of the scattering of relativistically intense laser radiation in plasmas in which there is no need to use any of the four approximations indicated above. Numerical methods are used to describe harmonic generation, stimulated Raman scattering on plasmons, the hydrodynamic analog of Compton scattering, and continuum generation, as well as the mutual effects of these processes. This study is based on a rigorous analysis of

the linearized Maxwell equations and equations for the relativistic hydrodynamics of the electron component of a plasma in a strong electromagnetic field. A proper analysis of this type requires the availability of an exact reference solution of the initial nonlinear system of equations. In the present case, that solution is a plane, monochromatic, circularly polarized electromagnetic wave of arbitrary intensity.[23] We note that in the earlier literature, there are a number of studies of the propagation instability of linearly polarized, monochromatic plane waves which are not exact solutions of the initial relativistic equations, and therefore can only be used when it is assumed that their intensity is low.

The problem of studying the propagation instability of a circularly polarized, monochromatic, plane reference wave of arbitrary intensity reduces to solving a system of linear partial differential equations with oscillating coefficients. After introducing a running variable along the propagation axis and taking the Fourier transform with respect to the spatial coordinate, we arrive at a linear system consisting of an infinite number of coupled ordinary differential equations (owing to the necessity of accounting for the generation of harmonics and their interactions). As these calculations show, a proper approximate solution of this problem is obtained when more than a hundred equations are included in the analysis. The temporal growth rate of this instability is defined as the maximum eigenvalue of the matrix for the linear system being solved. This approach, in particular, makes it possible to avoid writing down and analyzing the cumbersome dispersion relations. Hydrodynamic studies of this type have been done with a numerical analysis of the linear stage of the development of turbulence.[24] We have tested this method in a linear analysis of the instability of a plane wave in the approximation of a specified polarization.[10]

In this paper, therefore, for the first time we present the results of a rigorous linear analysis of the system formed by Maxwell's equations and the equations of relativistic electron hydrodynamics.

2. INITIAL EQUATIONS

The propagation of laser light at relativistic intensities in a plasma is described by Maxwell's equations and the equations of relativistic electron hydrodynamics[23,25]:

$$\Box A=\nabla\phi_t+\gamma^{-1}n(\mathbf{A}+\nabla\psi), \tag{1}$$

$$\nabla^2\phi=n-1, \tag{2}$$

$$\nabla\cdot\mathbf{A}=0, \tag{3}$$

$$\psi_t=\phi-\gamma, \tag{4}$$

$$n_t+\nabla\cdot[\gamma^{-1}n(\mathbf{A}+\nabla\psi)]=0, \tag{5}$$

$$\gamma=(1+|\mathbf{A}+\nabla\psi|^2)^{1/2}. \tag{6}$$

Here $\mathbf{A}$ and ϕ are the vector and scalar potentials of the electromagnetic field, ψ is the potential for the generalized electron momentum, and n is the electron density. Equation (6) defines the relativistic mass factor γ. The subscript t denotes a partial derivative with respect to time.

The system of Eqs. (1)–(6) is normalized as follows: $\mathbf{A}$ and ϕ to mc^2/e, n to the unperturbed value n_0, the momentum of the electron fluid to mc, the time to ω_p^{-1}, where ω_p is the unperturbed value of the plasma frequency, and the spatial coordinate to c/ω_p.

The exact solution of the system of Eqs. (1)–(6) is a plane, monochromatic, circularly polarized wave of arbitrary intensity, propagating, for example, along the $\mathbf{e}_3$ axis:[23]

$$\mathbf{A}_0=(1/2)(\mathbf{e}_1+i\mathbf{e}_2)A_0\,\exp(ik\xi)+\text{c.c.} \tag{7}$$

Here $\xi=x_3-qt$ is the running variable and $q=\omega/k$ is the phase velocity of the wave. In addition, the conditions $\omega^2-k^2=\gamma_0^{-1}$, $\gamma_0=(1+A_0^2)^{1/2}$, $n=1$, $\phi=\gamma_0$, and $\psi=0$ are satisfied. In the following we use the notation $k^{-1}=\varepsilon$.

Let us consider the development of small perturbations, indicated by the symbol ''$\sim$,'' in the plasma, through which the wave (7) propagates:

$$\mathbf{A}=\mathbf{A}_0+\tilde{\mathbf{A}},\quad n=1+\tilde{n},\quad \phi=\gamma_0+\tilde{\phi},\quad \psi=\tilde{\psi}. \tag{8}$$

The linearized system of equations for the perturbations has the form

$$\Box\tilde{\mathbf{A}}=\nabla\tilde{\phi}_t+\gamma_0^{-1}(\tilde{\mathbf{A}}+\nabla\tilde{\psi})+\gamma_0^{-1}\tilde{n}\mathbf{A}_0 -\gamma_0^{-3}[\mathbf{A}_0\cdot(\tilde{\mathbf{A}}+\nabla\tilde{\psi})]\mathbf{A}_0, \tag{9}$$

$$\nabla^2\tilde{\phi}=\tilde{n}, \tag{10}$$

$$\nabla\cdot\tilde{\mathbf{A}}=0, \tag{11}$$

$$\tilde{\psi}_t-\tilde{\phi}=-\gamma_0^{-1}[\mathbf{A}_0\cdot(\tilde{\mathbf{A}}+\nabla\tilde{\psi})]. \tag{12}$$

The continuity equation is obtained by taking the divergence of Eq. (9),

$$\tilde{n}_t+\gamma_0^{-1}(\mathbf{A}_0\cdot\nabla\tilde{n})=-\gamma_0^{-1}\Delta\tilde{\psi}+\gamma_0^{-3} \times\{\mathbf{A}_0\cdot\nabla[\mathbf{A}_0\cdot(\tilde{\mathbf{A}}+\nabla\tilde{\psi})]\}. \tag{13}$$

The system of Eqs. (9)–(13) is the same as the equations obtained in Ref. 9 to within the normalization.

Our procedure in the following differs from that used in Ref. 9. In the system of Eqs. (9)–(13) we transform to the comoving variables $(\mathbf{x}_\perp,\xi,t)$, after which the differential operators take the following form:

$$\nabla\to(\nabla_\perp,\partial_\xi),\quad \nabla^2\to\nabla_\perp^2+\partial_\xi^2,\quad \partial_t\to D_t=\partial_t-q\partial_\xi,$$
$$\Box\to D_\Box=\Delta_\perp-\partial_t^2+2q\partial_{\xi t}^2-\varepsilon^2\gamma_0^{-1}\partial_\xi^2.$$

As a result, we obtain a system of linear differential equations with coefficients that depend periodically on ξ, which we shall not write out. Given that we are studying the propagation of laser light in an unbounded uniform plasma, we transform this system of equations into the wave vector space of the perturbations (into momentum $\mathbf{k}$-space) by taking the Fourier transform with respect to $\mathbf{x}_\perp,\xi$ ($k=|\mathbf{k}_\perp|$):

$$(\tilde{\mathbf{A}},\tilde{\phi},\tilde{n},\overline{\psi})^T=(2\pi)^{-3/2}\int(\tilde{\mathbf{A}},\tilde{\phi},\tilde{n},\tilde{\psi})^T_{\mathbf{k}_\perp,\xi} \times\exp[i((\mathbf{k}_\perp,\mathbf{x}_\perp)+\chi\xi)]d^2\mathbf{k}_\perp d\chi.$$

In the following the symbol ''$\sim$'' on the small perturbations will be omitted for the sake of brevity.

The results of taking the Fourier transforms ($g_1 = a_0^2/2\gamma_0^3$, $g_2 = a_0/2\gamma_0$) are

$$\hat{D}_{\square}A_1 + g_1 \frac{ik_1\chi^2}{\mathbf{k}^2+\chi^2}\psi - g_1 \frac{k_1k_2}{\mathbf{k}^2+\chi^2}A_2 + g_1 \frac{k_2^2+\chi^2}{\mathbf{k}^2+\chi^2}A_1 + g_2((F_{1,1}^-)_{\chi-k} + (F_{1,1}^+)_{\chi+k}) + \frac{g_1}{2}((F_{1,2}^-)_{\chi-2k} + (F_{1,2}^-)_{\chi+2k}) = 0, \quad (14)$$

$$\hat{D}_{\square}A_2 + g_1 \frac{ik_2\chi^2}{\mathbf{k}^2+\chi^2}\psi - g_1 \frac{k_1k_2}{\mathbf{k}^2+\chi^2}A_1 + g_1 \frac{k_1^2+\chi^2}{\mathbf{k}^2+\chi^2}A_2 + g_2((F_{2,1}^-)_{\chi-k} + (F_{2,1}^+)_{\chi+k}) + g_1((F_{2,2}^-)_{\chi-2k} + (F_{2,2}^+)_{\chi+2k}) = 0, \quad (15)$$

$$-(\mathbf{k}^2+\chi^2)\phi = n, \quad (16)$$

$$\hat{D}_t\psi - \phi + g_2((\Pi_1 + i\Pi_2)_{\chi-k} + (\Pi_1 - i\Pi_2)_{\chi+k}) = 0, \quad (17)$$

$$\hat{D}_t n = (\gamma_0^{-1}\chi^2 + (\gamma_0^{-1} - g_1)k^2)\psi + ig_1(k_1A_1 + k_2A_2) - g_2((ik_1 - k_2)n_{\chi-k} + (ik_1 + k_2)n_{\chi+k}) + (g_1/2)[((ik_1 - k_2)\Pi_1 - (ik_2 + k_1)\Pi_2)_{\chi-2k} + ((ik_1 + k_2)\Pi_1 - (ik_2 - k_1)\Pi_2)_{\chi+2k}]. \quad (18)$$

Here we have used the notation

$$\mathbf{A} = (A_1, A_2, A_3), \quad \Pi_{1,2} = A_{1,2} + ik_{1,2}\psi, \quad \hat{D}_t = \partial_t - iq\chi,$$

$$\hat{D}_{\square} = -\partial_t^2 + 2iq\chi\partial_t + (\chi^2/\gamma_0 k^2 - \mathbf{k}^2) - \gamma_0^{-1},$$

$$F_{1,1}^- = \left(k_1 \frac{k_1 + ik_2}{\mathbf{k}^2+\chi^2} - 1\right)n, \quad F_{1,1}^+ = \left(k_1 \frac{k_1 - ik_2}{\mathbf{k}^2+\chi^2} - 1\right)n,$$

$$F_{1,2}^- = -\left(k_1 \frac{k_1 + ik_2}{\mathbf{k}^2+\chi^2} - 1\right)(\Pi_1 + i\Pi_2),$$

$$F_{1,2}^+ = -\left(k_1 \frac{k_1 - ik_2}{\mathbf{k}^2+\chi^2} - 1\right)(\Pi_1 - i\Pi_2),$$

$$F_{2,1}^- = \left(k_2 \frac{k_1 + ik_2}{\mathbf{k}^2+\chi^2} - i\right)n, \quad F_{2,1}^+ = \left(k_2 \frac{k_1 - ik_2}{\mathbf{k}^2+\chi^2} + i\right)n,$$

$$F_{2,2}^- = \left(ik_2 \frac{k_1 + ik_2}{\mathbf{k}^2+\chi^2} + 1\right)(i\Pi_1 - \Pi_2),$$

$$F_{2,2}^+ = \left(ik_2 \frac{k_1 - ik_2}{\mathbf{k}^2+\chi^2} - 1\right)(i\Pi_1 + \Pi_2).$$

Upon shifting the argument χ in these equations by $\pm nk$, where n is an integer, we arrive at an infinite chain of coupled, linear ordinary differential equations with respect to time for the amplitudes of the harmonics over which the perturbations are expanded. This latter system of equations can be written in the form

$$\mathbf{Y}_t = \mathbf{BY}, \quad (19)$$

where $\mathbf{Y}$ is an infinite-dimensional column vector, and $\mathbf{B}$ is an infinite-dimensional 30-diagonal matrix.

3. ONE-DIMENSIONAL PROBLEM

In the one-dimensional approximation, the components k_1 and k_2 of the perturbation wave vector must be set equal to zero in Eqs. (14)–(18). The equations for the one dimensional approximation are

$$\hat{D}_{\square,1}A_1 + g_1A_1 - g_2(n_{\chi-k} + n_{\chi+k}) + (g_1/2)((A_1 + iA_2)_{\chi-2k} + (A_1 - iA_2)_{\chi+2k}) = 0, \quad (20)$$

$$\hat{D}_{\square,1}A_2 + g_1A_2 + ig_2(-n_{\chi-k} + n_{\chi+k}) + (g_1/2)((iA_1 - A_2)_{\chi-2k} - (iA_1 + A_2)_{\chi+2k}) = 0, \quad (21)$$

$$\hat{D}_t\psi + \chi^{-2}n + g_2((A_1 + iA_2)_{\chi-k} + (A_1 - iA_2)_{\chi+k}) = 0, \quad (22)$$

$$\hat{D}_t n - \gamma_0^{-1}\chi^2\psi = 0, \quad (23)$$

where

$$\hat{D}_{\square,1} = -\partial_t^2 + 2iq\chi\partial_t + \chi^2/\gamma_0 k^2 - \gamma_0^{-1}.$$

The form of these equations, as well as the structure of the matrix $\mathbf{B}$, are analogous to those for the spatially three-dimensional case.

4. APPROXIMATION WITH A GIVEN CIRCULAR POLARIZATION

Equations (14)–(18), as well as their one-dimensional counterparts Eqs. (20)–(23), correspond to an arbitrary polarization of the perturbations in the electromagnetic field in the plasma, and in this regard, they differ from the problems presented in Refs. 8 and 10. In the latter, the polarization of the perturbation in the electric field was assumed to be circular, which greatly limited the generality of the approach.

Let us consider the spatially one-dimensional case. For a perturbation with given circular polarization,

$$\mathbf{A} = (1/2)(\mathbf{e}_1 + i\mathbf{e}_2)a\ \exp[i(kx_3 - \omega t)] + \text{c.c.},$$

which in terms of Fourier transforms means that

$$A_1 = \frac{1}{2}(a_{\chi-k} + b_{\chi+k}), \quad A_2 = \frac{i}{2}(a_{\chi-k} - b_{\chi+k}), \quad (24)$$

where a_χ and b_χ are the Fourier transforms of a and a^*, respectively.

Substituting these equations into Eqs. (20) and (21), we obtain the following equation for a:

$$-a_{tt} + 2iq\chi a_t + (\chi^2/\gamma_0 k^2)a + 2i(\omega a_t - (i\chi/\gamma_0 k)a) = 2g_2n - g_1(a + b). \quad (25)$$

Differentiating Eq. (23) with respect to time and using Eq. (22), as well as solving Eqs. (24) for a_χ and b_χ, we obtain the following equation for the response of plasma electrons to the propagating laser radiation:

$$((\partial_t - iq\chi)^2 + \gamma_0^{-1})n = -(a_0/2\gamma_0^2)\chi^2(a+b). \quad (26)$$

It is easy to confirm that Eqs. (25) and (26) are the Fourier transforms of the equations used in Ref. 8 to describe the corresponding instability of laser radiation in a plasma, under the assumption that its circular polarization is preserved.

5. NUMERICAL STUDY OF THE MAXIMUM INSTABILITY GROWTH RATES FOR LASER RADIATION IN A PLASMA

In examining the linearized problem with a matrix **B** of finite size, it is natural to take the maximum of the real part in its eigenvalue spectrum as the temporal growth rate. The linear algebra problem of finding the eigenvalues of the matrix can be solved using the QR algorithm.[26] The results below emphasize the advantage of the approach used here compared to the traditional methods of finding the growth rates, which involve solving the discriminant (dispersion) equations.

5.1. One-dimensional problem

Let us consider the problem of finding the maximum growth rates for the system of Eqs. (20)–(23). The results of the corresponding calculations are shown in Figs. 1 and 2. The size of the matrix **B** increases as $B = m \times m$, where $m = 6 + 12j$, with $j = 0,1,2,\ldots,17$. As the number j increases, the number of harmonics taken into account becomes larger, and this is reflected in Fig. 1. $j = 1$ corresponds to the frequency of the laser radiation. The wave vector of the scattered radiation equals χ. If $\chi > 0$, then the scattered radiation propagates in the positive x_3 direction, while if $\chi < 0$, then it propagates in the negative direction.

Figures 1 and 2 show the growth rates of the perturbations as functions of χ. The scattered radiation is a set of harmonics, each of which is a doublet consisting of Stokes and antistokes stimulated Raman scattering components. Thus, all the peaks in Figs. 1 and 2 correspond to values $\chi = \pm jk \pm k_p$ of the wave vector. Weak peaks are observed near $\chi = \pm jk$, corresponding to the hydrodynamic analog of Compton scattering.

The calculations show that as the dimensionality m of the matrix **B** increases, the changes in the magnitudes of the growth rates of the bulk of the harmonics become small, except for a few at the edges (see Fig. 1). In other words, in calculations using a matrix **B** with a finite dimensionality, there is an edge effect encompassing the extreme harmonics. For example, if $m = 210$, then the harmonics $j = 0,1,2,\ldots,14$ are given with adequate accuracy, while the harmonics $j = 15$, 16, and 17 are ''smeared out'' by the edge effect.

For the parameters of the calculations shown in Fig. 2, compared to the parameters corresponding to Fig. 1, the electron density is lower by a factor of two; this led to a drop in the plasma frequency ω_p and in k_p by a factor of $\sqrt{2}$. The stimulated Raman scattering components of the harmonics in Fig. 2 were closer by a factor of $\sqrt{2}$.

Figure 2 shows that as the amplitude of the field of the reference wave increases, the stimulated Raman scattering components are broadened. At fields $A_0 \simeq 1$, the two stimulated Raman scattering components merge into one. Thus, in the relativistic region with $A_0 > 1$, the harmonics are resolvable, but the stimulated Raman scattering components are not.

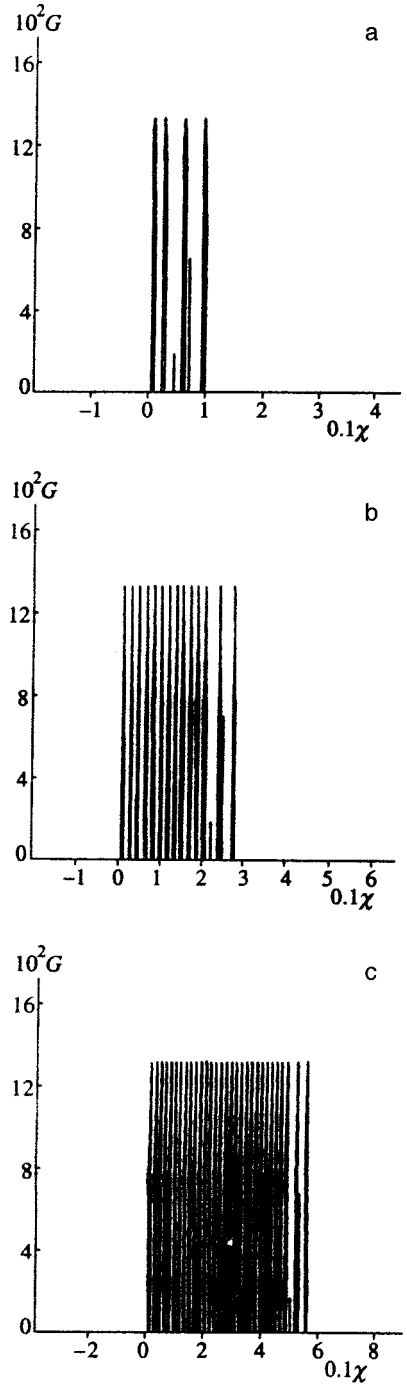


FIG. 1. The effect of the number of harmonics on the growth rate: $2j+1 = 5$ (a), 15 (b), 35 (c) harmonics; $a_0 = 0.1$, $\varepsilon^2 = 7.43 \cdot 10^{-2}$. The growth rate is shown here as a function of the longitudinal component of the perturbation wave vector for $\chi > 0$. The plot is symmetric in χ.

It is interesting to follow the correspondence between these calculations and those of Ref. 8, where it was assumed that in the framework of the one-dimensional problem the scattered radiation is circularly polarized. Our study shows that this leads to a significant error—an infinite sequence of harmonics is omitted. The problem reduces to studying a 6×6 matrix. As the computational results described above

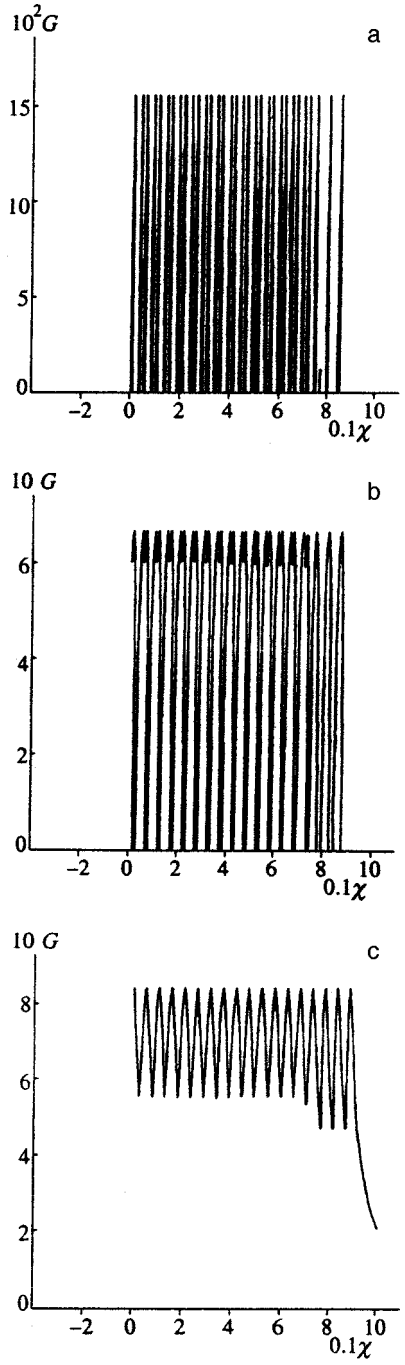


FIG. 2. The dependence of the growth rate on the intensity of the reference wave for $\chi>0$. The plot is symmetric in χ. $a_0=0.1$ (a), 0.5 (b), 5 (c). $\varepsilon^2=3.72\cdot10^{-2}$.

show, for arbitrarily polarized scattered light, in the one-dimensional problem all the harmonics of the radiation are generated and they all interact. In mathematical terms, here the problem reduces to the need to solve an infinite squence of coupled linear differential equations. Note that calculations using matrices with low dimensionality $m<20$ yield results close to those of Ref. 8.

5.2. Three-dimensional problem

We have studied the maximum growth rate for Eqs. (14)–(18) as a function of the three components of the perturbation wave vector **k**. Only a plot of the growth rate as a function of two variables is capable of providing a clear representation. For example, we considered distributions of the growth rate of the form $G(k_1,0,\chi)$, $G(0,k_2,\chi)$, and $G(k_1,k_2,0)$, as well as $G(k\cos\Phi,k\sin\Phi,\chi)$ for a fixed angle Φ and $k=(k_1^2+k_2^2)^{1/2}$. An investigation showed that the growth rate is a quasiperiodic function of χ (to within the accuracy of the numerical calculations) and is not axially symmetric with respect to the $\mathbf{e}_3$ axis, owing to the lack of such symmetry in the linearized equations over the period of the reference wave. The oscillatory coefficients that appear in the linearized equations depend on the choice of reference system. In our case, we calculated the growth rate under conditions such that at time $t=0$ the vector $\mathbf{A}_0$ is directed along $\mathbf{e}_1$ at the point $x_3=0$. However, the choice of initial time within a wave period is random. This means that the growth rate must be averaged over a wave period with respect to the initial time, which is equivalent in turn to averaging the growth rate over the azimuthal angle. Thus, the unaveraged computational results are intermediate and have the significance of ''unobservables,'' while the averaged results correspond to physically observable quantities.

Figure 3 shows the average growth rate as a function of k and χ. The distinct feature of this solution is that, against a continuum background, one can see 1) a system of interlocking rings, 2) repeating peaks located near the $\mathbf{e}_3$ axis, and 3) an increase in the growth rate as $k\to\infty$. Figure 3 illustrates the scattering of circularly polarized laser radiation in plasmas. Because of striction and the relativistic nonlinearity, harmonics $m\mathbf{k}_0-\delta\mathbf{k}$, with $m=0,\pm1,\pm2,\ldots$ (including a shift $\delta\mathbf{k}$ owing to electron recoil, with $|\delta\mathbf{k}|\ll|\mathbf{k}_0|$) are generated in a medium through which laser light with wave vector $\mathbf{k}_0$ propagates.

The initial equations obey the energy and momentum conservation laws.[25] Since their subsequent transformations are exact, the electron recoil effect must be included in this theory. Because of the decay instability,

$$m\mathbf{k}_0-\delta\mathbf{k}\to\mathbf{k}'_m+\mathbf{k}_e,$$

$$m\omega_0-\delta\omega=(m\omega_0-\delta\omega-\omega_p)+\omega_p,$$

each harmonic decays into an electromagnetic (Stokes stimulated Raman scattering) component and a plasma wave. Since the wave vector of plasma oscillations in a cold plasma can have any value, the vector $\mathbf{k}'_m$ is oriented arbitrarily in space. Thus, the growth rate for the Stokes stimulated Raman scattering component in **k**-space has a distribution close to a circle of radius $|\mathbf{k}'_m|$. Furthermore, wave interactions leading to the formation of scattered waves $\mathbf{k}''=n\mathbf{k}_0+\mathbf{k}'_m$ can take place, where n and m are arbitrary integers. In Fig. 3, one can see ring structures of type $\mathbf{k}''=n\mathbf{k}_0+\mathbf{k}'_1$. Ring structures with larger $m>1$ vanish during averaging, but they were seen in the intermediate (unaveraged) calculations.

In a medium without a reference plasma wave at frequency ω_p, the antistokes stimulated Raman scattering components do not appear, and this is shown by the calculations. Nevertheless, because of wave interactions between the harmonics $m\mathbf{k}_0$ and the backscattered Stokes stimulated Raman scattering components $\mathbf{k}'_n$, waves with wave vectors directed

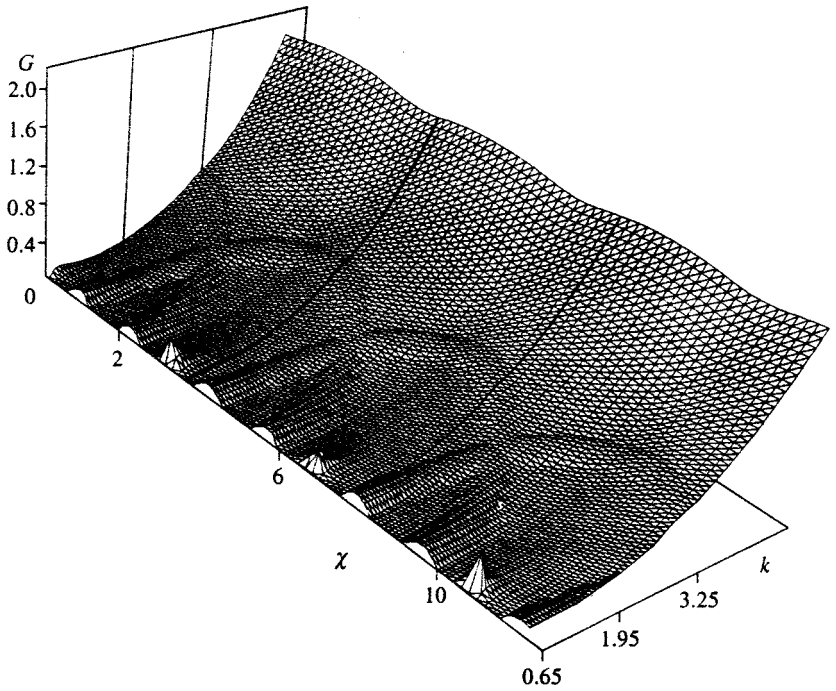


FIG. 3. Two-dimensional distribution of the growth rate as a function of k and χ averaged over azimuthal angle. $a_0=0.1$ and $\varepsilon^2=7.43\cdot 10^{-2}$.

along the propagation axis and absolute values $(m-n)k_0+k_p$ appear. This creates a pattern in which antistokes stimulated Raman scattering components appear in the one-dimensional problem. In the three-dimensional case, for this reason, scattering will be observed in the antistokes components only within narrow solid angles along the propagation axis of the reference wave.

The peaks along the $\mathbf{e}_3$ axis correspond to harmonics with wave vectors $\mathbf{k}''=m\mathbf{k}_0-\delta\mathbf{k}$, which have been shifted as a result of electron recoil.

The increase in growth rate as $k\to\infty$ corresponds to generation of a continuum, i.e., to scattered radiation in the continuum. When an electron revolves in a circle,[27] synchrotron radiation is produced at an infinite set of harmonics. When the circular trajectories of the electrons are distorted, the emission spectrum changes, leading to formation of a continuum. In experiments with laser plasmas, the continuum is produced by at least three mechanisms: 1) bremsstrahlung and partial photorecombination radiation from the plasma, 2) nonmonochromaticity of the laser radiation, which is especially important for ultrashort pulses, and 3) anharmonicity of the electron currents in the plasma. The latter factor is usually related to plasma turbulence and to an anomalous increase in radiation from the plasma.[20]

This theory leads to an interesting hypothesis about the polarization of the scattered light. Since a perturbation in the volume of the plasma at a given time consists of the sum of perturbations arising at different times within a wave period (the linear problem), while the asymptotic solutions for different initial times differ by a rotation of the **k**-space through an azimuthal angle φ (the phase angle) relative to the $\mathbf{e}_3$ axis, we should expect that the resulting vector potential would have components $\langle A_1\rangle=0$, $\langle A_2\rangle=0$, and $\langle A_3\rangle\neq 0$. The rms values of all the components are, of course, nonzero, i.e., the scattered radiation will be partially depolarized. A more detailed study of this question will require use of the methods of statistical mechanics,[28] which lies beyond the scope of this article.

The theory discussed above allows a limiting transition to the nonrelativistic approximation. In mathematical terms, it is necessary here to eliminate terms proportional to A_0^2 and set $\gamma_0=1$ in Eqs. (9)–(13). We then obtain an eigenvalue problem for a matrix **B** which is 15-diagonal, as opposed to the relativistic case. Calculations in the nonrelativistic approximation for $a_0=0.1$ agree with those shown in Figs. 1 and 2a to high accuracy.

The class of nonrelativistic problems was examined quite long ago in the literature and two different approaches were used: studying the instabilities in a framework of linearized equations based on a resonance approximation assuming that the phase synchronism conditions are satisfied,[20] and studying the dispersion relations without employing the phase synchronism conditions[14] (an initial or boundary value problem). As the present study has shown, the resonance approximation is not valid for relativistic problems because of the considerable broadening of the resonance structures.

There is some interest in comparing the results of this paper with Ref. 9, in which dispersion relations were obtained for the scattering of circularly polarized waves in a three-dimensional geometry. That study[9] was limited to the case $k_2=0$, assuming axial symmetry of the problem (the real situation is more complicated and requires averaging of the growth rate over the azimuthal angle in **k**-space). The series of Stokes stimulated Raman scattering harmonics[9] corresponds, in our case, to a system of rings for scattered waves with a wave vector $\mathbf{k}'_m$. The differences are the following: 1) in the unaveraged picture, as the polar angle $\theta\to 0$, π, the growth rate approaches the nonzero values obtained by solving the one-dimensional problem, as opposed to Ref.

9; 2) in the averaged picture, for scattering at large angles, one sees only the Stokes component at the fundamental frequency $\mathbf{k}_1'$ against a continuum background; and, 3) for scattering at small angles the theory yields Stokes and antistokes components over the entire sequence of harmonics.

6. CONCLUSION

A rigorous linear analysis of propagation instabilities of a plane, monochromatic, circularly polarized electromagnetic wave of arbitrary intensity in a plasma has been presented in this article in terms of a three-dimensional geometry. The theory describes the following wave processes: generation of harmonics of the propagating laser radiation in a nonlinear medium, scattering including the effect of electron recoil, decay instability of the harmonics with formation of scattered electromagnetic waves and plasmons, interactions of electromagnetic waves in the plasma, and generation of a continuum during scattering of the laser radiation. This set of wave processes has been studied in both the relativistic and nonrelativistic cases.

The calculations demonstrate the possibility of forward and backward scattering. Each of the radiation fluxes contains a set of harmonics. Scattering of the harmonics takes place into a set of overlapping spatial cones. A higher harmonic propagates into a narrower spatial cone. Outside these cones the theory predicts scattering with a continuous spectrum for the change in the wave vector. This effect is dominant.

As laser pulses propagate in a plasma, the lower-order harmonics are radiated into wider cones than the higher-order harmonics. Thus, the lower-order harmonics can leave the localization region of the pulse in the course of propagating. At the same time, the higher-order harmonics propagate along with the pulse. They can probably be detected by spectral analysis of the pulse in an experiment with a special geometry. The backscattering intensity is low because the interaction time for the radiation propagating in opposite directions is short.[10]

In this paper, we have examined for the first time a one-dimensional theory of scattering of laser radiation in plasmas without assuming that a particular type of polarization is preserved. It has been shown that in this case scattering leads to a sequence of harmonics of the laser light, each of which consists of a doublet made up of the Stokes and antistokes components. The mechanisms for generation of the harmonics of the scattered light are relativistic nonlinearity and electron striction.

In the framework of a multidimensional theory it has been established that during scattering of laser radiation at large angles, the most important effects that might be observed experimentally are the production of a Stokes stimulated Raman scattering component at the fundamental frequency and the generation of a scattered light continuum.

For scattering into small solid angles along the propagation axis of the reference wave, the harmonics contain both Stokes and antistokes stimulated Raman scattering components, with the latter resulting from a combination of higher-order harmonics with the Stokes stimulated Raman backscatter component. It has been shown that the harmonics have a doublet structure for nonrelativistic intensities, while for relativistic intensities the components of the harmonics are broadened and merge into a single line.

The results of the multidimensional theory for polar scattering angles of 0 and π transform to the corresponding results of the one-dimensional theory. (Unlike in Ref. 9, the growth rates for these scattering angles do not go to zero.)

In the numerical studies, we have, for the first time, calculated the maximum growth rates (real parts of the eigenvalues) for the matrix **B** of an infinite sequence of coupled ordinary differential equations as a technique for analyzing this class of problems. This approach is more effective than deriving and studying the extremely cumbersome dispersion relations.

In comparing theory with experiment, the following circumstances must be taken into account. First, the calculations do not account for absorption of the radiation in the plasma. The theoretical predictions may be somewhat different if absorption is included. Second, questions relating to the experimental geometry and the finite pulse duration may be important: a) the minimum size of the scattering region (transverse beam diameter) must exceed the wavelength, and b) the time for the instabilities under consideration to develop must be shorter than the pulse duration.

Linearly polarized radiation is mainly used in experiments with pulses at relativistic intensities, while all the results presented here apply to a circularly polarized reference wave. The group of physical effects outlined above, however, will also show up in the case of a linearly polarized reference wave. There are some differences. Thus, in the case of linear polarization, the scattering will not be azimuthally symmetric.

This work was partially supported by the Russian Fund for Fundamental Research (Grant Nos. 95-02-05194, 96-02-16401, and 96-02-18264) and by INTAS (Grant No. 94-1937).

*)E-mail: borovsky@kapella.gpi.ru

[1] D. S. Montgomery, S. H. Batha, K. S. Bradley *et al.*, in *ICF Annual Report*, UCRL-LR-105820-91, p. 13.
[2] D. Umstadter, W. B. Mori, and C. Joshi, Phys. Fluids B **1**, 183 (1989).
[3] S. H. Batha, D. D. Meyerhofer, A. Simon, and R. P. Drake, Phys. Fluids B **3**, 448 (1991).
[4] D. Milam and D. Eimerl, in *ICF Quarterly Report*, UCRL-LR-105821-92-4, p. 151.
[5] K. Estabrook, W. L. Kruer, and M. G. Haines, Phys. Fluids B **1**, 1282 (1989).
[6] S. V. Bulanov, I. N. Inovenkov, V. I. Kirsanov, N. M. Naumova, and A. S. Sakharov, Phys. Fluids B **4**, 1935 (1992).
[7] V. M. Gorbunov and V. I. Kirsanov, Trudy FIAN **219**, 3 (1992).
[8] V. I. Kirsanov and A. S. Sakharov, Fiz. Plazmy **21**, 623 (1995) [Plasma Phys. Rep. **21**, 587 (1995)].
[9] V. I. Kirsanov and A. S. Sakharov, Fiz. Plazmy **21**, 632 (1995) [Plasma Phys. Rep. **21**, 596 (1995)].
[10] A. V. Borovskiĭ, A. L. Galkin, V. V. Korobkin, and O. B. Shiryaev, Kvant. Élektron. **24**, 929 (1997).
[11] A. B. Borisov, A. V. Borovskiy, V. V. Korobkin, A. M. Prokhorov, O. B. Shiryaev, X. M. Shi, T. S. Luk, A. McPherson, J. L. Solem, K. Boer, and C. K. Rhodes, Phys. Rev. Lett. **68**, 2309 (1992).
[12] G. Mourous and D. Umstadter, Phys. Fluids B **4**, 2315 (1992).
[13] F. G. Patterson, R. Gonzales, and M. D. Perry, Opt. Lett. **16**, 1107 (1991).

[14] V. P. Silin, *Parametric Interaction of High Intensity Radiation with Plasmas* [in Russian], Nauka, Moscow (1973).
[15] N. Bloembergen, *Nonlinear Optics*, Benjamin, New York (1965).
[16] S. Kelikh, *Molecular Nonlinear Optics* [in Russian], Nauka, Moscow (1966).
[17] N. E. Andreev, Zh. Éksp. Teor. Fiz. **59**, 2105 (1970) [Sov. Phys. JETP **32**, 1141 (1970)].
[18] L. M. Gorbunov, Usp. Fiz. Nauk **109**, 631 (1973) [Sov. Phys. Usp. **16**, 217 (1973)].
[19] S. Jorna, Phys. Fluids **17**, 765 (1974).
[20] V. N. Tsytovich, *The Theory of Plasma Turbulence* [in Russian], Atomizdat, Moscow (1971).
[21] C. J. McKinstrie, A. Simon, and E. A. Williams, Phys. Fluids **27**, 2738 (1984).
[22] R. L. Berger, Phys. Fluids **27**, 1796 (1984).
[23] A. I. Akhiezer and R. V. Polovin, Zh. Éksp. Teor. Fiz. **30**, 915 (1956) [Sov. Phys. JETP **3**, 696 (1956)].
[24] V. G. Priĭmak, Doctoral Dissertation, Institute of Mathematical Modelling, Moscow (1996).
[25] A. V. Borovskiĭ and A. L. Galkin, *Linear Physics* [in Russian], IzdAT, Moscow (1996).
[26] J. H. Wilkinson and S. Reinisch, *Linear Algebra*, Springer, Berlin (1971).
[27] L. D. Landau and E. M. Lifshitz, *Theory of Fields* [in Russian], Nauka, Moscow (1982).
[28] Yu. L. Klimontovich, *Statistical Physics* [in Russian], Nauka, Moscow (1982).

Translated by D. H. McNeill

Ionization of atoms in weak fields and the asymptotic behavior of higher-order perturbation theory

V. S. Popov

Institute of Theoretical and Experimental Physics, 117218 Moscow, Russia

A. V. Sergeev

Department of Chemistry, Southern Methodist University, P.O. Box 750314, Dallas, Texas, USA

(Submitted 19 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2047–2055 (June 1998)

Using the imaginary time method, we study the structure of the perturbation series for the hydrogen atom in electric $\mathscr{E}$ and magnetic $\mathscr{H}$ fields. It is shown that there is a ''critical'' value of the ratio $\gamma=\mathscr{H}/\mathscr{E}$ at which the perturbation series for the ground state changes from having a fixed sign (for $\gamma<\gamma_c$) to having a variable sign (for $\gamma>\gamma_c$). This conclusion is confirmed by direct higher-order perturbation calculations. The change in the asymptotic regime is explained by competition among the contributions of the various complex trajectories that describe the subbarrier motion of the electrons. Here the parameter γ_c depends on the angle θ between the electric and magnetic fields. © *1998 American Institute of Physics.* [S1063-7761(98)01006-3]

1. The problem of the hydrogen atom in external fields is of fundamental importance in quantum mechanics and atomic physics and is often encountered in applications.[1–5]

Recently,[6–8] a semiclassical theory has been developed for the ionization of atoms and ions in constant and uniform electric $\mathscr{E}$ and magnetic $\mathscr{H}$ fields. The imaginary time method[9–11] was used to calculate the ionization probability $w(\mathscr{E},\mathscr{H})$, as it yields a convincing description of the subbarrier motion of the particles using the classical equations of motion, but with an imaginary ''time.'' [1)]

The ionization probability for the atomic s level with binding energy $|E_0|=\kappa^2/2$ is given by ($\hbar=m=e=1$, natural units)

$$w(\mathscr{E},\mathscr{H})=\kappa^2 R(\gamma,\theta)\ \exp\left\{-\frac{2}{3\epsilon}\,g(\gamma,\theta)\right\}, \tag{1}$$

which is asymptotically exact in the limit of weak fields ($\epsilon,h\ll 1$). Here $\epsilon=\mathscr{E}/\kappa^3\mathscr{E}_a$ and $h=\mathscr{H}/\kappa^2\mathscr{H}_a$ are the reduced electric and magnetic fields, θ is the angle between the fields, $\mathscr{E}_a=m^2e^5/\hbar^4$ and $\mathscr{H}_a=m^2ce^3/\hbar^3$ are the atomic units for the field strengths, m is the electron mass, $\gamma=h/\epsilon$,

$$g(\gamma,\theta)=\frac{3}{2}\,\beta\left[1-\frac{\sqrt{\beta^2-1}}{\gamma}\,\sin\,\theta-\frac{1}{3}\,\beta^2\,\cos^2\,\theta\right], \tag{2}$$

$\beta=\tau_0/\gamma$, $\tau_0=\tau_0(\gamma,\theta)$ is the positive root of the equation

$$\tau^2-\sin^2\,\theta(\tau\,\coth\,\tau-1)^2=\gamma^2, \tag{3}$$

and, finally, R is a (rather complicated) factor introduced in Ref. 7: $R=2^{2\eta}\epsilon^{1-2n}PQ^{\eta}$ in the notation given there. Equation (3) can be easily obtained using the imaginary time method, where τ_0 has a simple physical significance: $\tau_0=-i\omega_L t_0$, where $\omega_L=|e|\mathscr{H}/mc$ is the Larmor frequency and t_0 is the ''time'' (purely imaginary) for subbarrier motion of the electron. Note that for $\gamma\to 0$,

$$\beta=1+\frac{\sin^2\,\theta}{18}\,\gamma^2+\ldots,$$

while for $\gamma\to\infty$,

$$\beta(\gamma,\theta)=\begin{cases}\dfrac{1}{\cos\,\theta}-\tan^2\,\theta\gamma^{-1}+O(\gamma^{-2}), & 0\leqslant\theta<\pi/2,\\[2ex] \dfrac{\gamma}{2}\,(1+\gamma^{-2}+\ldots), & \theta=\pi/2.\end{cases} \tag{3a}$$

Thus, for the function g, which determines the principal (exponential) factor in the ionization probability, we obtain

$$g(\gamma,\theta)=1+O(\gamma^2),\quad \gamma\to 0,$$

$$g(\gamma,\theta)=\frac{1}{\cos\,\theta}-\frac{3}{2}\,\tan^2\,\theta\cdot\gamma^{-1}+\ldots,\quad \gamma\to\infty \tag{3b}$$

(for $\theta<\pi/2$; for $\theta=\pi/2$ the asymptote has a different form; see Eq. (16) below). The function $g(\gamma,\theta)$ increases monotonically[2)] along with γ (Fig. 1), so raising the magnetic field (at fixed $\mathscr{E}$) sharply reduces the ionization probability, stabilizing the atomic level.[6,7]

Using Eqs. (1)–(3) and invoking the same considerations as before,[15,16] one can obtain the asymptotic behavior of the higher orders of perturbation theory, which is the subject of this paper. We note that higher-order perturbation theory has been studied for use in many quantum mechanical problems: the anharmonic oscillator,[20–22] the Yukawa and funnel potentials,[23–26] the Stark[27–33] and Zeeman[34–36] effects in the hydrogen atom, the molecular hydrogen ion, etc., as well as for $1/n$-expansions.[14–17] The problem examined below is of interest in that the asymptotic regime undergoes a change at a certain value of $\gamma=\gamma_c$: the perturbation series switches from a constant sign series to an alternating series, which is

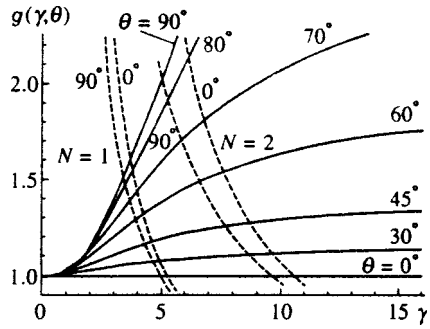


FIG. 1. The function $g(\gamma,\theta)$ (smooth curves next to which the angle θ is indicated) as a function of the parameter γ. The dashed curves are the values of $|g_c(\gamma,\theta)|$ for $N=1$ and 2, corresponding to the solution of Eqs. (10)–(12).

explained by examining a new class of complex subbarrier trajectories besides the usual subbarrier trajectory.

The asymptotic behavior of the higher orders of perturbation theory is of interest from a general standpoint, but is also of practical importance for calculating the shifts of atomic levels and their widths $\Gamma=\hbar\omega(\mathscr{E},\mathscr{H})$, using special procedures for summing diverging series, such as the Borel or Padé–Borel summation techniques.[28–33,37,38]

2. In calculating the energy levels of atoms in an electric field $\mathscr{E}$, the standard approach is to expand the energy in a perturbation series,

$$E(\mathscr{E})=\sum_{k=0}^{\infty} E_k\mathscr{E}^k. \tag{4}$$

According to Dyson's argument,[39] the instability of the state (complex energy $E=E_r-i\Gamma/2$, where Γ is the level width) is related to the divergence of the perturbation theory series. We shall study the behavior of the higher orders of perturbation theory in the presence of a magnetic field. To evaluate the behavior of the perturbation coefficients E_k as $k\to\infty$, we use the dispersion relation[20,28,33]

$$E_k=\frac{1}{2\pi i}\oint\frac{E(\mathscr{E})}{\mathscr{E}^{k+1}}\,d\mathscr{E}=-\frac{1}{2\pi}\int_0^{\infty}\frac{\Gamma(\mathscr{E})}{\mathscr{E}^{k+1}}\,d\mathscr{E} \tag{5}$$

(here we have taken advantage of the familiar analytic properties of the function $E(\mathscr{E})$, in particular its behavior on a large circle[32]: $|E(\mathscr{E})|\propto(\mathscr{E}\ln\mathscr{E})^{2/3}$ as $\mathscr{E}\to\infty$ for the ground state of hydrogen).

The asymptotic behavior of the higher orders of perturbation theory is determined by the level width $\Gamma(\mathscr{E})$ in an arbitrarily small field, so that it is possible to use the semiclassical Eq. (1). Equation (3) is even with respect to τ and has a pair of roots $\pm\tau_0$, for which the values of $g(\gamma,\theta)$ differ in sign. Given this, Eqs. (1) and (5) imply that

$$E_k\approx\frac{1+(-1)^k}{2}\,k!\,a^k k^{\beta}\left(c_0+\frac{c_1}{k}+\ldots\right),\quad k\to\infty, \tag{6}$$

$$a=3[2\kappa^3 g(\gamma,\theta)]^{-1} \tag{7}$$

(in the case of the ground state, the odd orders of the perturbation expansion for the energy $E(\mathscr{E})$ vanish identically). In the following we examine only the even orders of the perturbation expansion, omitting the factor $[1+(-1)^k]/2$. In particular, for the Stark effect in the hydrogen atom, we have[28]

$$E_k\approx-\frac{6}{\pi}\,k!\left(\frac{3}{2}\right)^k\left[1-\frac{107}{18k}+\frac{7363}{648k^2}+\ldots\right],\quad a=\frac{3}{2},\quad \beta=0 \tag{6a}$$

(in the ground state $E_0=-1/2$, $\kappa=\sqrt{-2E_0}=1$).

Besides the Stark expansion (4), let us consider the expansion of the ground state energy in powers of the magnetic field:

$$E=\sum_{k=0}^{\infty}\tilde{E}_k\mathscr{H}^k,\quad \tilde{E}_k=\gamma^{-k}E_k. \tag{8}$$

In the case of the Zeeman effect ($\gamma\to\infty$), the higher orders of this expansion also increase factorially:[35,36]

$$\tilde{E}_k\approx(-1)^{(k+2)/2}\left(\frac{4}{\pi}\right)^{5/2}\Gamma\left(k+\frac{3}{2}\right)\left(\frac{1}{\pi}\right)^k[1+O(k^{-1})] \tag{6b}$$

(k even), which corresponds formally to the asymptote (6) with a purely imaginary parameter $\tilde{a}=a/\gamma=\pm(\pi i)^{-1}$. At the same time, for $\gamma\gg1$, by virtue of Eqs. (3b) and (7), we have

$$\tilde{a}=3/2\,\gamma g(\gamma,\theta)\approx\frac{3}{2\gamma}\cos\theta\to0,$$

which is inconsistent with the previous result. This suggests the existence of other solutions (i.e., complex subbarrier trajectories for which the parameter $\tilde{a}$ does not vanish in the limit of a strong magnetic field). We shall show that this is indeed so, by solving Eq. (3) for $\gamma\to\infty$ in the complex plane.

Taking $\tau=i\tilde{\tau}$ and $\gamma=i\tilde{\gamma}$, we rewrite Eq. (3) in the form

$$\tilde{\tau}^2+\sin^2\theta(1-\tilde{\tau}\cot\tilde{\tau})^2=\tilde{\gamma}^2. \tag{9}$$

There are two possibilities as $\gamma\to\infty$: either $\tilde{\tau}_0\to\pm i\gamma/\cos\theta$ (here $\cot\tilde{\tau}_0\to\mp i$) or $\tilde{\tau}_0\to\pm N\pi$ for integer $N\neq0$ ($\cot\tilde{\tau}_0\to\infty$). The first possibility corresponds to the real solution considered above. In the second case we obtain

$$\tilde{\tau}_0=N\pi+N\pi\sin\theta\cdot\tilde{\gamma}^{-1}+\frac{1}{2}(N\pi)^3\sin\theta\times\left[\left(1-\frac{2}{3}\sin^2\theta\right)\tilde{\gamma}^{-3}+\sin\theta\left(1-\frac{1}{3}\sin^2\theta\right)\tilde{\gamma}^{-4}\right]+\ldots \tag{10}$$

($N=1,2,\ldots$), with $\tilde{\tau}_0^*$, $-\tilde{\tau}$, and $-\tilde{\tau}_0^*$ also solutions of Eq. (9). Introducing the function

$$G(\gamma,\theta)=\frac{2\tilde{\gamma}}{3\pi}\,g_c(\gamma,\theta)=\frac{2\tilde{\tau}_0^3}{3\pi\tilde{\gamma}^2}\left\{1+\frac{1}{2}\sin^2\theta\left[1+3\cot\tilde{\tau}_0\left(\cot\tilde{\tau}_0-\frac{1}{\tilde{\tau}_0}\right)\right]\right\} \tag{11}$$

and substituting Eq. (10) into it, some simple but cumbersome calculations yield

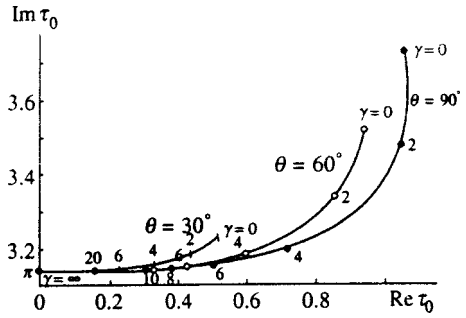


FIG. 2. Solutions of Eq. (9) in the complex plane for θ=30°, 60°, and 90° ($N=1$). The values of the parameter γ=0,2,4,... are indicated on the curves.

$$G(\gamma,\theta)=N\Bigg\{1-2\sin\theta\cdot\tilde{\gamma}^{-1}+\left(\sin^2\theta-\frac{(N\pi)^2}{3}\cos^2\theta\right)\tilde{\gamma}^{-2}+(N\pi)^2\sin\theta\left(\frac{2}{3}\sin^2\theta-1\right)\cdot\tilde{\gamma}^{-3}+(N\pi)^2\sin^2\theta\left(\frac{1}{3}\sin^2\theta-1\right)\tilde{\gamma}^{-4}-(N\pi)^2\sin\theta\left[\frac{1}{3}\sin^2\theta+(N\pi)^2\left(\frac{2}{15}\sin^4\theta-\frac{1}{3}\sin^2\theta+\frac{1}{4}\right)\right]\tilde{\gamma}^{-5}+\ldots\Bigg\}. \tag{12}$$

These expansions are valid for $\gamma\to\infty$. On the other hand, the value of $\tilde{\tau}_0$ at $\gamma=0$ can be determined from the equation $\cot\tilde{\tau}_0-1/\tilde{\tau}_0=\pm i/\sin\theta$. A numerical analysis of Eq. (9) shows that as γ increases, the point $\tau_0=i\tilde{\tau}_0(\gamma,\theta)$ describes the curves shown in Fig. 2 in the complex plane.

We shall mainly be interested in the case $N=1$, where the function $G(\gamma,\theta)$ has its minimum absolute value. The values of $|g_c|=|(3\pi/2\tilde{\gamma})G(\gamma,\theta)|$ for $N=1$ and 2 are shown in Fig. 1 as the dotted curves. For sufficiently large γ, when $|g_c|<g$, the asymptotic parameter a can be found from Eq. (7) by replacing g with g_c. Because of the existence of a pair of complex conjugate solutions $\tilde{\tau}_0$ and $\tilde{\tau}_0^*$, the asymptotes of the higher-order perturbation theory now have the form

$$\tilde{E}_k\sim(-1)^{k/2}\,\mathrm{Re}(CA^k)k!k^{\beta},\quad A=i\tilde{a}=3/2|g_c(\gamma,\theta)|, \tag{13}$$

so that the perturbation series is alternating for sufficiently large k.

In the limit $\theta\to0$ (parallel $\mathscr{E}$ and $\mathscr{H}$ fields), the expansion (12) terminates at the third term, so a solution can be obtained in analytic form:

$$G(\gamma,0)=N\left[1-\frac{(N\pi)^2}{3\tilde{\gamma}^2}\right],$$

$$g_c(\gamma,0)=i\,\frac{3N\pi}{2\gamma}\left[1+\frac{(N\pi)^2}{3\gamma^2}\right]. \tag{14}$$

The condition $|g_c(\gamma,0)|=g(\gamma,0)=1$ determines the ''critical'' value of γ (Ref. 6):

$$\gamma_c=\pi[(1+\sqrt{2})^{1/3}-(1+\sqrt{2})^{-1/3}]^{-1}=5.270495...,\quad N=1 \tag{15}$$

(see Appendix). For $\gamma<\gamma_c$, i.e., in sufficiently strong electric fields, the dominant contribution to the asymptotic E_k is from the subbarrier trajectory with real τ_0, corresponding to the function $g(\gamma,0)$, and the perturbation series has a fixed sign. If, however, $\gamma>\gamma_c$, then $a_c=1.5|g_c|^{-1}>a=1.5|g|^{-1}$, so the signs of the higher orders of perturbation theory should alternate according to Eq. (13). Thus, at $\gamma=\gamma_c$ the structure of the perturbation series changes.

We have verified this by direct calculation of the perturbation series coefficients E_k up to $k=80$ (see Table I). (For $k\leqslant10$ our calculations agree with an earlier paper[40] and for γ=0, with Refs. 28–33). Some of these results are shown in Fig. 3. It has been shown that between γ=5 and 5.5, the order of the signs[3)] of the coefficients E_k does indeed change. In addition, for $\gamma<\gamma_c$ the coefficients $E_k(\gamma)$ are all of the same order of magnitude (since $g(\gamma,\theta)\equiv1$ and the asymptotic parameter $a=3/2$ is independent of γ), while for $\gamma>\gamma_c$ they begin an additional (and very rapid!) growth in accordance with the reduction in $|g_c(\gamma)|$, which is clearly evident in Fig. 3 (see also Eq. (A4)).

In the case considered here (θ=0), the critical value of the parameter $\gamma=h/\epsilon$ can be found analytically. It is interesting to study the structure of the perturbation series in the more general case as well, especially for mutually perpendicular fields. The value of γ_c that determines the restructuring of the perturbation series can be found from the condition $g=|g_c|$, where[4)]

$$g\left(\gamma,\frac{\pi}{2}\right)=\frac{3}{8}\,\gamma(1+2\gamma^{-2}+\gamma^{-4}+\ldots),\quad \gamma\to\infty, \tag{16}$$

$$G\left(\gamma,\frac{\pi}{2}\right)=1-\gamma^{-2}-\frac{2\pi^2}{3}\,\gamma^{-4}-\frac{2i}{\gamma}\left[1-\frac{\pi^2}{6}\,\gamma^{-2}+\frac{\pi^2}{6}\left(1+\frac{3\pi^2}{20}\right)\gamma^{-4}\right]+\ldots, \tag{17}$$

whence

$$\left|g_c\left(\gamma,\frac{\pi}{2}\right)\right|=\frac{3\pi}{2\gamma}\times\left[1+2\gamma^{-2}-\left(\frac{8\pi^2}{3}-1\right)\gamma^{-4}+O(\gamma^{-6})\right] \tag{17a}$$

and $\gamma_c\approx3.54$. This simple estimate is in good agreement with the numerical calculations (see the point of intersection of the smooth and dashed ($N=1$) curves for θ=90° in Fig. 1).

Similarly, we can calculate γ_c for arbitrary angles θ. It would be interesting to confirm the existence of a switch in the asymptotic regime at $\gamma=\gamma_c(\theta)$ by direct calculation of the higher perturbation orders, as has been done above for the case of parallel fields.

TABLE I. Higher orders of perturbation theory (hydrogen atom in parallel fields).

	$-E_k(\gamma)$				
k	$\gamma=0$	$\gamma=2$	$\gamma=5$	$\gamma=5.5$	$\gamma=10$
0	0.500	0.500	0.500	0.500	0.5000
2	2.250	1.2500	−4.000	−5.3125	−22.750
4	55.547	40.089	1.039(2)	1.578(2)	2.319(3)
6	4.908(3)	3.351(3)	−6.448(3)	−1.437(4)	−9.358(5)
8	7.942(5)	5.201(5)	1.195(6)	2.930(6)	6.971(8)
10	1.945(8)	1.232(8)	−2.232(8)	−8.027(8)	−7.817(11)
20	1.121(22)	6.574(21)	1.033(22)	1.015(23)	8.114(28)
30	7.898(37)	4.529(37)	3.485(37)	−1.405(39)	−9.211(47)
40	1.478(55)	8.389(54)	5.674(54)	5.015(56)	2.642(68)
50	3.279(73)	1.850(73)	−3.502(72)	−2.054(75)	−8.726(89)
60	5.282(92)	2.968(92)	9.221(91)	6.026(97)	2.054(112)
66	3.973(104)	2.228(104)	−6.181(101)	−6.445(106)	−1.217(126)
68	4.084(108)	2.289(108)	5.568(107)	7.450(110)	5.355(130)
70	4.449(112)	2.493(112)	5.862(110)	−9.115(114)	−2.497(135)
72	5.130(116)	2.873(116)	6.295(115)	1.181(119)	1.232(140)
74	6.250(120)	3.499(120)	1.558(119)	−1.614(123)	−6.420(144)
76	8.033(124)	4.496(124)	8.981(123)	2.329(127)	3.528(149)
78	1.088(129)	6.085(128)	3.728(127)	−3.537(131)	−2.042(154)
80	1.550(133)	8.667(132)	1.598(132)	5.654(135)	1.243(159)

Note. The table lists the coefficients in the perturbation theory series (4) for the ground state of the hydrogen atom taken with the opposite sign; k is the perturbation theory order; $a(b)\equiv a\cdot 10^b$.

3. Therefore, at $\gamma=\gamma_c$ there is a change in the character of the asymptotic behavior of the higher orders of perturbation theory.[5)] Upon going from one asymptotic regime to the other, the perturbation series switches (when $k\geqslant k_0$) from an alternating series to one with a constant sign, which shows up in the position of the singularities in the Borel transformants that are closest to zero, and therefore in the choice of an efficient method for taking the sum.[17,33,37]

The complex solutions of Eq. (3) found above correspond to complex subbarrier trajectories which, therefore, can be important in determining the asymptotic behavior of the higher orders of perturbation theory. Their physical significance can be clarified using the example of parallel $\mathscr{E}$ and $\mathscr{H}$ fields. It is known that the asymptotic behavior of the higher orders of perturbation theory is directly related to the tunneling probability for a particle in a potential with the ''wrong'' sign on the coupling constant, e.g., $g\to -g$ in the case of an anharmonic oscillator,

$$V(x)=\frac{1}{2}x^2+g\,\frac{x^4}{4}$$

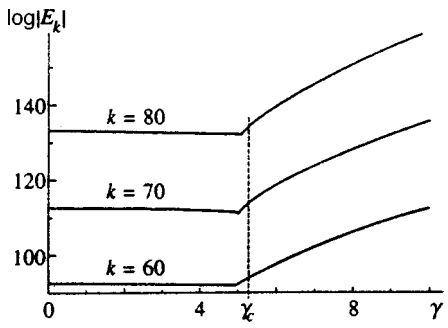


FIG. 3. Higher orders of perturbation theory (Eq. (4)) for the ground state of the hydrogen atom in parallel fields.

(the Dyson phenomenon[39,20]). In our problem, $\mathscr{H}^2$ plays the role of g. Going to purely imaginary values of the magnetic field ($\mathscr{H}=i\tilde{\mathscr{H}}$), we obtain a potential proportional to $-(1/8)\tilde{\mathscr{H}}^2\rho^2$, decreases without bound as $\rho=\sqrt{x^2+y^2}\to\infty$. It is evident that in such a potential, tunneling is possible both along the electric field (the z axis) and perpendicular to it. The complex solutions (10)–(12) probably correspond to an analytic continuation of ''perpendicular'' subbarrier trajectories of this sort from a region of purely imaginary magnetic fields into a region of real $\mathscr{H}$.

The authors thank V. M. Vaĭnberg and V. D. Mur for discussing this paper and for useful comments. This work was partially supported by the Russian Fund for Fundamental Research (Grant Nos. 95-02-05417 and 98-02-17007).

APPENDIX

The perturbation theory coefficients (4) for the energy of the ground state of the hydrogen atom are polynomials in γ^2:

$$E(\mathscr{E},\mathscr{H})=\sum_{i,j}c_{ij}\epsilon^{2i}h^{2j}=\sum_{k=0}^{\infty}E_{2k}(\gamma)\epsilon^{2k}, \tag{A1}$$

$$E_{2k}(\gamma)=\sum_{j=0}^{k}c_{k-j,j}\gamma^{2j}, \tag{A2}$$

where $2k$ is the order of perturbation theory, $\gamma=h/\epsilon=\alpha\mathscr{H}/\mathscr{E}$, and $\alpha=e^2/\hbar c$ is the fine structure constant. Several of the lowest orders of perturbation theory are known exactly, i.e., in the form of rational fractions, as

$$E_0=-\frac{1}{2},\quad E_2=-\frac{1}{4}(9-\gamma^2),$$

$$E_4=-\frac{1}{64}\left(3555-318\gamma^2+\frac{53}{3}\gamma^4\right),$$

$$E_6=-\frac{1}{512}\left(2\,512\,779-254\,955\gamma^2+\frac{49195}{3}\gamma^4-\frac{5581}{9}\gamma^6\right),$$

$$E_8=-2^{-12}\cdot\left(\frac{13\,012\,777\,803}{4}-\ldots-\frac{12\,368\,405}{9}\gamma^6+\frac{21\,577\,397}{540}\gamma^8\right), \quad \text{(A3)}$$

and were used to monitor the numerical calculations. The outer coefficients c_{k0} and c_{0k} in Eq. (A2) correspond to the Stark[27–31] and Zeeman[34] effects, while the cross terms $(1\leqslant j\leqslant k-1)$ were taken from Johnson *et al.*[40] and Lambin *et al.*[41] Here $c_{k-j,j}=2^{-3j}\epsilon^{(k-j,j)}$, where the $\epsilon^{(ij)}$ are coefficients tabulated (for the case of parallel fields) by Johnson *et al.*[40]

The asymptotes of the higher-order perturbation theory can be written in the form

$$E_k(\gamma)\approx -k!\left\{\frac{6}{\pi}c_0a^k+(-1)^{k/2}\left(\frac{4}{\pi}\right)^{5/2}c_1a_c^k k^{1/2}\right\}, \quad \text{(A4)}$$

where

$$a=\frac{3}{2}, \quad a_c=\frac{\gamma}{\pi}\left(1+\frac{\pi^2}{3\gamma^2}\right)^{-1}, \quad \text{(A5)}$$

$c_0=\gamma/\sinh\gamma$, and for c_1 we obtained (numerically) $c_1\approx 1-12.03\gamma^{-2}$ for $\gamma\gg 1$. The condition $a=a_c$ yields a cubic equation whose solution (according to the Cardano formula) leads to Eq. (15).

[1)]The imaginary time method was developed for the theory of multiphoton ionization of atoms and ions in strong optical fields,[9,10,12] and has also been used in the problem of electron–positron pair production from the vacuum in a variable electric field.[11,13] This method has been used[14–16] to study the asymptotic behavior of the higher orders of the $1/n$-expansion in multidimensional quantum mechanics problems, including the two-center Coulomb problem (another approach to this problem has appeared recently[17,18]).

[2)]This function was first calculated by Kotova *et al.*[19]

[3)]Numerical calculations show that sign $E_{2k}=(-1)^{k+1}$ for $\gamma\geqslant 5.5$ and $2k\leqslant 80$. On the other hand, for $\gamma\leqslant 5$ the coefficients $E_{2k}<0$ for sufficiently large $k\geqslant k_0$, where k_0 depends on γ and increases rapidly as it approaches γ_c. Thus, $2k_0=0$, 0, 4, and 68, respectively, for $\gamma=0$, 2, 4, and 5 (see Table I).

[4)]The first expansion follows from Eqs. (2) and (3), the second, from Eq. (12) with $N=1$. The parameter in these expansions is γ^{-2}, with $\gamma_c^{-2}\sim 0.08\ll 1$.

[5)]An analogous phenomenon occurs in the $1/n$-expansion in the problem of two Coulomb centers.[14–16] In this case the role of the parameter γ is played by the internuclear distance R.

[1] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Systems*, Handbuch der Physik, Vol. 35/1, Springer, Berlin (1957), p. 88.

[2] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Nonrelativistic Theory)*, 3rd ed., Pergamon, New York (1977).

[3] B. M. Smirnov, *Physics of Atoms and Ions* [in Russian], Énergoatomizdat, Moscow (1986).

[4] V. S. Lisitsa, Usp. Fiz. Nauk **153**, 379 (1987) [Sov. Phys. Usp. **30**, 927 (1987)].

[5] V. D. Mur and V. S. Popov, Laser Phys. **3**, 462 (1993).

[6] V. S. Popov and A. V. Sergeev, JETP Lett. **63**, 417 (1996).

[7] V. S. Popov, B. M. Karnakov, and V. D. Mur, Phys. Lett. A **229**, 306 (1997).

[8] B. M. Karnakov, V. D. Mur, and V. S. Popov, JETP Lett. **65**, 405 (1997).

[9] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].

[10] V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1966)].

[11] V. S. Popov, JETP Lett. **13**, 185 (1971); Zh. Éksp. Teor. Fiz. **61**, 1334 (1971) [Sov. Phys. JETP **34**, 709 (1972)].

[12] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions, and Decay in Nonrelativistic Quantum Mechanics* [in Russian], Nauka, Moscow (1971).

[13] M. S. Marinov and V. S. Popov, Yadernaya Fizika **15**, 1271 (1972); Fortschr. Phys. **25**, 373 (1977).

[14] V. S. Popov, A. V. Sergeev, and A. V. Shcheblykin, Zh. Éksp. Teor. Fiz. **102**, 1453 (1992) [Sov. Phys. JETP **75**, 787 (1992)].

[15] V. S. Popov and A. V. Sergeev, Phys. Lett. A **172**, 193 (1993); V. S. Popov, and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **105**, 568 (1994) [JETP **78**, 303 (1994)].

[16] V. S. Popov, in *New Methods in Quantum Theory*, Kluwer, Dordrecht (1996), p. 149.

[17] M. Lopez-Cabrera, D. Z. Goodson, D. R. Herschbach, and J. D. Morgan, Phys. Rev. Lett. **68**, 1992 (1992).

[18] S. Kais and D. R. Herschebach, J. Chem. Phys. **98**, 3990 (1993).

[19] L. P. Kotova, A. M. Perelomov, and V. S. Popov, Zh. Éksp. Teor. Fiz. **54**, 1151 (1968) [Sov. Phys. JETP **27**, 616 (1968)].

[20] C. M. Bender and T. T. Wu, Phys. Rev. Lett. **27**, 461 (1971); Phys. Rev. D **7**, 1620 (1973).

[21] A. D. Dolgov and V. S. Popov, Phys. Lett. B **79**, 403 (1978); A. D. Dolgov and V. S. Popov, Zh. Eksp. Teor. Fiz. **75**, 2010 (1978) [Sov. Phys. JETP **48**, 1012 (1978)].

[22] H. J. Silverstone, J. G. Harris, J. Cizek, and J. Paldus, Phys. Rev. A **32**, 1965 (1985).

[23] V. M. Vaĭnberg, V. L. Eletskiĭ, and V. S. Popov, Zh. Éksp. Teor. Fiz. **81**, 1567 (1981) [Sov. Phys. JETP **54**, 833 (1981)].

[24] A. V. Sergeev and A. I. Sherstyuk, Zh. Éksp. Teor. Fiz. **82**, 1070 (1982) [Sov. Phys. JETP **55**, 625 (1982)].

[25] V. Privman, Phys. Lett. A **81**, 326 (1981).

[26] E. R. Vrscay, Phys. Rev. A **33**, 1433 (1986).

[27] H. J. Silverstone, Phys. Rev. A **18**, 1853 (1978).

[28] H. J. Silverstone, B. G. Adams, J. Cizek *et al.*, Phys. Rev. Lett. **43**, 1498 (1979).

[29] V. L. Eletskiĭ and V. S. Popov, Dokl. Akad. Nauk SSSR **250**, 74 (1980).

[30] S. P. Alliluev, V. L. Eletsky, and V. S. Popov, Phys. Lett. A **73**, 103 (1979); **78**, 43 (1980).

[31] V. Privman, Phys. Rev. A **22**, 1833 (1980).

[32] L. Benassi, V. Grecchi, E. Harrell, and B. Simon, Phys. Rev. Lett. **42**, 704, 1430 (1979).

[33] S. P. Alliluev, V. M. Vaĭnberg, V. L. Eletskiĭ, and V. S. Popov, Zh. Éksp. Teor. Fiz. **82**, 77 (1982) [Sov. Phys. JETP **55**, 46 (1982)].

[34] A. Galindo and P. Pascual, Nuovo Cimento B **34**, 155 (1976).

[35] B. G. Adams, J. E. Avron, J. Cizek *et al.*, Phys. Rev. Lett. **43**, 691 (1979); Phys. Rev. A **21**, 1914 (1980).

[36] V. M. Vaĭnberg, V. A. Gani, and A. E. Kudryavtsev, Zh. Éksp. Teor. Fiz. **113**, 550 (1998) [JETP **86**, 305 (1998)].

[37] V. Franceschini, V. Grecchi, and H. J. Silverstone, Phys. Rev. A **32**, 1338 (1985).

[38] V. S. Popov, V. D. Mur, A. V. Sergeev *et al.*, Phys. Lett. A **149**, 418 (1990).

[39] F. J. Dyson, Phys. Rev. **85**, 631 (1952).

[40] B. R. Johnson, K. F. Scheibner, and D. Farrely, Phys. Rev. Lett. **51**, 2280 (1983).

[41] P. Lambin, J. C. van Hay, and E. Kartheuser, Am. J. Phys. **46**, 1144 (1978).

Translated by D. H. McNeill

On the possibility of using dark magneto-optical lattices to achieve Bose condensation of atoms

A. V. Taĭchenachev,*) A. M. Tumaĭkin, and V. I. Yudin

Novosibirsk State University, 630090 Novosibirsk, Russia
(Submitted 5 January 1998)
Zh. Éksp. Teor. Fiz. **113**, 2056–2064 (June 1998)

We show that in dark magneto-optical lattices, effects associated with the Bose statistics of atoms can be observed even at laser cooling temperatures (10^{-4}–10^{-6} K), which exceed evaporative cooling temperatures in magnetic traps by several orders of magnitude. Quasicondensation occurs, i.e., the wave function is formed over the distances on which atoms are localized near the bottom of a separate potential well. In addition, switching off the magnetic field adiabatically reduces the temperature significantly, as a result of which Bose condensation in the entire volume of the gas can be observed. We propose a configuration of the light and magnetic fields in which the shape of the three-dimensional magneto-optical potential is independent of the phases of the emerging light waves. © *1998 American Institute of Physics.* [S1063-7761(98)01106-8]

1. INTRODUCTION

In 1995, the efforts of many years to achieve Bose–Einstein condensation in atomic gases were crowned with success.[1–3] The new state of matter was created by trapping laser-cooled spin-polarized atoms in magnetic traps followed by evaporative cooling to temperatures T of order 10^{-7} K.

In the present paper we examine the possibility of using dark magneto-optical lattices to observe effects associated with the quantum statistics of the particles. Such lattices form because of the resonant interaction of atoms with transitions $F_g=F\to F_e=F$ (F is an integer) or $F_g=F\to F_e=F-1$ (F_g and F_e are the angular momenta of the ground and excited states) and an inhomogeneously polarized field under conditions of coherent trapping of populations in a static magnetic field.[4] The cold atoms, which are in the dark state of coherent trapping of populations, are localized near the points where this state is not destroyed by the magnetic field and hence the optical interaction is close to zero. Localization and cooling in such lattices for the one- and two-dimensional cases were studied in Refs. 5–7.

We suggest a field configuration for stable three-dimensional trapping of atoms with the $1\to1$ transition in a dark magneto-optical lattice. When the laser field is strong, the percentage modulation of the magneto-optical potential is determined by the Zeeman splitting of the ground state, $\hbar\Omega$, and the period, by the wavelength λ of the light, as in the case of a one-dimensional lattice considered by Konopleva *et al.*[7] In the limit $\Omega\gg\omega_r$ (here $\hbar\omega_r=\hbar^2k^2/2M$ is the one-photon recoil energy), tunneling between the wells can be ignored, and the distance between the low-lying vibrational levels in a potential well is of order $\hbar\sqrt{\Omega\omega_r}$. Then at temperatures $k_BT<\hbar\sqrt{\Omega\omega_r}$ the quantum statistical effects become significant at concentrations $n\lambda^3\geqslant1$, when on the average there is more than one atom to a well. For instance, Bose atoms, irrespective of their number, are practically all in the ground state. This leads to formation of a quasicondensate, where the phase of the wave function is fixed over distances of localization in a single well ($\ll\lambda$) and varies at random from well to well. Such quasicondensation manifests itself in sudden changes in the spatial distribution, heat capacity, etc. The jumps and fairly smooth, as in the transition at the critical point in a system of several particles in a single potential well.[8] Since the curvature of the magneto-optical potential near the minimum points is much larger than the curvature of the potential in the magnetic traps, the degeneracy of the atomic gas is substantial at temperatures ($T\sim10^{-4}$–10^{-6} K) that are much higher than those for magnetic traps.[1–3]

On the basis of a thermodynamic analysis we also show that when $n\lambda^3>1$, the temperature drops significantly (by several orders of magnitude) under adiabatic switch-off of the magnetic field. The possible result is Bose–Einstein condensation in the entire volume of the gas. In contrast to ordinary Bose condensation of free particles, a macroscopic wave function $|\Psi_{NC}\rangle$ is formed. This wave function is a spatially inhomogeneous coherent superposition of the ground-state Zeeman wave functions and is annihilated by the operator of the interaction with the laser field: $\hat{\mathbf{d}}\cdot\mathbf{E}|\Psi_{NC}\rangle=0$. From the practical viewpoint this method of adiabatic switch-off of the magnetic field can be considered a new, highly effective method of ultradeep cooling of atomic gases that uses dark magneto-optical lattices.

2. STATEMENT OF THE PROBLEM

We examine a gas consisting of Bose atoms resonantly interacting with a spatially inhomogeneous monochromatic laser field

$$\mathbf{E}(\mathbf{r},t)=\mathbf{E}(\mathbf{r})e^{-i\omega t}+\text{c.c.}$$

on the $F_g=1\to F_e=1$ transition in the presence of a constant magnetic field $\mathbf{B}$. As is known,[9] for all $F\to F$ transitions (F

is an integer) there are states of coherent trapping of populations, states that nullify the interaction with the field:

$$\hat{\mathbf{d}}\cdot\mathbf{E}(\mathbf{r})|\psi_{nc}(\mathbf{r})\rangle=0,$$

where $\hat{\mathbf{d}}$ is the is the atomic dipole moment operator. In our case ($F=1$) this state has the form[10]

$$|\psi_{nc}(\mathbf{r})\rangle=\frac{1}{|\mathbf{E}(\mathbf{r})|}\sum_{q=0,\pm1}E^q(\mathbf{r})|g,\mu=q\rangle, \tag{1}$$

where $E^q(\mathbf{r})$ are the components of the field $\mathbf{E}(\mathbf{r})$ in the cyclic base $\{\mathbf{e}_0=\mathbf{e}_z,\mathbf{e}_{\pm1}=\mp(\mathbf{e}_x\pm i\mathbf{e}_y)/\sqrt{2}\}$. The state (1) is a superposition of the ground-state Zeeman wave functions $|g,\mu\rangle$ with inhomogeneous coefficients and generally is an eigenstate of neither the operator of interaction with the magnetic field, $\hat{H}_B=-(\hat{\boldsymbol{\mu}}\cdot\mathbf{B})$, nor the kinetic-energy operator $\hat{H}_K=\hat{p}^2/2M$. However, the corrections introduced by translational motion and the magnetic field can be considered small perturbations in relation to the interaction with the laser field provided that

$$V(\mathbf{r})\sqrt{G}\gg k\bar{v},\Omega, \tag{2}$$

where $V(\mathbf{r})=|\langle\hat{d}\rangle\mathbf{E}(\mathbf{r})|/\hbar$ is the local Rabi frequency, $G=V^2(\mathbf{r})/[\gamma^2/4+\delta^2+V^2(\mathbf{r})]$ is the saturation parameter with allowance for field broadening of the levels (where δ is the detuning of the frequency of the light field from resonance, and γ is the radiative relaxation rate of the excited state), and $\bar{v}$ is the average velocity of the atoms. In this case the majority of atoms are in a dark state $|\psi_{nc}(\mathbf{r})\rangle$, since when the condition (2) is met the relative populations of atoms in the state of coherent population, n_{nc}, and in the excited state, n_e, satisfy

$$(1-n_{nc})\sim n_e\sim\left(\frac{\max\{k\bar{v},\Omega\}}{V(\mathbf{r})\sqrt{G}}\right)^2\ll1.$$

To the same accuracy, the evolution of an individual atom can be described by the effective one-particle Hamiltonian

$$\hat{H}^{(1)}_{\text{eff}}=\langle\psi_{nc}(\mathbf{r})|(\hat{H}_K+\hat{H}_B)|\psi_{nc}(\mathbf{r})\rangle. \tag{3}$$

Following the standard rules of many-particle quantum theory, we describe an ensemble of atoms by the following Hamiltonian in the second-quantization representation:

$$\hat{H}=\int\hat{b}^\dagger(\mathbf{r})\hat{H}^{(1)}_{\text{eff}}(\mathbf{r})\hat{b}(\mathbf{r})d^3r,$$

where $\hat{b}^\dagger(\mathbf{r})$ and $\hat{b}(\mathbf{r})$ are the creation and annihilation operators for atoms in the state (1) of coherent trapping of populations at the point $\mathbf{r}$.

Note that if condition (2) is met, due to the effect of coherent trapping of populations the strongest dipole–dipole interaction of atoms is strongly suppressed ($n_e n\lambda^3\gg1$), so that we can ignore it. We also ignore interatomic collisions, i.e., we use the ideal-gas approximation.

3. THE EFFECTIVE HAMILTONIAN FOR ATOMS IN A DARK STATE

Using the explicit form (1) for the state of coherent trapping of populations, we write the one-particle Hamiltonian (3) as a sum of four terms:

$$\hat{H}^{(1)}_{\text{eff}}=\frac{\hat{p}^2}{2M}+U(\mathbf{r})+\frac{1}{2M}\{\mathbf{A}(\mathbf{r})\cdot\hat{\mathbf{p}}+\hat{\mathbf{p}}\cdot\mathbf{A}(\mathbf{r})\}+W(\mathbf{r}). \tag{4}$$

The first term on the right-hand side of Eq. (1) is the kinetic energy operator. The second is the magneto-optical potential

$$U(\mathbf{r})=\hbar\Omega\frac{i\mathbf{B}\cdot(\mathbf{E}(\mathbf{r})\times\mathbf{E}^*(\mathbf{r}))}{|\mathbf{B}||\mathbf{E}(\mathbf{r})|^2}, \tag{5}$$

and below we assume that $\Omega>0$. The last two corrections in (4) reflect the spatial inhomogeneity of the dark state. The first is of order kv and can be interpreted as an interaction with the effective vector potential

$$A_j(\mathbf{r})=-i\hbar\left(\frac{\mathbf{E}^*}{|\mathbf{E}|}\cdot\frac{\partial}{\partial x_j}\frac{\mathbf{E}}{|\mathbf{E}|}\right). \tag{6}$$

The second is of order the one-photon recoil energy $\hbar\omega_r$ and contributes to the potential energy of an atom:

$$W(\mathbf{r})=\frac{\hbar^2}{2M}\sum_j\left|\frac{\partial}{\partial x_j}\frac{\mathbf{E}}{|\mathbf{E}|}\right|^2. \tag{7}$$

If the Zeeman splitting satisfies

$$\Omega\gg k\bar{v},\omega_r, \tag{8}$$

the last two terms in (4) can be ignored. In this case

$$\hat{H}^{(1)}_{\text{eff}}\approx\frac{\hat{p}^2}{2M}+U(\mathbf{r}),$$

and the problem reduces to that of the motion of a structureless particle in the magneto-optical potential (5), for which the depth is determined by the magnetic field and the spatial periodicity by the wavelength of the light.

4. FIELD CONFIGURATION FOR THREE-DIMENSIONAL TRAPPING OF ATOMS

Obviously, the potential (5) can be used to create a three-dimensional lattice of atoms if the laser field $\mathbf{E}(\mathbf{r},t)$ is a finite set of plane waves. However, a certain problem arises in this case: the shape of the potential (5) generally depends on the phases of the emerging plane waves, and usually these phases cannot be controlled. Changes in the phases of fields can lead, among other things, to a violation of the condition of stable three-dimensional atom trapping, i.e., the minima of the potential form continuous lines or flat surfaces. Hence of special interest from the standpoint of the experimenter are such configurations of the light fields for which the shape of the potential is independent of the choice of phases, so that it leads only to a shift in the spatial pattern as a whole.

In this paper we propose such a configuration, which is formed by five linearly polarized laser beams (Fig. 1):

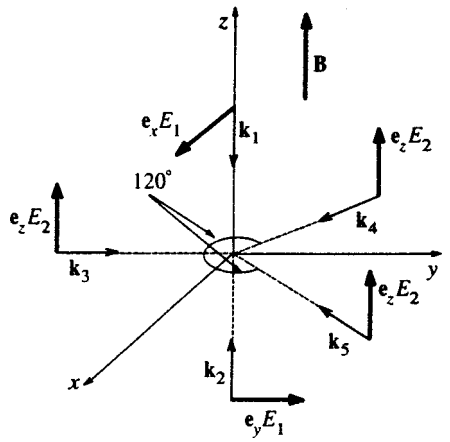


FIG. 1. Configuration of the optical and magnetic fields in which the shape of the magneto-optical potential (5) is independent of the phases of the emerging light waves and the choice of the origin of coordinates reduces the potential to (9).

$$\mathbf{E}(\mathbf{r})=\mathbf{e}_x E_1 \exp(i\mathbf{k}_1\cdot\mathbf{r})+\mathbf{e}_y E_1 \exp(i\mathbf{k}_2\cdot\mathbf{r}) + \mathbf{e}_z E_2\{\exp(i\mathbf{k}_3\cdot\mathbf{r})+\exp(i\mathbf{k}_4\cdot\mathbf{r})+\exp(i\mathbf{k}_5\cdot\mathbf{r})\}.$$

The two oppositely directed beams propagating along the z axis ($\mathbf{k}_2=-\mathbf{k}_1=|k|\mathbf{e}_z$) have the same amplitude E_1 and mutually orthogonal linear polarizations (the lin⊥lin configuration). The other three beams have equal amplitudes E_2 and are linearly polarized along $\mathbf{e}_z$; their wave vectors $\mathbf{k}_3$, $\mathbf{k}_4$, and $\mathbf{k}_5$ lie in the xy plane and form an angle of 120° with each other.

It can be shown that for any phases of the emerging waves the magneto-optical potential (5) can be reduced, via an appropriate selection of the origin of coordinates, to the form

$$U(\mathbf{r})=\frac{-2\hbar\Omega\,\cos(2kz)}{2+\left|\frac{E_2}{E_1}\right|^2\left\{\sin^2\frac{3ky}{2}+\left(\cos\frac{3ky}{2}+2\cos\frac{\sqrt{3}kx}{2}\right)^2\right\}}. \tag{9}$$

The minima of this potential are located in the planes $z=m\lambda/2$ ($m=0,\pm1,\ldots$), with each forming a two-dimensional lattice consisting of regular triangles with sides equal to $2\lambda/3$. At these points the polarization of the field $\mathbf{E}(\mathbf{r})$ is left-hand circular, and the state (1) of coherent trapping of populations coincides with the Zeeman sublevel $|g,\mu=-1\rangle$ and is not destroyed by the magnetic field $\mathbf{B}$ (Fig. 2). As shown in Ref. 7, the localization of atoms near these points leads to additional stability of the states of coherent trapping of populations opposed to the destructive effect of magnetic field. As a result, for localized atoms we have a stronger estimate:

$$(1-n_{nc})\sim n_e\sim\left(\frac{\Omega}{V\sqrt{G}}\right)^2\sqrt{\frac{\omega_r}{\Omega}}\ll 1,$$

which because of the factor $\sqrt{\omega_r/\Omega}\ll 1$ differs substantially from the general estimate of Sec. 1.

We can easily show that the potential (9) corresponds to three-dimensional periodic lattice with a unit-cell volume $u=\lambda^3/3\sqrt{3}$. It is well-known that in a periodic potential the energy spectrum has a band structure. However, in view of (8), for the lower bands we can ignore the tunneling of particles from one well to another. Here the widths of the lower bands are exponentially small in the parameter $\exp(-\sqrt{\Omega/\omega_r})$, and positions of these bands can be found from a harmonic expansion of the potential (9) near the bottom of an isolated well:

$$U(\mathbf{r})\approx-\hbar\Omega\left[1-\frac{9|E_2|^2}{8|E_1|^2}k^2(x^2+y^2)-2k^2z^2\right].$$

We see that the distance between the two lowest energy bands for commensurable values of the amplitudes $|E_1|$ and $|E_2|$ is of order $\hbar\sqrt{\Omega\omega_r}$ (Fig. 3). Putting $|E_2|=4|E_1|/3$, we arrive at the case of a three-dimensional isotropic harmonic oscillator with a frequency $\Omega_{\rm osc}=\sqrt{8\Omega\omega_r}$. For the sake of definiteness, below we examine this case.

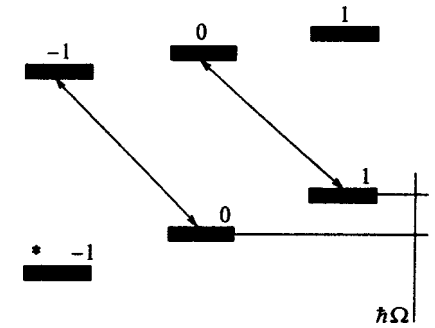


FIG. 2. Light-induced transitions (denoted by arrows) at the localization points. The dark state (1) coincides with the Zeeman sublevel $|g,\mu=-1\rangle$ (denoted by an asterisk).

5. QUASICONDENSATION

At temperatures $k_BT<\hbar\Omega$ the atoms are in finite motion in the magneto-optical potential (9). As is known, the effects of quantum statistics become appreciable when the de Broglie wavelength becomes equal to the average interatomic distance. In our case this corresponds to the condition $k_BT<\hbar\Omega_{\rm osc}(nu)^{1/3}$, where nu is the average number of atoms per unit cell ($nu>1$).

If we allow for the finite value of the width $\Delta\varepsilon_0$ of the lower energy band, we can isolate two regions. First,

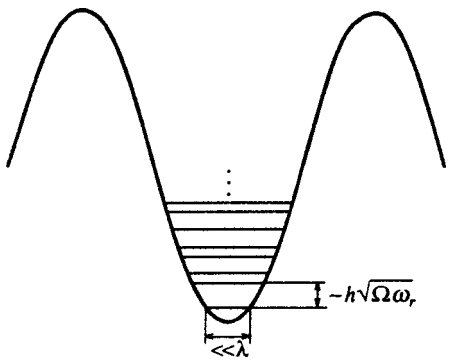


FIG. 3. A rough diagram of the low-lying vibrational levels in the magneto-optical potential (9).

$$\Delta\varepsilon_0 < k_B T < \hbar\Omega_{\rm osc}(nu)^{1/3}, \tag{10}$$

where practically all Bose atoms are in the lower energy band with an equiprobable distribution over the quasimomenta. This corresponds to quasicondensation, where the wave function of the atoms is formed over distances of localization ($\ll\lambda$) in a single well. When we go from one well to another, the phase of this function changes at random, i.e., there is global condensation in the entire volume of the lattice. In the interval (10), the volume density S of the system's entropy is weakly temperature-dependent and in the zeroth-order in $(\Delta\varepsilon_0/k_B T)\ll 1$ has the value

$$S \approx n\left[\frac{\ln(1+nu)}{nu} + \ln\left(1+\frac{1}{nu}\right)\right], \tag{11}$$

which can be obtained from combinatorial analysis:

$$S = \lim_{V\to\infty} \frac{\ln c}{V},$$

where c is the number of different arrangements of nV different objects (atoms) in V/u cells (quasimomenta of the lower band) when each cell can contain any number of objects (Bose statistics), i.e.,

$$c = \frac{(nV+V/u-1)!}{(nV)!(V/u-1)!}.$$

Using the well-known Stirling's asymptotic formula for $a!$ as $a\to\infty$ $(a!\approx a^a e^{-a}\sqrt{2\pi a})$, we arrive at (11). Formula (11) can also be obtained from the grand canonical distribution.

Second, if we have $k_B T < \Delta\varepsilon_0$, order in the quasimomenta of the lower band sets in. Here, as expected, the system's entropy tends to zero. However, if the conditions (8) are met, $\Delta\varepsilon_0$ is exponentially small (as noted earlier), so that the temperature range $k_B T < \Delta\varepsilon_0$ is of no practical interest.

6. BOSE CONDENSATION WHEN THE MAGNETIC FIELD IS SWITCHED OFF ADIABATICALLY

As the magnetic field gets weaker, the wave-function localization distances in each well increase, the tunneling of atoms from well to well becomes more intense, and the wave functions overlap. This may lead (at least in principle) to correlation in phases of the atomic wave functions and to Bose condensation in the entire volume of the gas.

Let us do a thermodynamic estimate of the possibility of such an effect when the magnetic field is switched off adiabatically. The thermodynamic condition that a process be adiabatic is the constancy of the entropy, $S=\text{const}$. For the initial conditions we take the temperature range (10). Then the value of S is given by (11). To estimate the values of the thermodynamic parameters in a zero magnetic field, we use the approximation of an ideal gas of free structureless particles. For instance, the entropy of an ideal Bose gas, S_I, at the critical point T_c is $S_I(T_c)\approx 1.28n$ (see Ref. 11). Then, if we want to reduce the temperature below the critical temperature by switching off the magnetic field adiabatically so that Bose condensation will occur, S must be smaller than $S_I(T_c)$ (this follows from the fact that entropy is a monotonic function of temperature). According to (11), this yields the following inequality:

$$\frac{\ln(1+nu)}{nu} + \ln\left(1+\frac{1}{nu}\right) < 1.28, \tag{12}$$

which is valid if $nu > 1.17$. Thus, adiabatic switch-off of the magnetic field sharply reduces the temperature, a process that outwardly resembles magnetic cooling. However, the physics of the process implies that we are dealing with cooling due to gas expansion. Indeed, in our case the magnetic field creates a magneto-optical potential and a strong spatial localization of the atoms in regions whose dimensions are much smaller than λ. When the magnetic field is switched off, the atoms become delocalized, which corresponds to effective expansion of the gas.

Note that to achieve a more rigorous estimate of the state of the system in the absence of a magnetic field we must take into account the last two corrections in (4), which are due to the periodic inhomogeneity of the state (1) of coherent trapping of populations over distances of the order of the light wavelength λ. However, for $nu>1$ these corrections do not change the thermodynamic quantities much in comparison to a gas of free particles. For instance, from general physical considerations is it clear that even if we allow for the corrections (6) and (7), the expression for the entropy at the critical value T_c is $S(T_c)=Dn$, where the coefficient D reaches a saturation plateau as the parameter nu increases and is of order unity, as in the case of a gas of free particles. Here the factor $(nu)^{-1}\ln(1+nu)+\ln(1+(nu)^{-1})$ in (11) fairly rapidly decreases as nu grows (actually, it decreases as $(nu)^{-1}$). Thus, we see that we can always select a value f such that for $nu>f$ this factor, $(nu)^{-1}\ln(1+nu)+\ln(1+(nu)^{-1})$, is smaller than D, i.e., when the temperature drops below T_c as the magnetic field is switched off adiabatically and Bose condensation is possible even if one allows for the spatial inhomogeneity of the state (1) of the coherent trapping of populations.

At the same time, in determining the ground-state wave function of the system one must carefully take into account the terms (6) and (7) in (4). Most likely, the ground state of the system in our case is the well-known macroscopic dark state,[10]

$$|\Psi_{NC}\rangle = R|\mathbf{E}(\mathbf{r})||\psi_{nc}(\mathbf{r})\rangle = R\sum_{q=0,\pm 1} E^q(\mathbf{r})|g,\mu=q\rangle,$$

which is not destroyed by translational motion:

$$\frac{\hat{p}^2}{2M}|\Psi_{NC}\rangle = \hbar\omega_r|\Psi_{NC}\rangle,$$

where R is the normalization constant. Indeed, the amplitude $|\mathbf{E}(\mathbf{r})|$ is an eigenfunction of the Hamiltonian (4) at $\mathbf{B}=0$, is real, and has no nodes in the configuration depicted in Fig. 1.

For an example we take the transition $1\to 1$ involving the D_1-line of ^{87}Rb ($\lambda = 7947.6\times 10^{-8}$ cm). If we take a magnetic field $B\approx 4$ G ($\Omega\approx\gamma/2$), we readily see that for temperatures $T\leqslant 10^{-5}$ K and concentrations $n\geqslant 1.2\times 10^{13}$ cm^{-3} the quasicondensation condition (10) and the

condition (12) are met, i.e., when Bose condensation in possible in the entire volume of the gas as a result of adiabatic switch-off of the magnetic field (for a gas of free particles with $n=1.2\times10^{13}$ cm^{-3} we have $T_c=9.7\times10^{-8}$ K). These estimates also demonstrate the high efficiency (several orders of magnitude) of the proposed method of adiabatic cooling of a gas using dark magneto-optical lattices. Note that for condition (2) to be met the light intensity must be greater than or of order 1 W cm^{-2}.

Such values of the concentrations have yet to be achieved in experiments in laser cooling and trapping of atoms. But the rapid progress in experimental techniques gives hope that in the near future the regime with $nu>1$ in dark optical and magneto-optical lattices will be achieved.

The case $nu<1$ requires special consideration.

This work was partially supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17794) and the Universities of Russia Program.

*)E-mail: llf@admin.nsu.ru

[1] M. H. Anderson, J. R. Ensher, M. R. Mathews *et al.*, Science **269**, 198 (1995).
[2] C. C. Bradley, C. A. Sackett, J. J. Tollett, and R. G. Hulet, Phys. Rev. Lett. **75**, 1687 (1995).
[3] K. B. Davis, M.-O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, Phys. Rev. Lett. **75**, 3969 (1995).
[4] A. M. Tumaĭkin and V. I. Yudin, Zh. Éksp. Teor. Fiz. **98**, 81 (1990) [Sov. Phys. JETP **71**, 43 (1990)].
[5] G. Grynberg and J.-Y. Courtois, Europhys. Lett. **27**, 41 (1994).
[6] A. Hemmerich, W. Weidemüller, T. Esslinger, C. Zimmermann, and T. Hänsch, Phys. Rev. Lett. **75**, 37 (1995).
[7] N. P. Konopleva, A. V. Taichenachev, A. M. Tumaikin, and V. I. Yudin, Quantum Semiclassic. Opt. **8**, 837 (1996).
[8] C. Herzog and M. Olshanii, E-preprint atom-ph/9609002 in xxx.lanl.gov.
[9] V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
[10] A. V. Taĭchenachev, A. M. Tumaĭkin, M. A. Ol'shanyĭ, and V. I. Yudin, JETP Lett. **53**, 351 (1991).
[11] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3rd ed., Pergamon Press, Oxford (1980).

Translated by Eugene Yankovsky

Two-photon ultrashort nonlinear dynamics of coherent bosonic quasiparticles in condensed media

A. Kh. Rotaru and K. V. Shura*)

Moldavian State University, MD2009 Kishinev, Moldova
(Submitted 18 January 1997)
Zh. Éksp. Teor. Fiz. **113**, 2065–2080 (June 1998)

We present a theoretical study of the ultrashort nonlinear dynamics of bosonic quasiparticles in condensed media participating in two-photon quantum transitions. In the resonance approximation we obtain exact solutions in the form of elliptic functions. We show that allowing for antiresonant terms in the Hamiltonian of the interaction of the electromagnetic field and the quasiparticles leads to an entirely new phenomenon: the development of ultrashort chaos. The dynamical stochasticity of the quasiparticles is due to the disintegration of constants of motion of the system. Finally, we predict the effect of nonlinear tunneling from one potential well to another. © *1998 American Institute of Physics.* [S1063-7761(98)01206-2]

1. INTRODUCTION

Lately there has been an upsurge of interest in the study of nonlinear dynamics, especially of dynamical chaos in physical, chemical, biological, and other systems.

The discovery of dynamical chaos in nonlinear deterministic systems has been one of the most important and striking scientific achievements in recent times.[1] Numerous papers devoted to deterministic chaos have been published. The theory of the stochastic behavior of dissipative and Hamiltonian dynamical systems has been studied in monographs, textbooks, and reviews.[1–5]

Of special interest is the study of optical dynamical chaos, especially in condensed media in connection with predictions of new physical phenomena in such media and their use in new optoelectronic devices. Another potential field of research is the application of these phenomena in optical processing of information and in designing a new generation of optical computers.

Because of the large nonlinearity values in semiconductors at the long-wavelength fundamental absorption edge of the crystal, aspects of optical self-organization are undergoing intensive study, including dynamical chaos in a system of coherent (in the Bogolyubov sense) excitons and biexcitons.

A theory of dissipative dynamical chaos in a system of excitons and biexcitons participating in various quantum transitions was developed in Refs. 6–12. There it was demonstrated (among other things) that in the region of the M luminescence band of a semiconductor, the dynamical evolution of the quasiparticles is described by a system of Lorenz equations in four-dimensional phase space. It has been proved that the transition to the chaotic mode is achieved through period-doubling bifurcations accompanied by the formation of strange attractors in the four-dimensional phase space. The dynamical evolution of coherent excitons and photons was studied in Refs. 13 and 14 with allowance for exciton–exciton interaction. In this case the quasiparticle dynamics is described by a generalized system of Keldysh equations.[15] It is shown that in contrast to Lorenz chaos, where the stochastic oscillations and the formation of a strange attractor are related to jumps between the corresponding equilibrium states, in the given case stochasticity is related to the formation of a strange attractor in four-dimensional phase space, which is filled by unstable phase trajectories in a complicated manner.

As for the dynamical chaos in Hamiltonian systems of coherent quasiparticles in semiconductors and insulators, the study of these phenomena has only begun. The possibility in principle of ultrashort dynamical chaos emerging in a system of excitons and biexcitons near the M luminescence band of the crystal and in two-photon creation of biexcitons from the ground state of the crystal has been proved in Ref. 16–18. There it was shown that at certain critical values of the parameters, stochastic instability sets in in the system due to destruction of constants of motion.

The present paper investigates a new cooperative nonlinear phenomenon: ultrashort dynamical chaos in a system of coherent bosons in the optical transitions between the levels. By way of an example, we study the nonlinear dynamics in two-photon transitions between exciton or exciton–biexciton levels in the course of time intervals shorter than the characteristic relaxation times.

Because these transitions are characterized by gigantic oscillator strengths, the effects of coherent nonlinear interaction of light and matter in this frequency range manifest themselves most vividly.

Wang and Haken[19] and Moskalenko *et al.*[20,21] have developed theories of a two-photon laser, respectively, for the model of two-level atoms and in two-photon optical conversion of biexcitons into excitons.

Moskalenko *et al.*[20,21] used the methods of the quantum theory of fluctuations and decay to obtain a master equation for the density matrix of excitons, photons, and biexcitons in solids. Using Glauber's P-representation, they derived a Fokker–Plank equation for the system of coherent quasiparticles. They found the condition needed for two-photon laser

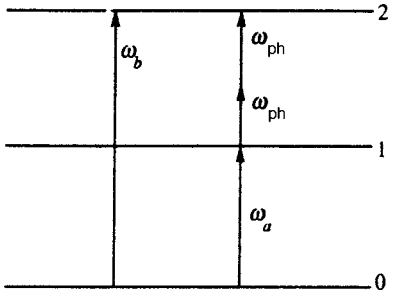


FIG. 1. Energy level diagram of the two-photon conversion of bosons from one excited state to another: 0 is the ground state of the crystal, 1 is the energy level of the first bosonic level, 2 is the energy level of the second bosonic level, ω_{ph} is the light frequency, ω_a is the frequency of the first bosonic transition, and ω_b is the frequency of the second bosonic transition.

generation of biexcitons in biexciton–exciton conversion. Finally, they showed that the transition from the disordered phase to the ordered one is equivalent to a first-order phase transition.

The general drawback of Refs. 19–21 is that they study only steady states of the particles and quasiparticles in two-photon quantum transitions. More interesting, however, is the study of nonlinear dynamics in the corresponding quantum transitions with and without allowance for dissipation processes.

2. HAMILTONIAN OF THE PROBLEM; DERIVATION OF THE DYNAMICAL EQUATIONS

Let us study the nonlinear ultrashort dynamics of coherent quasiparticles in condensed media in the case two-photon exciton–biexciton or exciton–exciton transitions.

The model we will be using is depicted in Fig. 1. The ground state of the quasiparticles is an exciton or the first exciton level, and the excited state is a biexciton or the second exciton level.

As is known, the characteristic relaxation times of excitons and biexcitons in semiconductors and insulators are of order 10^{-11}–10^{-10} s. However, by means of a newly developed method of generating and forming light pulses with a length down to 10^{-15} s, we can experimentally study coherent phenomena for times shorter than the relaxation time. In particular, to observe ultrashort phenomena in a system of excitons and biexcitons under two-photon excitation, pulses in the picosecond and subpicosecond ranges are needed.

Up till now the nonlinear dynamics of coherent excitons and biexcitons has been studied in the resonance approximation. Below we show that allowing for nonresonant (antiresonant) terms in the interaction Hamiltonian dramatically changes the nonlinear dynamics of the quasiparticles and under certain conditions results in a new cooperative phenomenon, ultrashort two-photon dynamical chaos in exciton–biexciton or exciton–exciton transitions.

We begin our study of two-photon nonlinear dynamics by introducing the Hamiltonian of coherent quasiparticles uniformly distributed over the crystal:

$$\frac{H}{\hbar}=\omega_a a_{\mathbf{p}}^{\dagger}a_{\mathbf{p}}+\omega_b b_{2\mathbf{k}+\mathbf{p}}^{\dagger}b_{2\mathbf{k}+\mathbf{p}}+\omega_{\mathrm{ph}}c_{\mathbf{k}}^{\dagger}c_{\mathbf{k}}+iq(a_{\mathbf{p}}^{\dagger}+a_{-\mathbf{p}})\times(c_{\mathbf{k}}^{\dagger}+c_{-\mathbf{k}})(c_{\mathbf{k}}^{\dagger}+c_{-\mathbf{k}})b_{2\mathbf{k}+\mathbf{p}}-iq(a_{\mathbf{p}}+a_{-\mathbf{p}}^{\dagger})(c_{\mathbf{k}}+c_{-\mathbf{k}}^{\dagger})(c_{\mathbf{k}}+c_{-\mathbf{k}}^{\dagger})b_{2\mathbf{k}+\mathbf{p}}^{\dagger}. \quad (1)$$

Here $a_{\mathbf{p}}^{\dagger}$ ($a_{\mathbf{p}}$) and $b_{\mathbf{p}}^{\dagger}$ ($b_{\mathbf{p}}$) are the creation (annihilation) operators for quasiparticles with a wave vector $\mathbf{p}$ on the first and second levels with generation energies $\hbar\omega_a$ and $\hbar\omega_b$, respectively; $c_{\mathbf{k}}^{\dagger}$ ($c_{\mathbf{k}}$) is the creation (annihilation) operator for a photon with a wave vector $\mathbf{k}$ and energy $\hbar\omega_{\mathrm{ph}}$; and q is the constant of two-photon conversion between the levels. From now on we set $\hbar=1$ and go over to amplitude–phase variables:

$$a_{\mathbf{p}}=\sqrt{n}\,\exp(-i\varphi_a+ipx),\quad a_{\mathbf{p}}^{\dagger}=\sqrt{n}\,\exp(i\varphi_a-ipx),$$
$$c_{\mathbf{k}}=\sqrt{f}\,\exp(-i\varphi_e+ikx),\quad c_{\mathbf{k}}^{\dagger}=\sqrt{f}\,\exp(i\varphi_e-ikx),$$
$$b_{2\mathbf{k}+\mathbf{p}}=\sqrt{N}\,\exp(-i\varphi_b+i(2k+p)x),$$
$$b_{2\mathbf{k}+\mathbf{p}}^{\dagger}=\sqrt{N}\,\exp(i\varphi_b-i(2k+p)x), \quad (2)$$

where n, φ_a, N, and φ_b are, respectively, the number and phase of the quasiparticles on the first and second levels, and f and φ_e are, respectively, the number and phase of the photons.

In terms of the new variables the system Hamiltonian is

$$H=H_0+H_r+H_{ar},$$
$$H_0=\omega_a n+\omega_b N+\omega_{\mathrm{ph}}f,$$
$$H_r=2q\sqrt{nN}f\,\sin(\varphi_b-\varphi_a-2\varphi_e),$$
$$H_{ar}=2q\sqrt{nN}f[\sin(\varphi_b-\varphi_a+2\varphi_e)+\sin(\varphi_b-\varphi_a)+\sin(\varphi_b+\varphi_a-2\varphi_e)+\sin(\varphi_b+\varphi_a+2\varphi_e)+\sin(\varphi_b+\varphi_a)], \quad (3)$$

where H_0 is the free-particle Hamiltonian, and H_r and H_{ar} are the resonant and nonresonant (antiresonant) terms in the Hamiltonian responsible for particle interaction. Here and in what follows we discard the subscripts on the wave vectors.

If we allow for (3), we can easily derive a system of nonlinear differential equations describing the dynamical evolution of the coherent quasiparticles:

$$\frac{dn}{dt}=-\frac{\partial H}{\partial\varphi_a}=2q\sqrt{nN}f[\cos(\varphi_b-\varphi_a-2\varphi_e)+\cos(\varphi_b-\varphi_a+2\varphi_e)+2\cos(\varphi_b-\varphi_a)-\cos(\varphi_b+\varphi_a-2\varphi_e)-\cos(\varphi_b+\varphi_a+2\varphi_e)-2\cos(\varphi_b+\varphi_a)], \quad (4)$$

$$\frac{dN}{dt}=-\frac{\partial H}{\partial\varphi_b}=-2q\sqrt{nN}f[\cos(\varphi_b-\varphi_a-2\varphi_e)+\cos(\varphi_b-\varphi_a+2\varphi_e)+2\cos(\varphi_b-\varphi_a)+\cos(\varphi_b+\varphi_a-2\varphi_e)+\cos(\varphi_b+\varphi_a+2\varphi_e)+2\cos(\varphi_b+\varphi_a)], \quad (5)$$

$$\frac{df}{dt}=-\frac{\partial H}{\partial\varphi_e}=4q\sqrt{nN}f[\cos(\varphi_b-\varphi_a-2\varphi_e)-\cos(\varphi_b-\varphi_a+2\varphi_e)+\cos(\varphi_b+\varphi_a-2\varphi_e)-\cos(\varphi_b+\varphi_a+2\varphi_e)], \quad (6)$$

$$\frac{d\varphi_a}{dt}=\frac{\partial H}{\partial n}=\omega_a+qf\sqrt{\frac{N}{n}}\,[\sin(\varphi_b-\varphi_a-2\varphi_e)+\sin(\varphi_b-\varphi_a+2\varphi_e)+\sin(\varphi_b-\varphi_a)+\sin(\varphi_b+\varphi_a-2\varphi_e)+\sin(\varphi_b+\varphi_a+2\varphi_e)+\sin(\varphi_b+\varphi_a)], \tag{7}$$

$$\frac{d\varphi_b}{dt}=\frac{\partial H}{\partial N}=\omega_b+qf\sqrt{\frac{n}{N}}\,[\sin(\varphi_b-\varphi_a-2\varphi_e)+\sin(\varphi_b-\varphi_a+2\varphi_e)+\sin(\varphi_b-\varphi_a)+\sin(\varphi_b+\varphi_a-2\varphi_e)+\sin(\varphi_b+\varphi_a+2\varphi_e)+\sin(\varphi_b+\varphi_a)], \tag{8}$$

$$\frac{d\varphi_e}{dt}=\frac{\partial H}{\partial f}=\omega_{\rm ph}+2q\sqrt{nN}[\sin(\varphi_b-\varphi_a-2\varphi_e)+\sin(\varphi_b-\varphi_a+2\varphi_e)+\sin(\varphi_b-\varphi_a)+\sin(\varphi_b+\varphi_a-2\varphi_e)+\sin(\varphi_b+\varphi_a+2\varphi_e)+\sin(\varphi_b+\varphi_a)]. \tag{9}$$

To separate the resonance phase, we introduce the following notation:

$$\psi=\varphi_b-\varphi_a-2\varphi_e,\quad \varphi_1=\varphi_b-\varphi_a+2\varphi_e,$$
$$\varphi_2=\varphi_b-\varphi_a,\quad \varphi_3=\varphi_b+\varphi_a+2\varphi_e,$$
$$\varphi_4=\varphi_b+\varphi_a,\quad \varphi_5=\varphi_b+\varphi_a-2\varphi_e.$$

Then the system of equations (4)–(9) becomes

$$\frac{dn}{d\tau}=\frac{\lambda}{5}\sqrt{nN}f[\cos\psi+\varepsilon(\cos\varphi_1+2\cos\varphi_2-\cos\varphi_3-2\cos\varphi_4-\cos\varphi_5)], \tag{10}$$

$$\frac{df}{d\tau}=\frac{2\lambda}{5}\sqrt{nN}f[\cos\psi-\varepsilon(\cos\varphi_1+\cos\varphi_3-\cos\varphi_5)], \tag{11}$$

$$\frac{dN}{d\tau}=-\frac{\lambda}{5}\sqrt{nN}f[\cos\psi+\varepsilon(\cos\varphi_1+2\cos\varphi_2+\cos\varphi_3+2\cos\varphi_4+\cos\varphi_5)], \tag{12}$$

$$\frac{d\psi}{d\tau}=\Delta+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}-f\sqrt{\frac{N}{n}}-2\sqrt{nN}\right]B, \tag{13}$$

$$\frac{d\varphi_1}{d\tau}=\Delta+4+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}-f\sqrt{\frac{N}{n}}+2\sqrt{nN}\right]B, \tag{14}$$

$$\frac{d\varphi_2}{d\tau}=\Delta+2+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}-f\sqrt{\frac{N}{n}}\right]B, \tag{15}$$

$$\frac{d\varphi_3}{d\tau}=\Delta+4+2\,\frac{\omega_a}{\omega_{\rm ph}}+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}+f\sqrt{\frac{N}{n}}+2\sqrt{nN}\right]B, \tag{16}$$

$$\frac{d\varphi_4}{d\tau}=\Delta+2+2\,\frac{\omega_a}{\omega_{\rm ph}}+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}+f\sqrt{\frac{N}{n}}\right]B, \tag{17}$$

$$\frac{d\varphi_5}{d\tau}=\Delta+2\,\frac{\omega_a}{\omega_{\rm ph}}+\frac{\lambda}{10}\left[f\sqrt{\frac{n}{N}}+f\sqrt{\frac{N}{n}}-2\sqrt{nN}\right]B, \tag{18}$$

where $B=\sin\psi+\varepsilon(\sin\varphi_1+2\sin\varphi_2+\sin\varphi_3+2\sin\varphi_4+\sin\varphi_5)$, $\tau=tck$, $\lambda=10q/\omega_{\rm ph}$, $\Delta=(\omega_b-\omega_a-2\omega_{\rm ph})/\omega_{\rm ph}$, and the parameter ε was introduced formally and takes a value of 0 or 1. For $\varepsilon=1$, Eqs. (10)–(18) are equivalent to Eqs. (4)–(9). The resonance approximation amounts to ignoring the terms containing φ_i, $i=1$–5, which is equivalent to $\varepsilon=0$.

3. RESONANCE APPROXIMATION $\varepsilon=0$

In the resonance approximation the system has three constants of motion for the numbers of particles:

$$2N+f=C,\quad n+N=C_1,\quad 2n-f=C_2. \tag{19}$$

Bearing this in mind and using Eqs. (10)–(18), we can easily obtain a reduced system of equations that describes the dynamical evolution of the coherent quasiparticles in the resonance approximation:

$$\frac{d\psi}{d\tau}=\frac{\partial P}{\partial N}=\Delta+\frac{\lambda}{10}\left[\sqrt{\frac{C_1-N}{N}}\,(C-2N)-\sqrt{\frac{N}{C_1-N}}\,(C-2N)-2\sqrt{N(C_1-N)}\right]\sin\psi, \tag{20}$$

$$\frac{dN}{d\tau}=-\frac{\partial P}{\partial\psi}=-\frac{\lambda}{5}(C-2N)\sqrt{N(C_1-N)}\,\cos\psi, \tag{21}$$

$$P=\frac{H-\omega_aC_1-\omega_{\rm ph}C}{\omega_{\rm ph}}=\Delta N+\frac{\lambda}{5}(C-2N)\sqrt{N(C_1-N)}\sin\psi, \tag{22}$$

where P an additional constant of motion acting as a Hamiltonian in the space of the variables N and ψ.

The evolution of the system varies depending on the relationship between the constants of motion and the detuning from resonance.

If we introduce the notation $\hat\lambda=\lambda C_1$, $\hat n=n/C_1$, $\hat N=N/C_1$, $\hat f=f/C_1$, and $\hat C=C/C_1$, go over to the variables N and $dN/d\tau$, substitute (22) in (21), and drop the ''hat,'' we arrive at the following equation for the unperturbed case:

$$\left(\frac{dN}{d\tau}\right)^2=\frac{\lambda^2}{25}N(1-N)(C-2N)^2-(P-\Delta N)^2. \tag{23}$$

The solution of this equation is

$$N=\frac{N_4N_{13}+N_1N_{34}\,\mathrm{sn}^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+F(\varphi_0);\kappa)}{N_{13}+N_{34}\,\mathrm{sn}^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+F(\varphi_0);\kappa)}. \tag{24}$$

Here $\varphi_0=\sin^{-1}\sqrt{N_{13}N_{04}/N_{34}N_{10}}$ is the argument of the elliptic integral of the first kind (F), $\kappa=\sqrt{N_{12}N_{34}/N_{13}N_{24}}$ is the modulus of the elliptic function, N_0 is the initial number of particles on the second level, and $N_{ij}=N_i-N_j$, where $N_4<N_3<N_2<N_1$ are the roots of the equation

$$\frac{\lambda^2}{25}N(1-N)(C-2N)^2-(P-\Delta N)^2=0. \tag{25}$$

The nonlinear oscillation frequency is

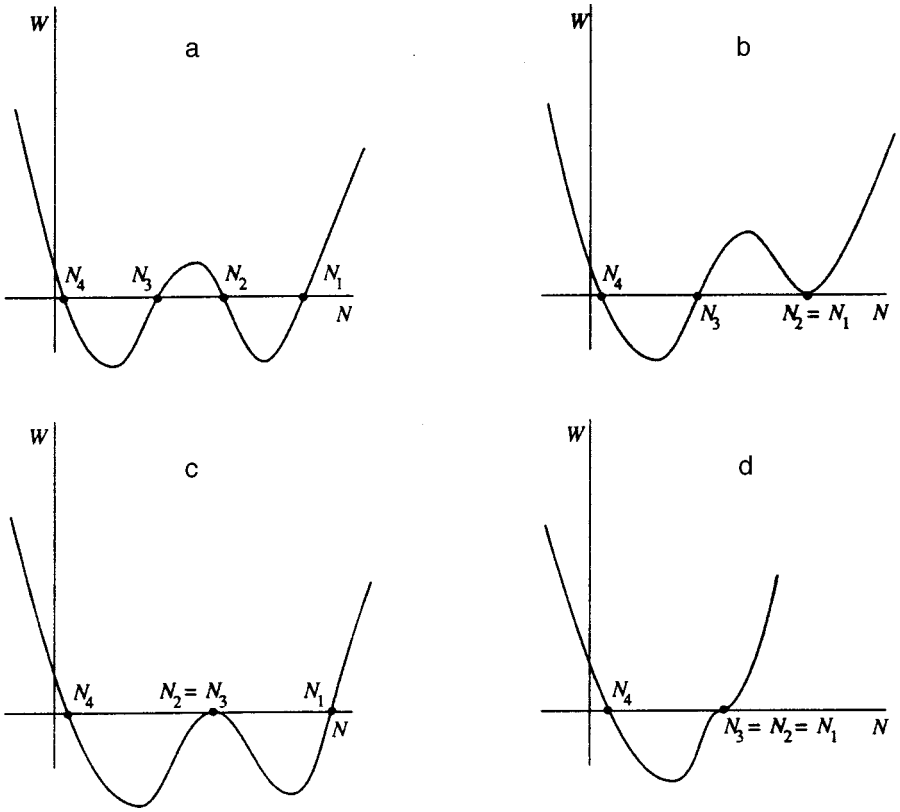


FIG. 2. Potential energy $W(N)$ of a nonlinear oscillator: in the general case (a), for $P=C\Delta/2$ and $C>2$ (b), for $P=C\Delta/2$ and $C<2$ (c), and for $P=\Delta=0$ and $C=2$ (d). The horizontal line corresponds to the total energy of the oscillator.

$$\omega(P)=\frac{\pi q\sqrt{N_{13}N_{24}}}{K(\kappa)}, \tag{26}$$

where $K(\kappa)$ is the complete elliptic integral of the first kind.

Thus, the number of quasiparticles is a periodic function of time with a frequency that depends on the values of the constants of motion and the detuning from resonance.

For $P/\Delta=C/2$ and $C>2$ we arrive at a particular solution of the form

$$N=\frac{N_4N_{13}+N_1N_{34}\sin^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+\varphi_0)}{N_{13}+N_{34}\sin^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+\varphi_0)}, \tag{27}$$

Here $N_4=25\Delta^2/4\lambda^2$, $N_{2,3}=C/2$, and $N_1=1$.

Note that the various solutions of Eq. (23) are determined by the shape of the potential curve (Fig. 2)

$$W(N)=-\frac{\lambda^2}{25}N(1-N)(C-2N)^2-(P-\Delta N)^2. \tag{30}$$

where $N_4=25\Delta^2/4\lambda^2$, $N_3=1$, and $N_{1,2}=C/2$. For $P=\Delta=0$, $N_1=N_2=N_3=1$, and $N_4=0$, we have

$$N=\frac{\lambda^2\tau^2/25}{1+\lambda^2\tau^2/25}. \tag{28}$$

For $P/\Delta=C/2$ and $C<2$ the phase trajectory of the systems is the separatrix. As is well known,[2,3] any small perturbation causes a dramatic disintegration of the trajectories near a separatrix and under certain conditions the quasiparticle motion becomes stochastic.

The solution on the separatrix is

$$N=\frac{N_4N_{13}+N_1N_{34}\tanh^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+\log(\tan(\varphi_0/2)+\pi/4))}{N_{13}+N_{34}\tanh^2((\lambda\tau/5)\sqrt{N_{24}N_{13}}+\log(\tan(\varphi_0/2)+\pi/4))}. \tag{29}$$

4. DISINTEGRATION OF CONSTANTS OF MOTION AND STOCHASTIZATION OF PHASE TRAJECTORIES

In the unperturbed case, the trajectories of motion are closed curves $N(\psi)$ determined by the parametric equation (22). If we allow for antiresonant terms $\varepsilon=1$, for small values of λ the constants of motion change and then, as λ increases, disintegrate.

The variation of the constant of motion P with time is given by the following expression:

$$\frac{dP(N,\psi,C,C_1)}{d\tau}=\frac{\partial P}{\partial N}\frac{dN}{d\tau}+\frac{\partial P}{\partial \psi}\frac{d\psi}{d\tau}+\frac{\partial P}{\partial C}\frac{dC}{d\tau}+\frac{\partial P}{\partial C_1}\frac{dC_1}{d\tau}. \tag{31}$$

Allowing for (12), (13), (20), and (21), for the perturbed equations we have

$$\frac{dN}{d\tau}=-\frac{\partial P}{\partial \psi}+\delta\left(\frac{dN}{d\tau}\right),\quad \frac{d\psi}{d\tau}=\frac{\partial P}{\partial N}+\delta\left(\frac{d\psi}{d\tau}\right), \tag{32}$$

where

$$\delta\left(\frac{dN}{d\tau}\right)=-\frac{\lambda}{5}(C-2N)\sqrt{N(C_1-N)}(\cos\varphi_1+2\cos\varphi_2+\cos\varphi_3+2\cos\varphi_4+\cos\varphi_5), \tag{33}$$

$$\delta\left(\frac{d\psi}{d\tau}\right)=\frac{\lambda}{10}\left[\sqrt{\frac{C_1-N}{N}}(C-2N)-\sqrt{\frac{N}{C_1-N}}(C-2N)-2\sqrt{N(C_1-N)}\right](\sin\varphi_1+2\sin\varphi_2+\sin\varphi_3+2\sin\varphi_4+\sin\varphi_5). \tag{34}$$

Substituting (32)–(34) into (31) yields

$$\frac{dP}{d\tau}=\left\{\frac{\lambda^2}{50}[(C-2N)^2(1-2N)-2(C-2N)(1-N)N]\times(\sin(\varphi_1-\psi)+2\sin(\varphi_2-\psi)+\sin(\varphi_3-\psi)+2\sin(\varphi_4-\psi)+\sin(\varphi_5-\psi))+\frac{\Delta\lambda}{5}(C-2N)\times\sqrt{N(C_1-N)}(\cos\varphi_1+2\cos\varphi_2+\cos\varphi_3+2\cos\varphi_4+\cos\varphi_5)\right\}\Big/\left[1+\frac{P-\Delta N}{C-2N}\right], \tag{35}$$

where for N, ψ, and φ_i we took the unperturbed expressions.

Near the separatrix $P/\Delta\to C/2$, $C<2$, we obtain the following expressions for the characteristic roots:

$$N_1=1-\frac{(P-\Delta)^2}{(2\lambda/5)^2(C/2-1)^2-2(P-\Delta)}, \tag{36}$$

$$N_{2,3}=\frac{C}{2}-\frac{(P-\Delta C/2)(\Delta\pm(\lambda/5)\sqrt{C(2-C)})}{(\lambda^2/25)C(2-C)-\Delta^2}, \tag{37}$$

$$N_4=\frac{P^2}{(\lambda C/5)^2+2\Delta P}. \tag{38}$$

The quantity $N(\tau)$ varies almost from 0 to $C/2$, where at a turning point of the hyperbolic type its period tends to infinity. For $\lambda\ll 1$ the dominator in (35) is of order unity, $(C-2N)^2(1-2N)-2(C-2N)(1-N)N\sim C^2$, and $(C-2N)\sqrt{N(C_1-N)}\sim(P-\Delta C/2)/\lambda$ holds far from the singular point $N=C/2$ within a small temporal interval of the small-oscillation period $2\pi/\omega_0$. In the neighborhood of the singular point, $(C-2N)^2(1-2N)-2(C-2N)(1-N)N\sim(P-\Delta C/2)/\lambda$ and $(C-2N)\sqrt{N(C_1-N)}\sim(P-\Delta C/2)/\lambda$ within a large temporal interval $2\pi/\omega(P)$, where $\omega(P)$ is the nonlinear oscillation frequency. In this case, instead of (35) we have

$$\frac{dP}{d\tau}=\frac{\lambda^2}{50}C^2A(\tau)(\sin\theta_1+2\sin\theta_2+\sin\theta_3+2\sin\theta_4+\sin\theta_5), \tag{39}$$

$$\frac{d\theta_1}{d\tau}=\Delta+4+O(\lambda), \tag{40}$$

$$\frac{d\theta_2}{d\tau}=\Delta+2+O(\lambda), \tag{41}$$

$$\frac{d\theta_3}{d\tau}=\Delta+4+2\frac{\omega_a}{\omega_{\rm ph}}+O(\lambda), \tag{42}$$

$$\frac{d\theta_4}{d\tau}=\Delta+2+2\frac{\omega_a}{\omega_{\rm ph}}+O(\lambda), \tag{43}$$

$$\frac{d\theta_5}{d\tau}=\Delta+2\frac{\omega_a}{\omega_{\rm ph}}+O(\lambda), \tag{44}$$

where $A(\tau)$ is a periodic function with a period $2\pi/\omega(P)$, a height of order unity, and a width of $2\pi/\omega_0$; $\omega_0=\lambda\sqrt{2+C}/5$ is the frequency of small oscillations of the system for $P\to\Delta C/2$ and $C<2$.

If in (39)–(44) we go over to a system of discrete transformations, we get

$$P_{m+1}=P_m+\bar{\Delta}P, \tag{45}$$

$$\theta_{i,m+1}=\theta_{i,m}+\frac{4\pi}{\omega(P_{m+1})}=\theta_{i,m}+\frac{4\pi}{\omega(P_m)}-\frac{4\pi}{\omega^2(P_m)}\frac{d\omega(P_m)}{dP_m},\quad i=1\text{–}5, \tag{46}$$

$$\bar{\Delta}P\approx\frac{\lambda^2}{50}C^2\int d\tau A(\tau)(\sin\theta_1+2\sin\theta_2+\sin\theta_3+2\sin\theta_4+\sin\theta_5). \tag{47}$$

The nature of the solution (45)–(47) is determined by the value of M (Refs. 2 and 3):

$$M=\frac{4\pi}{\omega^2(P)}\left|\frac{d\omega(P)}{dP}\right|. \tag{48}$$

When $M\ll 1$, the system of coherent particles performs quasi-periodic oscillations. When $M\gg 1$, the motion becomes stochastic during the phase decorrelation time

$$R(\tau)=\left\langle\exp\left(\sum_{i=1,5}[\theta_i(\tau_1)-\theta_i(\tau_1+\tau)]\right)\right\rangle\sim\exp\left(-\frac{5\tau}{\tau_c}\right), \tag{49}$$

where $\tau_c=1/\omega(P)\ln M$.

The change in P in one transformation step is

$$\max\bar{\Delta}P\sim\frac{28}{10}C^2\lambda\pi\exp\left(-\frac{5\alpha\pi}{2\lambda}\right), \tag{50}$$

where the constant α is of order unity. The nonlinear oscillation frequency $\omega(P)$ near the separatrix is

$$\omega(P)=\frac{\pi\lambda\sqrt{N_{13}N_{24}}}{5\ln\dfrac{5\sqrt{C(2-C)}((\lambda^2/25)C(2-C)-\Delta^2)}{\lambda(P-\Delta C/2)}}. \quad (51)$$

Plugging (50) and (51) into (48), we get

$$M=\frac{56C^2\pi\ \exp(-5\alpha\pi/2\lambda)}{(P-\Delta C/2)\sqrt{N_{13}N_{24}}}. \quad (52)$$

The boundary of the stochastic layer is determined from the condition $M(P_0,H)\sim 1$, or

$$P_0=\frac{\Delta C}{2}+\frac{56C^2\pi\ \exp(-5\alpha\pi/2\lambda)}{\sqrt{N_{13}N_{24}}}, \quad (53)$$

and the nutation decay constant is given by the following expression:

$$\gamma_c=\frac{\pi\lambda\sqrt{N_{13}N_{24}}\ \ln M}{5\ln\dfrac{5\sqrt{C(2-C)}((\lambda^2/25)C(2-C)-\Delta^2)}{\lambda(P-\Delta C/2)}}. \quad (54)$$

5. COMPUTER EXPERIMENT

The numerical analysis was done, without loss of generality, at $\Delta=0$. Generally, the system of equations (4)–(9) has one constant of motion, the system energy H, and the region of motion is a five-dimensional hypersurface defined by (3) in a six-dimensional phase space.

When the antiresonant terms are ignored, four additional constants of motion emerge, C, C_1, C_2, and P, of which two are linearly independent. The quantities H, C, and P determine a three-parameter family of trajectories that are either closed curves on the surface or asymptotic curves. The temporal evolution of coherent quasiparticles in this case are nonlinear periodic oscillations with a finite or infinite period. The period of nutation oscillations decreases with increasing λ. As the separatrix is approached ($C<2$ and $P\to 0$), the angles of the phase trajectories become sharper and assume a rectangular shape. The number of quasiparticles varies from minimum to maximum (or from maximum to minimum) as the phase remains constant, $\psi=n\pi$ ($n=0,\pm1,\pm2,\ldots$, depending on the initial conditions). When the quasiparticle number reaches its maximum or minimum, the phase undergoes a jump: $\psi_{t+0}=\psi_{t-0}\pm\pi$.

Allowing for the antiresonant terms in the Hamiltonian leads to changes in the shape of the potential energy, so that all existing aperiodic oscillation regimes cease to exist. In this case, for small values of λ and any initial conditions the system performs quasiperiodic oscillations (Fig. 3). In the phase space this motion is depicted by a trajectory that winds itself around a toroidal tube. Here the fundamental harmonic is modulated by subharmonics, whose amplitudes increase with λ, making the motion more complicated.

As λ increases, the motion of the system becomes more complicated and the quasiperiodic oscillations begin to change into chaotic oscillations. The constants of motion completely disintegrate in the process. Figure 4 depicts the chaotic dynamics of quasiparticle and phonon numbers n and f and the corresponding phase portrait. Note that the quasi-

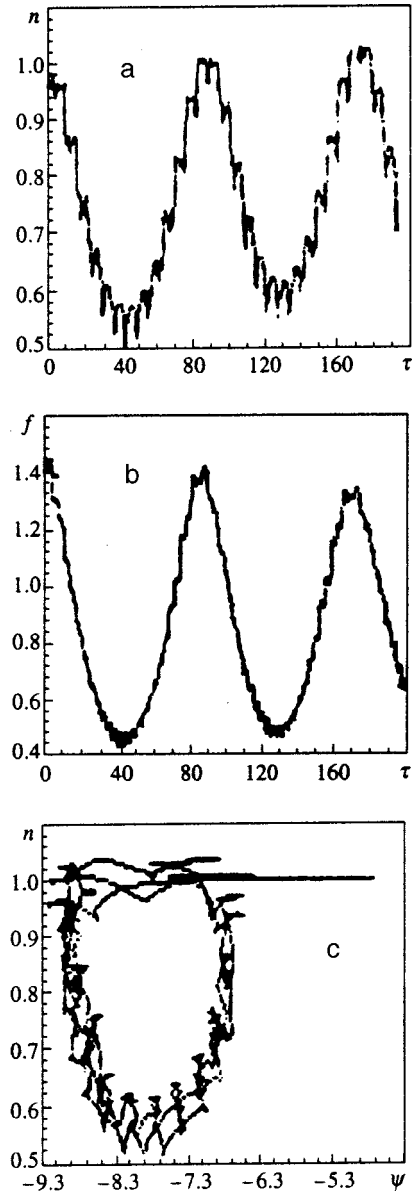


FIG. 3. Temporal evolution of the number of (a) coherent bosons on the first excited level and (b) photons, and (c) the projection of the phase trajectory on the (n,ψ) plane at $n_0=1$, $N_0=10^{-8}$, $f_0=1.4$, $\varphi_{a0}=\varphi_{e0}=1.571$, $\varphi_{b0}=10^{-8}$, and $\lambda=0.5$.

particle numbers change within the potential wells inside which they begin their motion. An interesting phenomenon was detected in computer experiments for $C\sim 1.3$–1.7. At such values, tunneling of the nonlinear motion from one well to another was observed. The tunneling was accompanied by a spread of the trajectories over the phase space (Fig. 5). Figure 5 clearly shows that there is a rapid change in the lower limit of the number of quasiparticles on the first level (n) and the upper limit of the number of photons (f) in comparison to Fig. 4, where the motion is only in one potential well. The phase trajectory in this case spreads out in the phase space and occupies a greater phase volume.

The calculations have shown that the evolution of the quasiparticles may be localized to a single potential well with λ constant by changing the detuning from resonance, Δ, i.e., by moving away from the condition $P=\Delta C/2$.

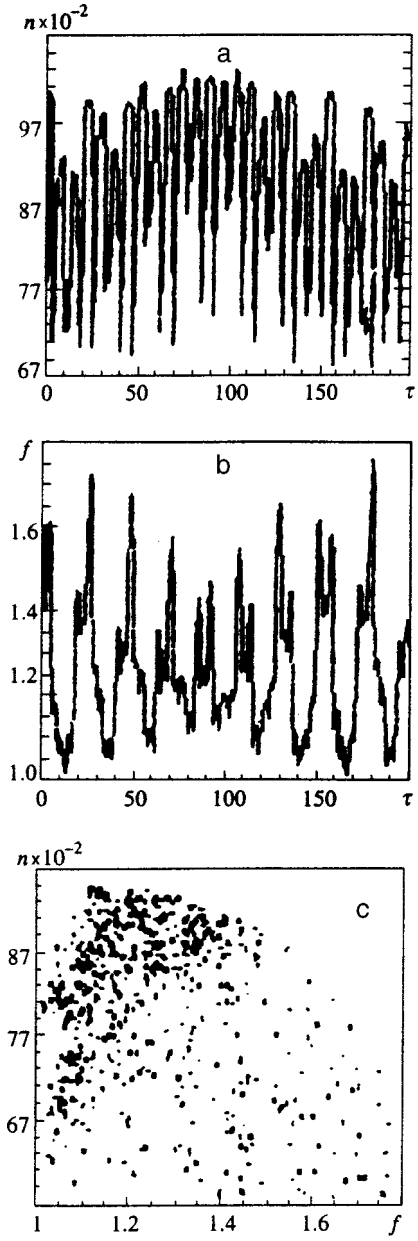


FIG. 4. Temporal evolution of the number of (a) coherent bosons on the first excited level and (b) photons, and (c) the projection of the phase trajectory on the (n,f) plane at $n_0=1$, $N_0=10^{-8}$, $f_0=1.4$, $\varphi_{a0}=\varphi_{e0}=-1.571$, $\varphi_{b0}=10^{-8}$, and $\lambda=1.025$.

This model has one distinctive feature. The condition $C<2$ determines not one separatrix, as it does in Refs. 17 and 18, but a family of separatrices. Each separatrix in this family corresponds to a set of initial quasiparticle concentrations. By changing the quasiparticle concentration we can move into the neighborhood of another separatrix. Near different separatrices chaos emerges at different values of λ. Chaos sets in most rapidly when C is large. As C decreases, chaos emerges at large values of λ.

Depending on the specific physical situation, we may find that only one nonresonant term in (6) is finite.

Figure 6 depicts the dynamical evolution in the case of a finite nonresonant term containing $\sin\varphi_1$. We see that as the resonance phase reaches the value $\psi=-\pi/2$, which in the unperturbed case corresponds to a hyperbolic turning point, the system goes into the stochastic-motion mode, which confirms that points of the hyperbolic type play a special role in the formation of a stochastic instability. We also note that although the evolution of the numbers n and ψ is stochastic, the constant of motion $C=n+N$ is preserved, i.e., the nonresonant term $\sin\varphi_1$ does not destroy it.

Figure 7 depicts the development of a local instability at various values of λ. The distance between two initially close trajectories is given by the following expression:

$$D=\left[(n_1-n_2)^2+(N_1-N_2)^2+(f_1-f_2)^2+\left\{\frac{\varphi_{a1}-\varphi_{a2}}{2\pi}\right\}^2+\left\{\frac{\varphi_{b1}-\varphi_{b2}}{2\pi}\right\}^2+\left\{\frac{\varphi_{e1}-\varphi_{e2}}{2\pi}\right\}^2\right]^{1/2}. \quad (55)$$

The decay constant is given by the formula

$$D=D_0\exp(\gamma_c\tau). \quad (56)$$

When λ is small, the distance between initially close trajectories remains small with the passage of time. As λ increases and reaches a certain value λ_{cr}, the value of D becomes of order unity. For $\lambda>\lambda_{cr}$ the decay constant γ_c rapidly increases: $\gamma_c=0.008$ and 0.04 at $\lambda=1.025$ and 1.05, and $C=1.4$. Each separatrix has its own value of λ_{cr}. Calculations have shown that $C\lambda_{C,cr}\approx$const.

The predicted dynamical chaos is an example of such chaos in Hamiltonian systems.

Generally, the dynamical equations for a system with n degrees of freedom are not completely integrable.

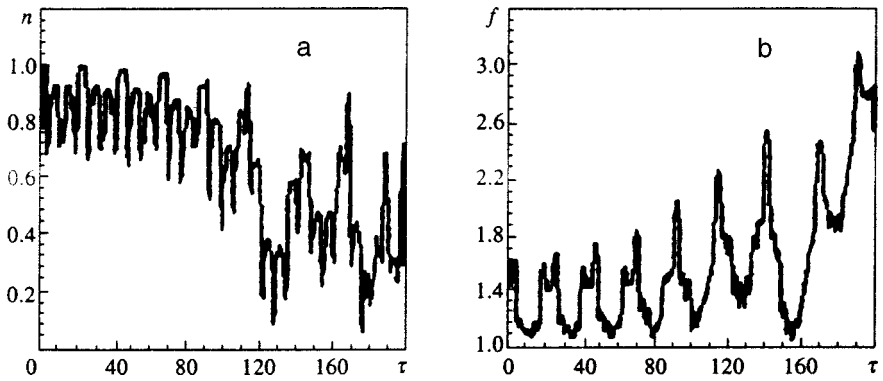


FIG. 5. Temporal evolution of the number of (a) coherent bosons on the first excited level and (b) photons at $n_0=1$, $N_0=10^{-8}$, $f_0=1.4$, $\varphi_{a0}=\varphi_{e0}=-1.571$, $\varphi_{b0}=10^{-8}$, and $\lambda=1.05$.

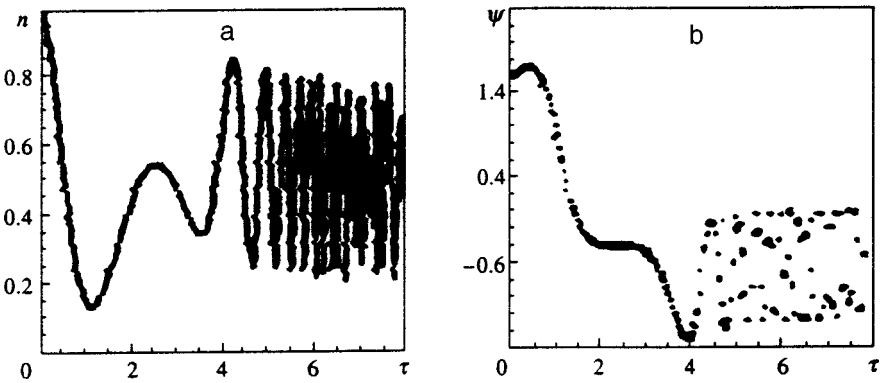


FIG. 6. Temporal evolution of the number of (a) coherent bosons on the first excited level and of (b) the coherent phase at $n_0=1$, $N_0=10^{-8}$, $f_0=2$, $\psi_0=\varphi_{01}=1.571$, and $\lambda=7.5$.

In our case the system of equations describing the dynamical evolution of coherent bosons consists of an integrable part (at $\varepsilon=0$) and a nonintegrable part for which nonresonant terms ($\varepsilon=1$) are responsible.

For $\varepsilon=0$ Eqs. (10)–(18) are completely integrable in terms of periodic trigonometric and elliptic functions and aperiodic hyperbolic functions on the separatrix.

The Kolmogorov–Arnold–Moser theorem[22–25] answers the question as to what extent the nature of integrable systems changes when perturbations are taken into account ($\varepsilon=1$). In this case the energy hypersurface is partitioned into two finite-volume regions. One region contains tori deformed by the antiresonant terms in the Hamiltonian. In the other (smaller) region, whose volume is zero at $\varepsilon=0$, the motion is extremely complicated. The tori that exist in this region at $\varepsilon=0$ are destroyed and the motion becomes chaotic.

As is well known, when the number of degrees of freedom $n>2$ (in our case $n=3$), the n-dimensional undestroyed tori fail to partition the $2n-1$-dimensional energy surface into nonintersecting parts and the regions of destroyed resonant tori merge, forming a unique complex net, which is known as an Arnold web. Here the phase points may wander far from their unperturbed positions, i.e., a phenomenon known as Arnold diffusion[2,26,27] is observed. A characteristic feature of such diffusion is that globally the system is unstable.

The emerging chaotic instability is not related to the action of random forces; rather, it is due to an intrinsic property of the system, the local instability.

In the model we have studied the expression for the perturbation is exact, so that the chaos is not related to poor convergence.

The computer experiment was performed with varying accuracy. We found that the chaotic solution is stable with respect to the choice of the calculation step, i.e., we are not dealing with chaos produced by the computer experiment itself. We checked our program by reproducing the results of Zaslavskiĭ and Sagdeev[3] concerning dynamical chaos and applying it to the system of Lorentz equations (see p. 88 of Ref. 10).

Thus, we have proved that in principle ultrashort dynamical chaos may occur in two-photon quantum transitions between the energy levels of bosonic quasiparticles in condensed media. We have shown that allowing for antiresonant

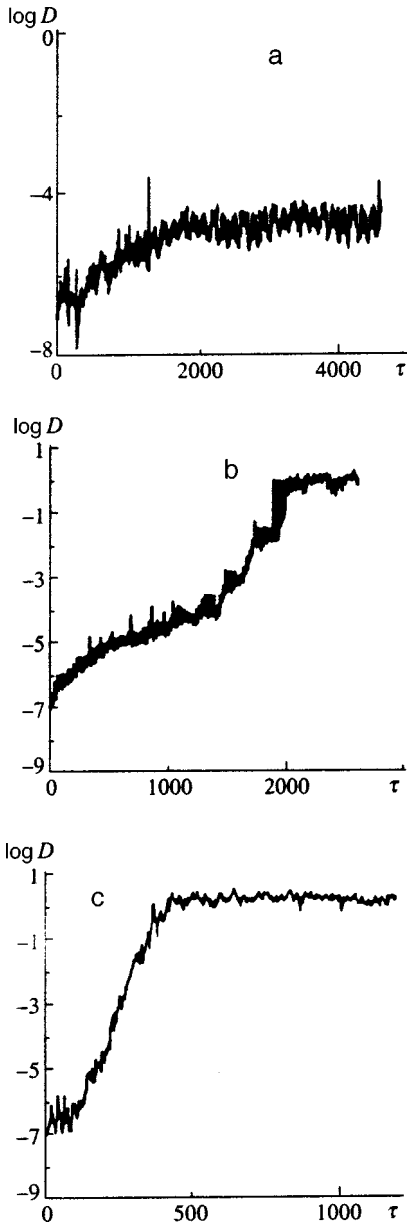


FIG. 7. Temporal dependence between initially close trajectories at $n_0=1$, $N_0=10^{-8}$, $f_0=1.4$, $\varphi_{a0}=\varphi_{e0}=1.571$, $\varphi_{b0}=10^{-8}$, and $\lambda=1$ (a), $\lambda=1.25$ (b), and $\lambda=1.05$ (c).

terms in the Hamiltonian responsible for the interaction between the electromagnetic field and the system of quasiparticles at certain values of the parameters, makes the system chaotic. The value $\lambda \sim 1$ is critical (at $C = 1.4$) and leads to a large stochastization region. We predict the presence of tunneling of nonlinear motion from one potential well to another.

In conclusion we note that two-photon ultrashort dynamical chaos may also occur in the model of two-level atoms.

*)E-mail: shura@usm.md

[1] Yu. N. Neĭmark and P. S. Landa, *Stochastic and Chaotic Oscillations*, Kluwer Academic; Dordrecht, Boston (1992) [Russ. orig., Nauka, Moscow (1987)].
[2] G. M. Zaslavsky, *Chaos in Dynamical Systems*, Harwood Academic, New York (1985).
[3] G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics*, [in Russian], Nauka, Moscow (1988).
[4] M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Theory of Oscillations and Waves* [in Russian], Nauka, Moscow (1984).
[5] *Strange Attractors* [in Russian], A. N. Kolmogorov (ed.), Mir Publishers, Moscow (1981) [a collection of translations into Russian of the basic articles on strange attractors].
[6] A. H. Rotaru and G. D. Shibarshina, Phys. Lett. A **109**, 292 (1985).
[7] A. Kh. Rotaru, Fiz. Tverd. Tela (Leningrad) **28**, 2492 (1986) [Sov. Phys. Solid State **28**, 1393 (1986)].
[8] S. A. Moscalenco, A. H. Rotaru, and V. A. Zaloj, Phys. Status Solidi B **150**, 454 (1988).
[9] A. Kh. Rotaru and V. A. Zalozh, Fiz. Tverd. Tela (Leningrad) **31**, 234 (1989) [Sov. Phys. Solid State **31**, 1231 (1989)].
[10] A. Kh. Rotaru and V. A. Zalozh, *Optical Self-organization of Excitons and Biexcitons in Semiconductors* [in Russian], Shtiintsa, Kishinev (1990).
[11] V. A. Zalozh, A. Kh. Rotaru, and V. Z. Tronchu, Zh. Éksp. Teor. Fiz. **103**, 994 (1993) [JETP **76**, 487 (1993)].
[12] V. A. Zalozh, A. Kh. Rotaru, and V. Z. Tronchu, Zh. Éksp. Teor. Fiz. **105**, 260 (1994) [JETP **78**, 138 (1994)].
[13] V. A. Zalozh, S. A. Moskalenko, and A. Kh. Rotaru, Zh. Éksp. Teor. Fiz. **95**, 601 (1989) [Sov. Phys. JETP **68**, 338 (1989)].
[14] B. Sh. Parkanskiĭ and A. Kh. Rotaru, Zh. Éksp. Teor. Fiz. **99**, 899 (1991) [Sov. Phys. JETP **72**, 499 (1991)].
[15] L. V. Keldysh, in *Current Problems of Physics* [in Russian], Nauka, Moscow (1992), p. 433.
[16] A. Kh. Rotaru and K. V. Shura, Fiz. Tverd. Tela (Leningrad) **33**, 1973 (1991) [Sov. Phys. Solid State **33**, 1111 (1991)].
[17] A. Kh. Rotaru and S. V. Shura, Zh. Éksp. Teor. Fiz. **104**, 2374 (1993) [JETP **77**, 60 (1993)].
[18] A. Kh. Rotaru and S. V. Shura, Zh. Éksp. Teor. Fiz. **107**, 450 (1995) [JETP **80**, 240 (1995)].
[19] Z. C. Wang and H. Z. Haken, Z. Phys. B: Cond. Matter **55**, 364 (1984).
[20] S. A. Moskalenko, A. Kh. Rotaru, and Yu. M. Shvera, Fiz. Tverd. Tela (Leningrad) **29**, 2396 (1987) [Sov. Phys. Solid State **29**, 1379 (1987)].
[21] S. A. Moskalenko, A. Kh. Rotaru, and Yu. M. Shvera, Teor. Mat. Fiz. **75**, 295 (1987).
[22] V. I. Arnold, Usp. Mat. Nauk **18**, 13 (1963).
[23] V. I. Arnold, Usp. Mat. Nauk **18**, 91 (1963).
[24] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **98**, 527 (1954).
[25] J. K. Moser, Nachr. Akad. Wiss. Göttingen, math.-phys. Kl. 1 (1962).
[26] V. I. Arnold, Dokl. Akad. Nauk SSSR **156**, 9 (1964) [*sic*].
[27] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion*, Springer, Berlin (1982).

Translated by Eugene Yankovsky

The method of moments and the nonequilibrium thermodynamics of rarefied gases

V. M. Zhdanov and V. I. Roldugin*)

Institute of Physical Chemistry, Russian Academy of Sciences, 117915 Moscow, Russia
(Submitted 3 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 2081–2095 (June 1998)

A generalized variant of the nonequilibrium thermodynamics of rarefied gases based on the linearized equations of Grad's method of moments is studied. It is shown that despite the more complicated form of the thermodynamic forces, which include spatial derivatives of the fluxes, entropy production remains a bilinear combination of generalized thermodynamic fluxes and forces. Using perturbation theory in the small Knudsen number, the expressions obtained can be reduced to the well-known results of the Chapman–Enskog method at the level of the linearized Burnett approximation. © *1998 American Institute of Physics.* [S1063-7761(98)01306-7]

1. INTRODUCTION

The methods of the kinetic theory of gases are often used to substantiate the limits of applicability of phenomenological nonequilibrium thermodynamics.[1,2] A long-standing limitation on the application of nonequilibrium thermodynamics—that the classical form is compatible with the kinetic theory of rarefied gases only at the level of the first approximation in the well-known Chapman–Enskog method[7,8]—has been removed in a recent series of works.[3–6] It has been shown that nonequilibrium thermodynamics is also compatible with higher approximations of the method, for example, with the results of the linearized Burnett approximation. In the latter case additional terms, which are proportional to the second derivatives of the velocity and temperature, and the corresponding ''nonphysical'' fluxes, whose introduction ensures satisfaction of the Onsager relations,[5,6] appear in the expression for the local entropy production.

As far back as 1948, Grad noted on the basis of his method of moments[9,10] that nonequilibrium thermodynamics is applicable in more general situations where the nonequilibrium state of a gas (and the nonequilibrium entropy) is determined not only by the local values of the density and internal energy (temperature) of the gas, as in the standard classical scheme of nonequilibrium thermodynamics, but also by any number of additional state variables (moments of the distribution function). This idea was then implemented in attempts to construct a extended irreversible thermodynamics, which employs the heat flux and the viscous-stress tensor,[11–13] and sometimes even a larger number of moments,[14,15] as additional variables.

The specific area of application of extended irreversible thermodynamics has been the analysis of situations where the characteristic time in the problem is comparable to the relaxation time due to intermolecular collisions, which calls for the use of nonstationary moment equations in the theory. At the same time, questions concerning the problem of taking into account the spatial derivatives of fluxes (moments) of a different tensor order in the linearized equations remained on the sidelines. The presence of such derivatives in the expressions for the heat flux and the viscous-stress tensor as well as in other nonphysical fluxes which follow from the solution of the moment equations, sets these expressions apart from the standard linear transport relations of classical nonequilibrium thermodynamics. However, it can be shown that such a generalization does not result in any conflict with the canonical results of nonequilibrium thermodynamics. Despite the fact that the new terms make the expressions for the thermodynamic forces somewhat more complicated, the expression for entropy production remains a bilinear combination of the generalized thermodynamic forces and fluxes. At the same time, allowing for additional moments of the distribution function greatly expands the system of phenomenological equations for the fluxes and forces, while the cross coefficients in the expressions for the fluxes with the same tensor order satisfy the Onsager symmetry relations.

Our aim in the present paper is to construct such a generalization of nonequilibrium thermodynamics. An expression for the local entropy production is derived on the basis of an expansion of the distribution function into a series in orthogonal tensor polynomials, which, to within a normalization, are products of Sonine polynomials and tensor spherical harmonics, as well as the use of the moment equations obtained from a linearized kinetic equation. An application of perturbation theory in the small Knudsen number to the system of moment equations then yields results that are virtually identical to the well-known results of the Chapman–Enskog method at the level of both the first and second (Burnett) approximations. This pertains both to the generalized expressions for the fluxes and the corresponding representation for the entropy production and to the concrete results which make it possible to calculate all required transport coefficients with an accuracy that corresponds to retaining an arbitrary number of Sonine polynomials in the expansion of the distribution function.

2. EXPANSION OF THE DISTRIBUTION FUNCTION AND THE MOMENT EQUATIONS

Let us consider a stationary state of a rarefied monatomic gas described by a distribution function $f(\mathbf{v},\mathbf{r})$, where $\mathbf{v}$ is the particle velocity and $\mathbf{r}$ is the particle position. Let the state of the gas deviate slightly from equilibrium. Then $f(\mathbf{v},\mathbf{r})$ can be represented in the form

$$f=f^{(0)}(1+\Phi), \quad f^{(0)}=n(\beta/\pi)^{3/2}\exp(-\beta c^2), \tag{1}$$

where $f^{(0)}$ is the local Maxwellian velocity distribution of the particles, Φ is a small correction ($|\Phi|\ll 1$), $\beta=m/2kT$, n is the density, T is the temperature, m is the particle mass, $\mathbf{c}=\mathbf{v}-\mathbf{u}$ is the velocity of the particles relative to the center of mass, and $\mathbf{u}$ is the macroscopic velocity of the gas.

The correction $\Phi(\mathbf{v},\mathbf{r})$ satisfies the linearized Boltzmann equation[8,10]

$$(\mathbf{v}\cdot\nabla)\ln f^{(0)}+(\mathbf{v}\cdot\nabla)\Phi=L\Phi, \tag{2}$$

where

$$\nabla\ln f^{(0)}=\nabla\ln p+(\beta c^2-5/2)\nabla\ln T+2\beta\mathbf{c}(\nabla\cdot\mathbf{u}). \tag{3}$$

In addition,

$$L\Phi=\int\int(\Phi'+\Phi_1'-\Phi-\Phi_1)f_1^{(0)}g\sigma(g,\Omega)d\Omega\,d\mathbf{v}_1 \tag{4}$$

is the linearized collision operator for molecules with the velocities $\mathbf{v}$ and $\mathbf{v}_1$ (the prime indicates that the distribution function depends on the postcollision velocities of the molecules), $p=nkT$ is the pressure, $\sigma(g,\Omega)$ is the differential collision cross section, g is the relative velocity of the colliding molecules, and Ω is the scattering angle.

It is helpful to introduce the scalar product of functions in a Hilbert space:

$$(g,h)=\frac{1}{n}\int f^{(0)}(\mathbf{c})g(\mathbf{c})h(\mathbf{c})d\mathbf{c}.$$

The parameters of the local Maxwellian distribution $(n,\mathbf{u},T)$ are defined in the same manner with respect to both the complete distribution function f and $f^{(0)}$. This leads to the conditions

$$(1,\Phi)=0, \quad (\mathbf{c},\Phi)=0, \quad (c^2,\Phi)=0. \tag{5}$$

Two well-known properties of the linearized collision operator will be employed below:[8] the symmetry relation

$$(\Psi,L\Phi)=(L\Psi,\Phi) \tag{6}$$

and the condition

$$(\Phi,L\Phi)\leqslant 0. \tag{7}$$

The equal sign in the latter relation corresponds to the case where Φ is an invariant of the particle collisions.

Let us expand the nonequilibrium correction Φ into a series in an orthonormal system of tensor polynomials $\mathbf{P}^{ps}(\mathbf{W})$ of the dimensionless relative velocity of the particles $\mathbf{W}=\mathbf{c}\sqrt{\beta}$:

$$\Phi=\sum_{p=0}^{\infty}\sum_{s=0}^{\infty}\mathbf{a}^{ps}(\mathbf{r})\otimes\mathbf{P}^{ps}(\mathbf{W}), \tag{8}$$

where the symbol $\otimes$ denotes the scalar product of the tensors. The polynomials have the form (see Refs. 16 and 17)[1)]

$$\mathbf{P}^{ps}(\mathbf{W})=\gamma_{ps}S_{p+1/2}^{(s)}(W^2)\mathbf{R}^p(\mathbf{W}). \tag{9}$$

Here $S_{p+1/2}^{(s)}(W^2)$ are Sonine polynomials,[1] the $\mathbf{R}^p(\mathbf{W})$ are tensor spherical harmonics,[16] and γ_{ps} is a normalization factor:

$$\gamma_{ps}=(-1)^s\sqrt{\frac{2^{p+s}s!(2p+1)!!}{p!(2p+2s+1)!!}}.$$

The polynomials $\mathbf{P}^{ps}$ are normalized by the condition

$$(\mathbf{P}^{ps},\mathbf{P}^{p's'})=\delta_{pp'}\delta_{ss'}\Delta^{(p)}, \tag{10}$$

where $\Delta^{(p)}$ is the unit projection tensor.[17]

The first few polynomials are:

$$\begin{aligned}&P^{00}=1, \quad P^{01}=\sqrt{2/3}(W^2-3/2), \quad \mathbf{P}^{10}=\sqrt{2}\mathbf{W},\\ &\mathbf{P}^{11}=\sqrt{4/5}\mathbf{W}(W^2-5/2), \quad \mathbf{P}^{20}=\sqrt{2}\,\overline{\mathbf{WW}},\\ &\mathbf{P}^{30}=\sqrt{4/3}\,\overline{\mathbf{WWW}},\cdots\end{aligned} \tag{11}$$

Here $\overline{aaa\ldots}$ is used to denote irreducibility of the tensor. For example,

$$(\overline{\mathbf{WW}})_{ik}=W_iW_k-\frac{1}{3}\delta_{ik}W^2,$$

$$(\overline{\mathbf{WWW}})_{ikl}=W_iW_kW_l-\frac{1}{5}W^2(\delta_{ik}W_l+\delta_{il}W_k+\delta_{kl}W_i).$$

We note that

$$\mathbf{a}^{ps}=\frac{1}{n}\int\mathbf{P}^{ps}f^{(0)}\Phi d\mathbf{c}=(\mathbf{P}^{ps},\Phi) \tag{12}$$

follows from the orthogonality of the polynomials $\mathbf{P}^{ps}$ with a Maxwellian weight function. On account of the conditions (5) we have $a^{00}=\mathbf{a}^{10}=a^{01}=0$, i.e., the expansion (8) actually starts from the polynomials $\mathbf{P}^{11}$ and $\mathbf{P}^{20}$ with the corresponding coefficients $\mathbf{a}^{11}$ and $\mathbf{a}^{20}$.

The equations for the coefficients $\mathbf{a}^{ps}$ (the linearized moment equations) can be obtained by multiplying the kinetic equation (2) by $f^{(0)}\mathbf{P}^{ps}$ and integrating over the velocities.

For simplicity, below we shall consider only slow gas flows, for which, along with the condition that the gradients of the principal thermodynamic quantities $(n,\mathbf{u},T)$ be small, the condition

$$|\mathbf{u}|\ll\sqrt{kT/m} \tag{13}$$

also holds. In this case the terms of order $\mathbf{u}\cdot\nabla\ln f^{(0)}$ and $(\mathbf{u}\cdot\nabla)\Phi$ in Eq. (2) can be neglected. The latter means that the variable $\mathbf{c}$ or $\mathbf{v}$ can be used equally well in Eqs. (2) and (3). The corresponding system of moment equations then becomes

$$(\mathbf{P}^{ps},(\mathbf{c}\cdot\nabla)\ln f^{(0)})+\sum_{p's'}(\mathbf{P}^{ps},\mathbf{c}\mathbf{P}^{p's'})\otimes\nabla\mathbf{a}^{p's'}$$

$$=\sum_{p's'}(\mathbf{P}^{ps},L\mathbf{P}^{p's'})\otimes\mathbf{a}^{p's'}. \tag{14}$$

It is easy to see that $(\mathbf{c}\cdot\nabla)\ln f^{(0)}$ contains a linear combination of the polynomials $\mathbf{P}^{10}$, $\mathbf{P}^{11}$, and $\mathbf{P}^{20}$. Then, since the polynomials satisfy the orthogonality condition (10), we have

$$(\mathbf{P}^{ps},(\mathbf{c}\cdot\nabla)\ln f^{(0)})=(\sqrt{2}/2)\beta^{-1/2}\nabla\ln p\,\delta_{p1}\delta_{s0}+\sqrt{5/4}\beta^{-1/2}\nabla\ln T\delta_{p1}\delta_{s1}+\sqrt{2}\overline{\nabla\mathbf{u}}\,\delta_{p2}\delta_{s0}. \tag{15}$$

It can be shown for symmetric irreducible tensors $\mathbf{P}^{ps}$ that the flux term on the left-hand side of Eq. (14) is a linear combination of the derivatives of the coefficients of tensor orders $p+1$ and $p-1$ with respect to the coordinates (Refs. 16 and 17)

$$\sum_{p's'}(\mathbf{P}^{ps},\mathbf{c}\mathbf{P}^{p's'})\otimes\nabla\mathbf{a}^{p's'}=\sum_{s'=0}^{\infty}(A_{ps}^{p+1,s'}\nabla\mathbf{a}^{p+1,s'}+B_{ps}^{p-1,s'}\overline{\nabla\mathbf{a}}^{p-1,s'}), \tag{16}$$

where the operators A and B are given by the expressions

$$A_{ps}^{p+1,s'}=\frac{1}{2p+3}(\mathbf{P}^{ps}\otimes\mathbf{c}\mathbf{P}^{p+1,s'}),$$
$$B_{ps}^{p-1,s'}=\frac{1}{2p+1}(\mathbf{P}^{ps}\otimes\mathbf{c}\mathbf{P}^{p-1,s'}). \tag{17}$$

The notation $\overline{\nabla\mathbf{a}}^{p-1,s'}$ denotes a symmetric irreducible tensor. Thus, if $\mathbf{a}^{p-1,s'}$ is a vector ($p=2$), then $\overline{\nabla\mathbf{a}}^{1s'}$ is an irreducible tensor of rank 2,

$$\frac{1}{2}\left(\frac{\partial a_i^{1s'}}{\partial x_j}+\frac{\partial a_j^{1s'}}{\partial x_i}\right)-\frac{1}{3}\,\delta_{ij}\,\frac{\partial a_l^{1s'}}{\partial x_l}.$$

It is convenient to express the right-hand side of the system (14), or the ''moment with respect to the collision integral,'' by means of the so-called integral brackets of the corresponding polynomials. To this end we make use of the relation[17]

$$(\mathbf{P}^{ps},L\mathbf{P}^{p's'})=-\delta_{pp'}n[\mathbf{P}^{ps},\mathbf{P}^{p's'}]\Delta^{(p)}, \tag{18}$$

where

$$[\mathbf{P}^{ps},\mathbf{P}^{p's'}]=\frac{1}{4n^2}\int f^{(0)}f_1^{(0)}\Delta P^{ps}\Delta P^{p's'}g\sigma\,d\Omega\,\mathbf{dc}\,\mathbf{dc}_1. \tag{19}$$

Here

$$\Delta F=F(\mathbf{c}')+F(\mathbf{c}_1')-F(\mathbf{c})-F(\mathbf{c}_1).$$

Using Eq. (18) we find

$$\sum_{p's'}(\mathbf{P}^{ps},L\mathbf{P}^{p's'})\otimes\mathbf{a}^{p's'}=-\sum_{s'=0}^{\infty}n\Lambda_{pss'}\mathbf{a}^{ps'}, \tag{20}$$

where $\Lambda_{pss'}$ can be expressed in terms of the known integral brackets of Sonine polynomials.[7,8] Allowing for the explicit form (9) of the polynomials $\mathbf{P}^{ps}$, we have

$$\Lambda_{pss'}=\frac{1}{2p+1}\,\gamma^{ps}\gamma^{ps'}\times[S_{p+1/2}^{(s)}(W^2)\mathbf{R}^{(p)}(\mathbf{W}),S_{p+1/2}^{(s')}(W^2)\mathbf{R}^{(p)}(\mathbf{W})]. \tag{21}$$

The expressions (20) and (21) take into account that the total trace of the projection tensor $\Delta^{(p)}$ equals $2p+1$.

The moment equations thus assume the following final form:

$$(\mathbf{P}^{ps},(\mathbf{c}\cdot\nabla)\ln f^{(0)})+\sum_{s'=0}^{\infty}(A_{ps}^{p+1,s'}\nabla\mathbf{a}^{p+1,s'}+B_{ps}^{p-1,s'}\overline{\nabla\mathbf{a}}^{p-1,s'})=-\sum_{s'=0}^{\infty}n\Lambda_{pss'}\mathbf{a}^{ps'}, \tag{22}$$

where the first term on the left-hand side is defined by (15).

The direct application of the moment equations in the form (22) is of interest in cases where their right-hand side is nonzero. It is easy to see for the polynomials P^{00}, $\mathbf{P}^{10}$, and P^{01} that the right-hand side of the initial moment equations (14) equals zero by virtue of the condition (7), since these polynomials are collision invariants corresponding to the conservation of the mass, momentum, and energy of the particles in collisions. In this case the standard conservation equations (the equations of hydrodynamics) for slow gas flows follow from Eq. (14):

$$\nabla\cdot\mathbf{u}=0,\quad \nabla p+\nabla\boldsymbol{\pi},\quad \nabla\cdot\mathbf{q}=0. \tag{23}$$

Here $\boldsymbol{\pi}$ is the viscous-stress tensor, and $\mathbf{q}$ is the heat flux. They are defined by the expressions

$$\boldsymbol{\pi}=mn(\overline{\mathbf{cc}},\Phi)=\sqrt{2}\,p\mathbf{a}^{20},$$
$$\mathbf{q}=p((\beta c^2-5/2)\mathbf{c},\Phi)=(5/4)^{1/2}\beta^{-1/2}p\mathbf{a}^{11}. \tag{24}$$

Here $(\nabla\boldsymbol{\pi})_i=\partial\pi_{ik}/\partial x_k$.

3. SOLUTIONS OF THE MOMENT EQUATIONS

The equations (22) form an infinite system of coupled equations (on account of the presence of the flux terms on the left-hand side) for scalar ($p=0$), vector ($p=1$), and tensor ($p=2,3,\ldots$) quantities. They can be solved if the expansion (8) is limited to a finite number of terms.

Let us consider an approximation where terms including tensor polynomials of rank no higher than 3 are retained in the expansion, i.e., let us set $\mathbf{a}^{ps}=0$ for $p\geqslant 4$. The corresponding systems of equations (22) for $p=1,2,3,\ldots$ become

$$(5/4)^{1/2}\beta^{-1/2}\nabla\ \ln T\delta_{s1}+\sum_{s'=0}^{\infty}(A_{1s}^{2s'}\nabla\mathbf{a}^{2s'}+B_{1s}^{0s'}\nabla\mathbf{a}^{0s'})$$
$$=-\sum_{s'=0}^{\infty}n\Lambda_{1ss'}\mathbf{a}^{1s'}, \tag{25}$$

$$\sqrt{2}\overline{\nabla\mathbf{u}}\delta_{s0}+\sum_{s'=0}^{\infty}(A_{2s}^{3s'}\nabla\mathbf{a}^{3s'}+B_{2s}^{1s'}\overline{\nabla\mathbf{a}^{1s'}})$$
$$=-\sum_{s'=0}^{\infty}n\Lambda_{2ss'}\mathbf{a}^{2s'}, \tag{26}$$

$$\sum_{s'=0}^{\infty}B_{3s}^{2s'}\overline{\nabla\mathbf{a}^{2s'}}=-\sum_{s'=0}^{\infty}n\Lambda_{3ss'}\mathbf{a}^{3s'}. \tag{27}$$

The solution of the equations can be further simplified by truncating the series at finite values of s and s'. As an example, we shall consider the well-known 20-moment Grad approximation,[9,10] where, along with the coefficients $\mathbf{a}^{11}$ and $\mathbf{a}^{20}$, the coefficient $\mathbf{a}^{30}$ with the corresponding polynomial $\mathbf{P}^{30}=\sqrt{4/3}\,\overline{\mathbf{WWW}}$ is retained in the expansion. The coefficient $\mathbf{a}^{30}$ is related to the third-order moment $\hat{\mathbf{S}}=m\int\overline{\mathbf{ccc}}\,f\mathbf{dc}$ of the distribution function, so that

$$\hat{\mathbf{S}}=\sqrt{3}\beta^{-1/2}p\mathbf{a}^{30}. \tag{28}$$

In this case Eqs. (25)–(27) assume the form

$$\sqrt{5/4}\beta^{-1/2}\nabla\ \ln T+A_{11}^{20}\nabla\mathbf{a}^{20}=-n\Lambda_{111}\mathbf{a}^{11},$$
$$\sqrt{2}\overline{\nabla\mathbf{u}}+A_{20}^{30}\nabla\mathbf{a}^{30}+B_{20}^{11}\overline{\nabla\mathbf{a}^{11}}=-n\Lambda_{200}\mathbf{a}^{20}, \tag{29}$$
$$B_{30}^{20}\overline{\nabla\mathbf{a}^{20}}=-n\Lambda_{300}\mathbf{a}^{30}.$$

The coefficients A and B, as well as $\Lambda_{pss'}$, can be calculated on the basis of the expressions (17) and (21). Using the known expressions for the integral brackets relating them to $\Omega^{(l,s)}$ integrals,[7,8] as well as Eqs. (24) and (28), we arrive at the following expressions for $\mathbf{q}$, $\boldsymbol{\pi}$, and $\hat{\mathbf{S}}$:

$$\mathbf{q}=-L_{11}\left[\frac{\nabla T}{T}+\frac{2}{5}\frac{1}{p}\nabla\pi\right], \tag{30}$$

$$\pi=-L_{22}\left[\overline{\nabla\mathbf{u}}+\frac{2}{5}\frac{1}{p}\overline{\nabla\mathbf{q}}+\frac{1}{2p}\nabla\hat{\mathbf{S}}\right], \tag{31}$$

$$\hat{\mathbf{S}}=-L_{33}\frac{\overline{\nabla\pi}}{2p}, \tag{32}$$

where

$$(\overline{\nabla\pi})_{ijk}=\left(\frac{\partial\pi_{jk}}{\partial x_i}+\frac{\partial\pi_{ik}}{\partial x_j}+\frac{\partial\pi_{ij}}{\partial x_k}\right)$$
$$-\frac{2}{5}\left(\frac{\partial\pi_{il}}{\partial x_l}\,\delta_{jk}+\frac{\partial\pi_{jl}}{\partial x_l}\,\delta_{ik}+\frac{\partial\pi_{kl}}{\partial x_l}\,\delta_{ij}\right).$$

The coefficients L_{11}, L_{22}, and L_{33} are related to the viscosity $[\eta]_1$ and the thermal conductivity $[\lambda]_1$, which appear in the first approximation (with respect to the number of Sonine polynomials in the Chapman–Cowling expansion):

$$L_{11}=[\lambda]_1,\quad L_{22}=2[\eta]_1,\quad L_{33}=(4kT/3m)[\eta]_1,$$

where

$$[\eta]_1=\frac{5}{8}\frac{kT}{\Omega^{(2,2)}},\quad [\lambda]_1=\frac{75}{32}\frac{k^2T}{m\Omega^{(2,2)}}. \tag{33}$$

Generally speaking, the system of equations (30) and (31) for $\mathbf{q}$ and $\boldsymbol{\pi}$ must be solved self-consistently, since the thermodynamic forces contain derivatives of the fluxes. This situation is somewhat different from the conventional representation of the phenomenological equations of nonequilibrium thermodynamics, where the fluxes appear on the left-hand side and the gradients of the standard hydrodynamic variables appear on the right-hand side. However, it is easy to see that this system can formally be brought into a classical form, but with transport coefficients represented in operator form. It is equally correct to use such a representation in the canonical nonequilibrium thermodynamics along with the standard representation.[1]

The expressions for $\mathbf{q}$ and $\boldsymbol{\pi}$, corresponding to the known 13-moment approximation[9] follow from Eqs. (30)–(32) if we set $\hat{\mathbf{S}}=0$. In this case a different representation of Eqs. (30) and (31) in a form close to the canonical representation is possible. The heat flux $\mathbf{q}$ can be expressed in terms of the pressure and temperature gradients [see Eq. (45) below]. Then only derivatives of the standard thermodynamic variables (including the second derivatives of temperature and pressure) remain in the expression for $\boldsymbol{\pi}$ on the right-hand side.

Including a larger number of terms in the expansion makes the structure of the solutions appreciably more complicated, though the formal representation of the solutions for the coefficients $n\mathbf{a}^{ps}$ reduces mainly to finding the inverse matrix of coefficients $(\Lambda_{pss'})^{-1}$. The solutions for the 26-moment approximation were studied in, for example, Refs. 14 and 15.

4. ENTROPY BALANCE EQUATION

As is generally known, nonequilibrium thermodynamics is based on a local entropy balance equation, which can be obtained from the complete (not linearized) kinetic equation for f by multiplying by $\ln f$ and integrating over the velocities. The local entropy density ρs (ρ is the density) is then specified by the expression[1,19]

$$\rho s=-k\int f\ \ln f\mathbf{dc}+kn \tag{34}$$

(k is Boltzmann's constant, n is the particle density), while the balance equation for the stationary case has the form

$$\nabla\cdot\mathbf{J}_s=\sigma, \tag{35}$$

where

$$\mathbf{J}_s=-k\int\mathbf{v}f\ \ln f d\mathbf{c}+kn\mathbf{u} \tag{36}$$

is the total entropy flux density and

$$\sigma = -k\int f^{(0)} \ln f L\Phi d\mathbf{c} \tag{37}$$

is the local entropy production.

Substituting f in the form (1) into Eqs. (34), (36), and (37) and using the condition (5), we obtain up to terms quadratic in the small correction Φ

$$\rho s = (\rho s)_0 - \frac{kn}{2}(1,\Phi^2), \quad \mathbf{J}_s = \rho s\mathbf{u} + \frac{\mathbf{q}}{T} - \frac{kn}{2}(\mathbf{c},\Phi^2),$$
$$\sigma = -kn(\Phi, L\Phi). \tag{38}$$

Here the index 0 corresponds to the entropy density determined in the state of local equilibrium.

Using the expansion (8), Eqs. (34) and (36) can be represented as

$$\rho s = (\rho s)_0 - \frac{kn}{2}\sum_{p,s}\mathbf{a}^{ps}\mathbf{a}^{ps},$$
$$\mathbf{J}_s = \rho s\mathbf{u} + \frac{\mathbf{q}}{T} - \frac{kn}{2}\sum_{p,s}\sum_{p',s'}\mathbf{a}^{ps}(\mathbf{P}^{ps},\mathbf{c}\mathbf{P}^{p's'})\mathbf{a}^{p's'}. \tag{39}$$

In the 20-moment approximation these expressions assume the form[14]

$$\rho s = (\rho s)_0 - \frac{1}{4pT}\pi\pi - \frac{m}{5pkT^2}\mathbf{q}\mathbf{q} - \frac{1}{12}\frac{m}{pkT^2}\hat{\mathbf{S}}\hat{\mathbf{S}},$$
$$J_{si} = \rho s u_i + \frac{q_i}{T} - \frac{2}{5pT}q_j\pi_{ij} - \frac{5}{14pT}\hat{S}_{ijk}\pi_{jk}. \tag{40}$$

We are henceforth interested mainly in the expression (38) for the local entropy production σ. We substitute the expansion for Φ into it and replace the linearized collision integral $L\Phi$ by the left-hand side of the kinetic equation (2). The result is

$$T\sigma = -kT\sum_{p,s} n\mathbf{a}^{ps}\otimes\left[(\mathbf{P}^{ps},(\mathbf{c}\cdot\nabla)\ln f^{(0)}) + \sum_{p',s'}(\mathbf{P}^{ps},\mathbf{c}\mathbf{P}^{p's'})\otimes\nabla\mathbf{a}^{p's'}\right], \tag{41}$$

or, taking into account the definitions (24) and (16),

$$T\sigma = -\frac{1}{T}q\nabla T - \boldsymbol{\pi}\otimes\overline{\nabla\mathbf{u}} - kT\sum_{p,s} n\mathbf{a}^{ps} \otimes\left[\sum_{s'=0}^{\infty} A_{ps}^{p+1,s'}\nabla\mathbf{a}^{p+1,s'} + B_{ps}^{p-1,s'}\overline{\nabla\mathbf{a}}^{p-1,s'}\right]. \tag{42}$$

The first two terms in Eq. (42) correspond to the standard representation of the local entropy production in the form of a bilinear combination of the fluxes $\mathbf{q}$ and $\boldsymbol{\pi}$ and the thermodynamic forces conjugate to them, which can be obtained within the classical scheme of nonequilibrium thermodynamics for vector and tensor phenomena.[1] A new feature is that the same fluxes ($p\mathbf{a}^{11}\sim\mathbf{q}$ and $p\mathbf{a}^{20}\sim\boldsymbol{\pi}$) appear together with the additional forces conjugate to them in the next terms in (42).

For the 20-moment approximation, a calculation of the entropy production in this case gives

$$T\sigma = -\mathbf{q}\left(\frac{1}{T}\nabla T + \frac{2}{5p}\nabla\pi\right) - \boldsymbol{\pi}\otimes\left(\nabla\mathbf{u} + \frac{2}{5}\frac{1}{p}\nabla\mathbf{q} + \frac{1}{2p}\nabla\hat{\mathbf{S}}\right) - \frac{1}{2p}\hat{\mathbf{S}}\otimes\overline{\nabla\boldsymbol{\pi}}. \tag{43}$$

The linear phenomenological relations of nonequilibrium thermodynamics corresponding to the representation (43) are completely consistent with the expressions (30)–(32) following from a direct solution of the system of moment equations.

We note that the complication of the form of the thermodynamic forces on account of the derivatives of the fluxes in (30) and (31) did not necessitate a revision of the values of the coefficients L_{11} and L_{12}, which once again are determined by the values of the ordinary viscosity and thermal conductivity. Moreover, just as in the standard classical scheme,[1] the entropy production is found to be a linear combination of terms which are quadratic in the fluxes:

$$T\sigma = \frac{1}{L_{11}}\mathbf{q}\mathbf{q} + \frac{1}{L_{22}}\boldsymbol{\pi}\otimes\boldsymbol{\pi} + \frac{1}{L_{33}}\hat{\mathbf{S}}\otimes\hat{\mathbf{S}}. \tag{44}$$

The condition $\sigma\geqslant 0$ (positive entropy production), which follows from the properties of the linearized collision integral (7), is guaranteed by the positive values of the coefficients L_{ii} or the obvious conditions $\eta>0$ and $\lambda>0$.

We note for slow flows that the relation $\nabla\boldsymbol{\pi} = -\nabla p$ follows from the equation of motion (23). Then the expression (30) for the heat flux assumes the form

$$\mathbf{q} = -L_{11}\left(\frac{\nabla T}{T} - \frac{2}{5}\frac{\nabla p}{p}\right). \tag{45}$$

Curiously, in the well-known monograph by Landau and Lifshitz (Ref. 20) the important assumption that the heat flux can depend only on the temperature gradient is used in the derivation of the expression for the entropy production. The corresponding proof is as follows (Ref. 20, p. 274): ''If $\mathbf{q}$ contained a term proportional to ∇p, the expression... for the rate of change of entropy would include another term having the product $\nabla p\cdot\nabla T$ in the integrand. Since the latter might be either positive or negative, the time derivative of the entropy would not necessarily be positive, which is impossible.''

In fact, it fact from the expressions (44) and (45) obtained above that the part of the entropy production that corresponds to the heat flux can be represented in the form

$$T\sigma_q = L_{11}\left(\frac{\nabla T}{T} - \frac{2}{5}\frac{\nabla p}{p}\right)^2, \tag{46}$$

i.e., it remains essentially positive despite the fact that the expression for the heat flux contains a term with a pressure gradient.

The cross terms, whose coefficients ordinarily satisfy the Onsager relations, are absent in the system of equations (30)–(32). This is because these equations correspond to the minimum number of polynomials in the expansion for Φ of the same tensor order (first, second, and third). It is easy to show that when a larger number of expansion coefficients

and corresponding fluxes (which no longer have any direct physical meaning) are taken into account, it is possible to obtain a set of linear phenomenological relations containing the cross terms for the fluxes. Thus, if the coefficients $n\mathbf{a}^{ps}$ are chosen as the fluxes $\mathbf{J}^{ps}$ and the expressions

$$\mathbf{F}^{ps}=kT\Bigg[(\mathbf{P}^{ps},(\mathbf{c}\cdot\nabla)\ln f^{(0)})+\sum_{s'}A_{ps}^{p+1,s'}\nabla\mathbf{a}^{p+1,s'}+B_{ps}^{p-1,s'}\overline{\nabla\mathbf{a}}^{p-1,s'}\Bigg], \tag{47}$$

are taken as the thermodynamic forces, then the system of linear phenomenological equations of different tensor order relating the thermodynamic fluxes and forces becomes

$$\begin{aligned}&\mathbf{J}^{0s}=\sum_{s'}L_{ss'}^{(0)}\mathbf{F}^{0s'},\\ &\mathbf{J}^{1s}=\sum_{s'}L_{ss'}^{(1)}\mathbf{F}^{1s'},\\ &\cdots\cdots\cdots\cdots\\ &\mathbf{J}^{ps}=\sum_{s'}L_{ss'}^{(p)}\mathbf{F}^{ps'},\end{aligned} \tag{48}$$

where the cross kinetic coefficients $L_{ss'}^{(p)}$ satisfy the Onsager relations ($L_{ss'}^{(p)}=L_{s's}^{(p)}$).

Equations (48) have the same structure as the direct solution of Eqs. (22) for the coefficients $n\mathbf{a}^{ps}$. The matrix of phenomenological coefficients L_{ij} is identical to the inverse matrix of coefficients Λ. The symmetry of the cross coefficients then follows from the symmetry of the coefficients $\Lambda_{pss'}$ defined by the expression (21).

5. RELATION TO THE CHAPMAN-ENSKOG RESULTS

It is helpful to clarify the assumptions under which the results obtained by the method of moments agree with the results obtained by the standard Chapman–Enskog method.[7,8] The latter method is known to be based on the application of perturbation theory with the Knudsen number as the small parameter ($\mathrm{Kn}=l/L\ll 1$, where l is the mean free path of the particles and L is the characteristic dimension in the problem). The parameter Kn^{-1} is introduced on the right-hand side of the kinetic equation, and the distribution function is expanded into a series in the small Knudsen number. We apply this procedure not to the distribution function, but rather to the expansion coefficients $\mathbf{a}^{mn}$. This is made possible, because the coefficients $\Lambda_{pss'}$ on the right-hand side of the linearized moment equations are of the order of the reciprocal τ^{-1} of the characteristic time between particle collisions, where $\tau=1/\langle v\rangle=[\eta]_1/p$ and $\langle v\rangle$ is the average thermal velocity of the particles.

Using the formal expansion of the coefficients $\mathbf{a}^{ps}$ into a series of the form

$$\mathbf{a}^{ps}=\mathbf{a}_{(1)}^{ps}+\mathrm{Kn}\,\mathbf{a}_{(2)}^{ps}+\mathrm{Kn}^2\mathbf{a}_{(3)}^{ps}+\ldots,$$

substituting these series into Eqs. (25)–(27), and equating the terms with like powers of Kn, we arrive at systems of equations of first-, second-, and higher-order perturbation theory. The first approximation corresponds to a system of equations on whose left-hand sides the flux terms with derivatives of the coefficients $\mathbf{a}^{ps}$ ($p=0,1,2,3,\ldots$) are omitted. For the vector and tensor coefficients $\mathbf{a}_{(1)}^{1k}$ and $\mathbf{a}_{(1)}^{2k}$ we have, specifically, two independent systems of algebraic equations of the form

$$\begin{aligned}&-\sqrt{5/4}\beta^{-1/2}\nabla\ln T\,\delta_{p1}=\sum_{s=1}^{\infty}n\Lambda_{1ps}\mathbf{a}_{(1)}^{1s},\\ &-\sqrt{2}\overline{\nabla\mathbf{u}}\,\delta_{p2}=\sum_{s=0}^{\infty}n\Lambda_{2ps}\mathbf{a}_{(1)}^{2s}.\end{aligned} \tag{49}$$

These equations can be solved with respect to the fluxes $\mathbf{J}_{1s}=n\mathbf{a}_{(1)}^{1s}$ and $\mathbf{J}_{2s}=n\mathbf{a}_{(1)}^{2s}$ for any finite value $p=\xi$. Specifically, values are obtained for $n\mathbf{a}_{(1)}^{11}$ and $n\mathbf{a}_{(1)}^{20}$ or the corresponding heat flux $\mathbf{q}$ and viscous-stress tensor $\boldsymbol{\pi}$ to them to any approximation in ξ. The latter can be represented as

$$\mathbf{q}=-\lambda\nabla T,\quad \pi=-2\eta\overline{\nabla\mathbf{u}}. \tag{50}$$

It is easy to show that these solutions are completely consistent with the results of the first (non-Burnett) approximation in the Chapman–Enskog method.[8]

We recall that in this method the first-approximation correction to the distribution function is sought in the form

$$\Phi_1=\Phi_t(\mathbf{c}\cdot\nabla)\ln T+\Phi_p\overline{\mathbf{cc}}\otimes\overline{\nabla\mathbf{u}}, \tag{51}$$

where expansions of the form

$$\mathbf{c}\Phi_t=\frac{1}{n}\sum_{k=1}^{\infty}A_{1k}\mathbf{P}^{1k},\quad \overline{\mathbf{cc}}\,\Phi_p=\frac{1}{n}\sum_{k=0}^{\infty}A_{2k}\mathbf{P}^{2k}. \tag{52}$$

are used for Φ_t and Φ_p. Comparing Eqs. (51) and (52) with the expansion (8) for Φ shows that the coefficients $\mathbf{a}^{ik}$ and A_{ik} are related by the expressions

$$\mathbf{a}_{(1)}^{1k}=\frac{1}{n}A_{1k}\nabla\ln T,\quad \mathbf{a}_{(1)}^{2k}=\frac{1}{n}A_{2k}\overline{\nabla\mathbf{u}}. \tag{53}$$

Then it follows from the definitions (24) that

$$\lambda=-\sqrt{\frac{5}{4}}k\beta^{-1/2}A_{11},\quad \eta=-\frac{\sqrt{2}}{2}kT\,A_{20}.$$

In the Chapman–Enskog method a system of algebraic equations for the coefficients A_{ik} is obtained from the integral equations for Φ_t and Φ_p[7,8] by multiplying the latter by $f^{(0)}\mathbf{P}^{1n}$ and $f^{(0)}\mathbf{P}^{2n}$, respectively, and then integrating over the velocities. It is easily found that these equations are completely equivalent to the system (49), if the relation (53) is substituted into the latter.

Thus, the coefficients A_{11} and A_{20}, in terms of which λ and η are expressed, can, in fact, be found (to any approximation in ξ) in both methods from similar systems of algebraic equations.

We now turn to the next (Burnett) approximation in the Chapman–Enskog method. The correction Φ_2 in the linearized Burnett approximation in known to have the form[5,6]

$$\Phi_2=\mathbf{\Phi}_B^T\otimes\overbrace{\frac{\nabla\nabla T}{T}}+\mathbf{\Phi}_B^u\otimes\overbrace{\nabla\nabla\mathbf{u}}+\Phi_B^v\mathbf{c}\cdot\nabla^2\mathbf{u}. \tag{54}$$

The expression obtained for the entropy production can be written using Φ_1 (51) and Φ_2 (54) in the form[6]

$$T\sigma=-\left[(\mathbf{q}\cdot\boldsymbol{\nabla})\ln T+\boldsymbol{\pi}\otimes\overbrace{\nabla\mathbf{u}}+\mathbf{J}^T\otimes\overbrace{\frac{\nabla\nabla T}{T}}+\mathbf{J}^u\otimes\overbrace{\nabla\nabla\mathbf{u}}+\mathbf{J}^v\cdot\nabla^2\mathbf{u}\right], \tag{55}$$

where the additional ''nonphysical'' fluxes are defined as

$$\mathbf{J}^T=nkT(\Phi_t\overbrace{\mathbf{cc}},\Phi),\quad \mathbf{J}^v=nkT\left(\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right)\mathbf{c},\Phi\right),$$
$$\mathbf{J}^u=nkT(\Phi_p\overbrace{\mathbf{ccc}},\Phi). \tag{56}$$

The representation (55) corresponds to three independent systems of phenomenological equations of the form

$$\mathbf{q}=-\Lambda_{11}\nabla\ln T-\Lambda_{12}\nabla^2\mathbf{u}, \tag{57a}$$
$$\mathbf{J}^v=-\Lambda_{21}\nabla\ln T-\Lambda_{22}\nabla^2\mathbf{u},$$
$$\boldsymbol{\pi}=-\lambda_{11}\overbrace{\nabla\mathbf{u}}-\lambda_{12}\overbrace{\frac{\nabla\nabla T}{T}}, \tag{57b}$$
$$\mathbf{J}^T=-\lambda_{21}\overbrace{\nabla\mathbf{u}}-\lambda_{22}\overbrace{\frac{\nabla\nabla T}{T}},$$
$$\mathbf{J}^u=-L_{11}\overbrace{\nabla\nabla\mathbf{u}}, \tag{57c}$$

where the cross coefficients of each pair of equations (57a) and (57b) satisfy the Onsager relations ($\Lambda_{12}=\Lambda_{21}$ and $\lambda_{12}=\lambda_{21}$).

We now show for the example of the vector fluxes $\mathbf{q}$ and $\mathbf{J}^v$ that expressions with completely similar structure can be obtained from the results of the method of moments presented above. We note first that the solution of the equation for Φ_B^v, just as for Φ_t, can be sought in the form of an expansion in the polynomials $\mathbf{P}^{1k}$:

$$\mathbf{c}\Phi_B^v=\frac{1}{n}\sum_k B_{1k}\mathbf{P}^{1k}. \tag{58}$$

As we have already noted, the matrix of the coefficients L_{ij} in the expression (48) for $\mathbf{J}^{ps}=n\mathbf{a}^{ps}$ is the inverse matrix of the coefficients $\Lambda_{pss'}$. Using the integral equations of the Chapman–Enskog method[6–8] for Φ_B^v and Φ_t, as well as the expansions (52) and (58), it is easy to obtain the following relations:

$$\sum_k\left(\mathbf{c}\left(\beta c^2-\frac{5}{2}\right),\mathbf{P}^{1k}\right)L_{kn}^{(1)}=A_{1n},$$
$$\sum_k\left(\mathbf{c}\left(\frac{c^2}{5}\Phi_p+\frac{\eta}{p}\right),\mathbf{P}^{1k}\right)L_{kn}^{(1)}=B_{1n}. \tag{59}$$

We now write the expressions for $\mathbf{q}$ and $\mathbf{J}^v$ in the form

$$\mathbf{q}=nkT\left(\mathbf{c}\left(\beta c^2-\frac{5}{2}\right),\Phi\right)$$
$$=kT\sum_s\left(\mathbf{c}\left(\beta c^2-\frac{5}{2}\right),\mathbf{P}^{1s}(\mathbf{W})\right)n\mathbf{a}^{1s}$$
$$=kT\sum_s\sum_{s'}\left(\mathbf{c}\left(\beta c^2-\frac{5}{2}\right),\mathbf{P}^{1s}(\mathbf{W})\right)L_{ss'}^{(1)}\mathbf{F}^{1s'},$$
$$\mathbf{J}^v=nkT\left(\mathbf{c}\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right),\Phi\right)$$
$$=kT\sum_s\left(\mathbf{c}\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right),\mathbf{P}^{1s}(\mathbf{W})\right)n\mathbf{a}^{1s}=kT$$
$$\times\sum_s\sum_{s'}\left(\mathbf{c}\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right),\mathbf{P}^{1s}(\mathbf{W})\right)L_{ss'}^{(1)}\mathbf{F}^{1s'}. \tag{60}$$

If Eqs. (59) are taken into account, the relations (60) can be rewritten as

$$\mathbf{q}=kT\sum_k A_{1k}\mathbf{F}^{1k},\quad \mathbf{J}^v=kT\sum_k B_{1k}\mathbf{F}^{1k}. \tag{61}$$

Using Eqs. (52) and (58) and the explicit form of the expressions for F^{1k} (47) we have

$$\mathbf{q}=kTn\left\{(c\Phi_t,(\mathbf{c}\cdot\boldsymbol{\nabla})\ln f^{(0)})+\frac{1}{5}\sum_k(\mathbf{c}\Phi_t,\mathbf{cP}^{2k})\nabla\mathbf{a}^{2k}\right\},$$
$$\mathbf{J}^v=kTn\left\{(c\Phi_B^v,(\mathbf{c}\cdot\boldsymbol{\nabla})\ln f^{(0)})+\frac{1}{5}\sum_k(\mathbf{c}\Phi_B^v,\mathbf{cP}^{2k})\nabla\mathbf{a}^{2k}\right\}. \tag{62}$$

Using the perturbation expansion of $\mathbf{a}^{ps}$, we can now replace the terms with the derivatives $\nabla\mathbf{a}^{ps}$ with the aid of the second relation in Eq. (53), which corresponds to the first approximation. The result is

$$\mathbf{q}=nkT\left\{(\mathbf{c}\Phi_t,(\mathbf{c}\cdot\boldsymbol{\nabla})\ln f^{(0)})+\frac{1}{5}(\mathbf{c}\Phi_t,\overbrace{\mathbf{ccc}}\Phi_p)\nabla\overbrace{\nabla\mathbf{u}}\right\},$$
$$\mathbf{J}^v=nkT\left\{(\mathbf{c}\Phi_B^v,(\mathbf{c}\cdot\boldsymbol{\nabla})\ln f^{(0)})+\frac{1}{5}(\mathbf{c}\Phi_B^v,\mathbf{c}\overbrace{\mathbf{cc}}\Phi_p)\nabla\overbrace{\nabla\mathbf{u}}\right\}. \tag{63}$$

Using (3), retaining the term with $\nabla p=\eta\Delta\mathbf{u}$ in $\nabla\ln f^{(0)}$, and isolating the vector part in the terms $\nabla\overbrace{\nabla\mathbf{u}}$, we arrive at the final result

$$\mathbf{q}=nk\left[\left(\mathbf{c}\Phi_t,\mathbf{c}\left(\beta c^2-\frac{5}{2}\right)\right)\nabla T+\left(\mathbf{c}\Phi_t,\mathbf{c}\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right)\right)\nabla^2\mathbf{u}\right],$$
$$\mathbf{J}^v=nk\left[\left(\mathbf{c}\Phi_B^v,\mathbf{c}\left(\beta c^2-\frac{5}{2}\right)\right)\nabla T+\left(\mathbf{c}\Phi_t,\mathbf{c}\left(\Phi_B^v\frac{c^2}{5}+\frac{\eta}{p}\right)\right)\nabla^2\mathbf{u}\right]. \quad (64)$$

It is easy to see that the expressions obtained are in complete agreement with Eq. (57a). In accordance with the definition of λ,[7,8] we have $\Lambda_{11}=\lambda T$. We now prove that the kinetic coefficients are symmetric. Indeed,

$$\left(\mathbf{c}\Phi_B^v,\mathbf{c}\left(\beta c^2-\frac{5}{2}\right)\right)=(\mathbf{c}\Phi_B^v,L(\mathbf{c}\Phi_t)).$$

where the known[7,8] integral equation for Φ_t was employed.

On the basis of the symmetry condition (6) we have

$$(\mathbf{c}\Phi_B^v,L(\mathbf{c}\Phi_t))=(\mathbf{c}\Phi_t,L(\mathbf{c}\Phi_B^v))=\left(\mathbf{c}\Phi_t,\mathbf{c}\left(\Phi_p\frac{c^2}{5}+\frac{\eta}{p}\right)\right).$$

where the integral equation for $\mathbf{c}\Phi_B^v$ was also used.[6] It follows from the relations obtained that $\Lambda_{12}=\Lambda_{21}$.

Using transformations similar to those presented above it is easy to show that the expressions for the tensor fluxes $\boldsymbol{\pi}$ and $\mathbf{J}^T$ can also be represented in the form of the linear relations (57b). Then, for the coefficients λ in these relations we obtain $\lambda_{11}=2\eta$ and

$$\lambda_{21}=\lambda_{12}=\frac{mn}{5}(\mathbf{cc}\otimes\Phi_B^T)=\frac{kT}{5}(\Phi_t\mathbf{cc}\otimes\mathbf{cc}\Phi_p),$$
$$\lambda_{22}=\frac{kT}{5}(\Phi_t\mathbf{cc}\otimes\Phi_B^T). \quad (65)$$

The results presented above completely close the scheme for obtaining the phenomenological relations of nonequilibrium thermodynamics at the level of the Chapman–Enskog method from the linearized equations of Grad's method of moments.

This work was supported by the Russian Fund for Fundamental Research under Grants Nos. 97-01-00611 and 95-03-08092.

*)E-mail: roldughin@lmm.phyche.msk.su

1)These polynomials correspond to an orthonormal system of eigenfunctions of the linearized collision operator for the Maxwell molecules.[16] They are also identical (to within a normalization) to the irreducible Hermite tensor polynomials first used by Grad.[18]

[1] S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics*, North Holland, Amsterdam, 1962 [Russian translation, Mir, Moscow, 1964].
[2] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes*, Springer-Verlag, New York, 1987 [Russian translation, Mir, Moscow, 1990].
[3] V. I. Roldughin, Dokl. Akad. Nauk SSSR **263**, 606 (1982) [Sov. Phys. Dokl. **27**, 244 (1982)].
[4] V. I. Roldughin, J. Non-Equilib. Thermodyn. **9**, 71 (1984).
[5] I. Kuscer, Physica A **133**, 397 (1985).
[6] V. M. Zhdanov and V. I. Roldughin, Zh. Éksp. Teor. Fiz. **109**, 1267 (1996) [JETP **82**, 683 (1996)].
[7] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, Cambridge University Press, New York, 3rd ed., 1974 [Russian translation, Inostr. Lit., Moscow, 1960].
[8] J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases*, North Holland, Amsterdam, 1972 [Russian translation, Mir, Moscow, 1976].
[9] H. Grad, Commun. Pure and Appl. Math. **2**, 331 (1949).
[10] H. Grad, in *Handbuch der Physik*, Springer-Verlag, Berlin, 1958, Vol. 12, p. 205.
[11] I. Muller, Z. Phys. **198**, 329 (1967).
[12] D. Jou, J. Casas-Vazquez, and G. Lebon, Rep. Prog. Phys. **51**, 1105 (1988).
[13] D. Jou, J. Casas-Vazquez, and G. Lebon, *Extended Irreversible Thermodynamics*, Springer-Verlag, Berlin, 1993.
[14] R. M. Velasco and L. S. Garsia-Colin, J. Stat. Phys. **69**, 217 (1992).
[15] R. M. Velasco and L. S. Garsia-Colin, J. Non-Equilib. Thermodyn. **18**, 157 (1993).
[16] L. Waldmann, in *Handbuch der Physik*, Springer-Verlag, Berlin, 1958, Vol. 12, p. 295.
[17] F. R. W. McCourt, J. J. M. Beenakker, W. E. Kohler, and I. Kuscer, *Non-Equilibrium Phenomena in Polyatomic Gases*, Clarendon Press, Oxford, 1990, Vol. 1.
[18] H. Grad, Phys. Fluids **6**, 147 (1963).
[19] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, Pergamon, Elmsford, N. Y., 1981 [Russian original, Nauka, Moscow, 1979].
[20] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon, Oxford, 1987 [Russian original, Nauka, Moscow, 1986].

Translated by M. E. Alferieff
Edited by P. Shelnitz

Dynamic fluctuation phenomena in double membrane films

E. I. Kats*)

Max-Planck Institute for Physics of Complex Systems, D-01187 Dresden, Germany

V. V. Lebedev

Department of Physics, Weizmann Institute of Science, 76100 Rehovot, Israel

S. V. Malinin

L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia
(Submitted 18 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2096–2108 (June 1998)

Dynamics of double membrane films is investigated in the long-wavelength limit $qh \ll 1$ (q is the wave vector, and h is the thickness of the film) including the overdamped squeezing mode. We demonstrate that thermal fluctuations essentially modify the character of the mode due to its nonlinear coupling to the transverse shear hydrodynamic mode. The renormalization can be analyzed if the condition $g \ll 1$ is satisfied (where $g \sim T/\kappa$, T is the temperature, and κ is the bending modulus). The corresponding Green's function acquires as a function of the frequency ω a cut along the imaginary semiaxis. At $qh > \sqrt{g}$ the effective length of the cut is $\sim Tq^3/\eta$ (where η is the shear viscosity of the liquid). At $qh < \sqrt{g}$ the fluctuations lead to an increase in the attenuation of the squeezing mode: it is larger than the 'bare' value by a factor $1/\sqrt{g}$. We also present the analysis of the elastic modes. © *1998 American Institute of Physics.* [S1063-7761(98)01406-1]

1. INTRODUCTION

The most distinctive property of amphiphilic molecules is their ability to spontaneously self-assemble into aggregates of various shapes. Typically, the molecules spontaneously self-assemble into membranes which are bilayers of a thickness of the order of a molecular length. Different lyotropic structures consisting of these membranes have generated considerable current interest (see Refs. 1–3 and review articles[4–6]). Films composed of two bilayer membranes sandwiching a thin layer of a liquid are widely used in the lyotropic systems. They play also an essential role in various biological processes (note the so-called flickering phenomena in erythrocytes or red blood cells). In this paper we will examine the dynamic properties of such double membrane films.

The main peculiarity of a membrane is its negligible surface tension. The membrane is immersed into a liquid and consequently its area can vary. Zero surface tension is the equilibrium condition with respect to the variations. In this case the shape fluctuations of the membrane are determined by the bending elasticity; the corresponding energy is[7,8]

$$\mathscr{H}_{\rm curv} = \frac{\kappa}{2} \int dA \left(\frac{1}{R_1} + \frac{1}{R_2} \right)^2, \tag{1}$$

where the integral is taken over the membrane which is considered as a two-dimensional object, R_1 and R_2 are its local curvature radii, and κ is the bending rigidity modulus. Corrugations of the membrane induced by the thermal noise lead to the loss of the orientation correlation of the membrane pieces at separations larger than the so-called persistent length[9] ξ_p, which is estimated to be

$$\xi_p \sim a \, \exp(2\pi\kappa/T),$$

where T is the temperature, and a is the thickness of the membrane. The shape fluctuations of the membrane lead to the logarithmic renormalization of the bending modulus κ, which was examined first by Helfrich[10] and later by Förster.[11] The correct renormalization-group (RG) equation was derived by Peliti and Liebler,[12] Kleinert,[13] and Polyakov.[14] The explicit form of the one-loop RG equation is

$$\frac{d\kappa}{d\xi} = -\frac{3T}{4\pi}.$$

Here $\xi = \ln(r/a)$, and r is the characteristic scale. As follows from the equation, the role of the dimensionless coupling constant is played by the quantity

$$g = \frac{3T}{4\pi\kappa}. \tag{2}$$

Note that $\ln(\xi_p/a) \sim g^{-1}$. For real membranes $g \sim 10^{-2}$–10^{-3} and, consequently, we can treat g as a small parameter. The smallness of g means that there exists a wide range of scales $r < \xi_p$ where the thermal fluctuations can be treated in the framework of the perturbation theory.

Below we consider a double membrane film. We assume that at equilibrium the film is parallel to the xy plane. Corrugations of the membranes in a double film can be decomposed into undulation (or bending) deformations and squeez-

ing deformations. The bending deformations are characterized by the displacement u of the film as a whole from its equilibrium position along the z axis and the squeezing deformation is characterized by variations of the film thickness h (which is the separation between the membranes). In the harmonic approximation we obtain from (1) the energy

$$\mathscr{H}=\int dx\ dy\left[\kappa(\nabla^2 u)^2+\frac{\kappa}{4}(\nabla^2 h)^2\right], \tag{3}$$

where u and h are treated as functions of x and y, and ∇ is the two-dimensional gradient.

In deriving (3) we disregarded the interaction between the membranes. First, one should remember the steric interaction, which is associated with a certain restriction of accessible configurations for one membrane in the presence of the second membrane.[8] The explicit expression for the energy is[15]

$$\mathscr{H}_{\rm ster}=\int dx\ dy\ \frac{3\pi^2 T^2}{128\kappa h^2}. \tag{4}$$

Due to the interaction (4) two membranes can be treated independently only on scales smaller than $g^{-1/2}h$. Therefore (3) is the main contribution to the energy if

$$qh>\sqrt{g}, \tag{5}$$

where q is the characteristic wave vector. Second, we should take into account the van der Waals interaction. We assume that the same liquid is inside and outside the film. We can then write the van der Waals energy as[16]

$$\mathscr{H}_{\rm vdw}=\int dx\ dy\ \frac{Ha^2}{2\pi h^4}, \tag{6}$$

where H is the Hamaker constant. We can disregard this energy, in contrast with (3), if

$$(qh)^4>\frac{H}{\kappa}\left(\frac{a}{h}\right)^2.$$

Let us assume that the thickness of the film is large enough to satisfy the inequality

$$g^2>\frac{H}{\kappa}\left(\frac{a}{h}\right)^2.$$

Then (5) is the only restriction that enables us to treat the energy (3) as the main contribution to the film energy.

2. DYNAMICS

We will examine the dynamics of the double membrane film in the long-wavelength limit $qh\ll 1$, where q is the wave vector of the eigenmodes of the film. Note that the inequality $qh\ll 1$ is compatible with (5) since $g\ll 1$. In the limit $qh\ll 1$ we should take into account the following variables which describe the dynamics: the velocity of the film $\mathbf{v}$, the displacement of the film u, the film thickness h, and the densities of the two membranes since they are conserved quantities. We are interested mainly in the squeezing mode, which is associated with the relaxation of the thickness h.

To find dynamical characteristics of the film one should solve the conventional hydrodynamic equations in bulk supplemented by boundary conditions on both membranes. In the linear approximation the problem was solved by Brochard and Lennon,[17] who found the dispersion relation of the squeezing mode

$$\omega=-i\ \frac{\kappa h_0^3 q^6}{24\eta}, \tag{7}$$

where ω is the frequency of the mode, h_0 is the equilibrium separation between the membranes, and η is the viscosity of the liquid surrounding the membranes. In deriving (7) it was assumed that at equilibrium the film is flat. The dispersion relation

$$\omega=-i\ \frac{\kappa q^3}{2\eta} \tag{8}$$

of the bending mode was also found in the linear approximation. Note that the dispersion relation (7) is correct only if one neglects the direct interaction of the membranes upon satisfaction of the condition (5), whereas the region of applicability of the dispersion relation (8) does not depend on the interaction of the membranes, since they move in-phase in the bending mode. The elastic modes associated with variations of the membrane densities are harder than (7) and (8).[18] Therefore the only effect of the elastic degrees of freedom upon examination of the squeezing mode is the incompressibility condition

$$\nabla_\alpha v_\alpha=0. \tag{9}$$

Here and below we believe that all variables characterizing the film are functions of x and y and we assume that the Greek subscripts run over x and y.

We will consider the renormalization of the dispersion law (7) of the squeezing mode due to fluctuational effects. Nonlinear dynamical equations of the film should be utilized for this purpose. In the long-wavelength limit $qh_0\ll 1$ the equations can be derived phenomenologically. The reactive (nondissipative) part of the equations can be found by using the Poisson brackets method (see Ref. 19 and also Ref. 18), whereas the dissipative part of the equations is expressed in terms of the kinetic coefficients. One should know the expression for the energy $\mathscr{H}$ of the system to write both contributions. Actually, we will need the expression for one Poisson bracket:

$$\{j_\alpha(x_1,y_1),h(x_2,y_2)\}=h(x_1,y_1)\nabla_\alpha[\delta(x_1-x_2)\delta(y_1-y_2)], \tag{10}$$

where j_α is the two-dimensional momentum density of the film. The expression (10) (which is characteristic of two-dimensional density of any conserved scalar quantity on a film[18]) is motivated by the fact that the two-dimensional mass density of the film is ρh, where ρ is the three-dimensional density of the liquid. Note that $j_\alpha\approx\rho h v_\alpha$ since we believe that the membrane thickness a can be ignored in comparison with the film thickness h.

The dynamic equation for the thickness h has the standard form following from (10)

$$\partial_t h + \nabla_\alpha (v_\alpha h) = \Gamma \nabla^2 \frac{\delta \mathscr{H}}{\delta h}, \tag{11}$$

where $\partial_t \equiv \partial/\partial t$, and Γ is the kinetic coefficient. The second power of the gradient appears in (11) since the equation should support the conservation law of the liquid inside the film and therefore the right-hand side of the equation should be a full derivative at any $\mathscr{H}$. Due to (9) the second term on the left-hand side of (11) describes the sweeping of h by the velocity v_α. In the linear approximation we can ignore the sweeping term. Substituting the harmonic expression (3) for the energy $\mathscr{H}$ into (11) and comparing the result with (7), we obtain

$$\Gamma = h_0^3/12\eta. \tag{12}$$

Note that Γ is inversely proportional to the shear viscosity coefficient. The point is that the dissipation described by Γ comes from viscous motion of the liquid surrounding the double membrane film which is excited only slightly at large η.

The dynamic equation for j_α has the form[21]

$$\partial_t j_\alpha - \{\mathscr{H}, j_\alpha\} = J_\alpha, \tag{13}$$

where **J** is the momentum flow from the bulk to the film. Since this term supplies the main dissipation of the film momentum, we ignored the internal viscosity. The Poisson bracket $\{\mathscr{H}, j_\alpha\}$ can be reduced to the divergence of the symmetric stress tensor for any energy $\mathscr{H}$.[18] Actually, only the contribution associated with the Poisson bracket (10) and created by the harmonic energy (3) is relevant for us. We can then write Eq. (13) in the form

$$\partial_t j_\alpha + \frac{\kappa}{2} h \nabla_\alpha \nabla^4 h + \nabla_\alpha P_s = J_\alpha, \tag{14}$$

where P_s is the two-dimensional pressure, which is related to the elastic degrees of freedom (see the Appendix). In the linear approximation relevant for us we can write[21]

$$J_\alpha = -2\eta \hat{q} v_\alpha, \tag{15}$$

where $\hat{q}$ is the nonlocal operator, which is reduced to multiplying by the absolute value of the wave vector q in the Fourier representation. The expression (15) implies the inequality $\omega \ll \eta q^2/\rho$, which is satisfied for the squeezing mode.

We will not present here dynamical equations for the variables j_z and u and for the densities of the membranes. The reason is that the equations for j_z and u, which describe the bending mode, decouple in the approximation used from Eqs. (11) and (14). Actually, the equations describing the bending motion of the double film are the same as for a single membrane and the corresponding nonlinear equations can be found in Ref. 18 and also in Refs. 20 and 21. One should remember only that the bending modulus of the double film is 2κ, as follows from (3). As to the equations for the densities of the membranes, they need a separate analysis, which is presented in the Appendix. The only role of the degrees of freedom in analyzing the squeezing mode reduces to the incompressibility condition (9).

3. RENORMALIZATION OF THE SQUEEZING MODE

As can be seen from (7), in the long-wavelength limit the squeezing mode is very soft. This is the reason why one anticipates that fluctuational effects which are related to the mode are relevant. The effects are associated with the nonlinear terms in the dynamic equations and can be examined in terms of the diagrammatic technique of the first type which was developed by Wyld,[22] who studied the velocity fluctuations in a turbulent fluid. In Ref. 23 the Wyld technique was generalized for a broad class of dynamical systems. A textbook description of the diagram technique can be found in the book by Ma.[24] The diagram technique can be formulated in terms of path integrals, as was first suggested by de Dominicis[25] and Janssen.[26] In the framework of this approach apart from conventional dynamic variables one should also introduce auxiliary fields conjugated to the variables. The dynamic correlation functions of the variables can then be presented as functional integrals over both types of fields: conventional and auxiliary. The integrals are taken with the weight $\exp(i\mathscr{T})$, where $\mathscr{T}$ is the effective action which is constructed on the basis of nonlinear dynamic equations of the system.

Since we are interested in the renormalization of the squeezing mode of the double membrane film, we will take into account only the variables h and v_α and the corresponding auxiliary conjugated fields p and μ_α. We should also remember the incompressibility condition (9) and impose an analogous constraint $\nabla_\alpha \mu_\alpha = 0$ on the field μ_α. We can then write the correlation function of the film thickness h in the form

$$\langle h_1 h_2 \rangle = \int \mathscr{D}h \mathscr{D}\mathbf{v}_{tr} \mathscr{D}p \mathscr{D}\mu_{tr} \exp(i\mathscr{T}) h_1 h_2, \tag{16}$$

where the subscript ''*tr*'' implies that in the Fourier representation we should use only the components of the fields **v** and $\boldsymbol{\mu}$, which are transverse to the wave vector **q**. The explicit expression for the effective action in (16) can be found by using the dynamic equations (11) and (14). It can be written as the sum of the reactive part and the dissipative part $\mathscr{T} = \mathscr{T}_{\rm reac} + \mathscr{T}_{\rm diss}$, where

$$\mathscr{T}_{\rm reac} = \int dt\, d^2 r \Big\{ p\partial_t h + p v_\alpha \nabla_\alpha h + \mu_\alpha \partial_t j_\alpha - \frac{\kappa}{2} \mu_\alpha \nabla^4 h \nabla_\alpha h \Big\}, \tag{17}$$

$$\mathscr{T}_{\rm diss} = \int dt\, d^2 r \Big\{ -\frac{1}{2} \Gamma \kappa p \nabla^6 h + iT\Gamma (\nabla p)^2 + 2\eta\mu\hat{q}(\mathbf{v} + iT\mu) \Big\}. \tag{18}$$

The detailed derivation of the effective action for the problem can be found in Refs. 20 and 21.

We introduce the notation for the pair correlation functions. Taking into account only the transverse components of the fields **v** and $\boldsymbol{\mu}$, we can write

$$\langle h(t,\mathbf{r})p(0,0)\rangle = \int \frac{d\omega\, d^2q}{(2\pi)^3} \exp(-i\omega t + i\mathbf{qr}) G(\omega,\mathbf{q}),$$

$$\langle v_\alpha(t,\mathbf{r})\mu_\beta(0,0)\rangle = \int \frac{d\omega\, d^2q}{(2\pi)^3} \exp(-i\omega t + i\mathbf{qr}) \times\left[\delta_{\alpha\beta} - \frac{q_\alpha q_\beta}{q^2}\right] G_{tr}(\omega,\mathbf{q}), \tag{19}$$

$$\langle h(t,\mathbf{r})h(0,0)\rangle = \int \frac{d\omega\, d^2q}{(2\pi)^3} \exp(-i\omega t + i\mathbf{qr}) D(\omega,\mathbf{q}),$$

$$\langle v_\alpha(t,\mathbf{r})v_\beta(0,0)\rangle = \int \frac{d\omega\, d^2q}{(2\pi)^3} \exp(-i\omega t + i\mathbf{qr}) \times\left[\delta_{\alpha\beta} - \frac{q_\alpha q_\beta}{q^2}\right] D_{tr}(\omega,\mathbf{q}). \tag{20}$$

The correlation functions $\langle pp\rangle$ and $\langle \boldsymbol{\mu}\boldsymbol{\mu}\rangle$ are equal to zero (for the general property of the technique, see, e.g., Ref. 18). The functions D and D_{tr} determine the pair correlation functions of the observable quantities and the functions G and G_{tr} are the response functions. Therefore, the function $G(\omega)$ is analytic in the upper ω half-plane.

It is possible to formulate the diagram technique for calculating the correlation functions (19) and (20). The harmonic part of the effective action $\mathscr{T}=\mathscr{T}_{\rm reac}+\mathscr{T}_{\rm diss}$ determines the bare values of the response functions

$$G_0(\omega,\mathbf{q}) = -\frac{1}{\omega + i\Gamma\kappa q^6/2}, \quad G_{tr,0}(\omega,\mathbf{q}) = -\frac{1}{\rho h\omega + 2i\eta q}. \tag{21}$$

The values of the 'bare' pair correlation functions satisfy the relations

$$\mathrm{Im}\ G = \frac{\kappa q^4}{4T} D, \quad \mathrm{Im}\ G_{tr} = \frac{1}{2T} D_{tr}, \tag{22}$$

which are the consequences of the fluctuation-dissipation theorem. In addition to the harmonic part, the effective action $\mathscr{T}$ contains terms of the third order, which determine the third-order vertices which figure on the diagrams representing the perturbation series for the correlation functions (19) and (20). One can check the relations (22) order by order. Consequently, these relations are valid for the ''dressed'' correlation functions (19) and (20). Note that the relation

$$\int \frac{d\omega}{2\pi} D(\omega,q) = \frac{2T}{\kappa q^4}, \tag{23}$$

which can be proved by using (22), the analyticity of $G(\omega)$ in the upper half-plane, and the asymptotic law $G(\omega) \approx -\omega^{-1}$, which is correct for large ω. Actually, (23) is a direct consequence of (3), since the integral over frequencies is just the simultaneous correlation function.

Analysis of the diagrams shows that they contain infrared logarithms, which are related to the lines representing the correlation function D in Eq. (20). The lines produce the factors

$$(\nabla_\alpha h(t,\mathbf{r})\nabla_\beta h(t,0)) = \frac{TL}{2\pi\kappa}\,\delta_{\alpha\beta}, \tag{24}$$

where $L=\ln[hg^{-1/2}/r]$, and r^{-1} is determined by the characteristic external wave vector of the diagram. The expression (24) can be found from (23) if one recalls condition (5). The presence of the logarithmic contributions implies that the main renormalization of a correlation function like $G(\omega,\mathbf{q})$ is produced by the degrees of freedom with the wave vectors much smaller than q. Therefore, we should extract from the diagrammatic expressions for $G(\omega,\mathbf{q})$ only the contributions corresponding to the interaction with the degrees of freedom.

The program can be realized directly in using the language of the functional integral. Let us separate the variables h, p, $\mathbf{v}$, and $\boldsymbol{\mu}$ into fast parts (with wave vectors larger than q), basic parts (with wave vectors of the order of q), and slow parts (with wave vectors smaller than q). In calculating $G(\omega,\mathbf{q})$ we can forget about the fast parts and keep the interaction of the basic part with the slow part. We then obtain the following expression from (17) and (18):

$$\mathscr{T} = \int dt\, d^2r \left\{ p\partial_t h + pv_\alpha m_\alpha + \mu_\alpha \partial_t j_\alpha - \frac{\kappa}{2}\mu_\alpha \nabla^4 h m_\alpha - \Gamma\,\frac{\kappa}{2}\, p\nabla^6 h + 2\eta\mu\hat{q}\mathbf{v}\right\} + \ldots, \tag{25}$$

where h, p, $\mathbf{v}$, and $\boldsymbol{\mu}$ denote the basic parts of the fields, m_α is the gradient of the slow part of h, and the dots designate the irrelevant terms. The action (25) is of the second order over h, p, $\mathbf{v}$, and $\boldsymbol{\mu}$ and, consequently, the integrals over the fields can be taken explicitly. Since $\mathbf{m}$ varies only weakly along the length q^{-1}, we obtain

$$G(\omega,\mathbf{q}) = -\langle(\rho h\omega + 2i\eta q)\Delta^{-1}\rangle_m, \tag{26}$$

$$G_{tr}(\omega,\mathbf{q}) = -\langle(\omega + i\kappa\Gamma q^6/2)\Delta^{-1}\rangle_m, \tag{27}$$

$$\Delta = (\rho h\omega + 2i\eta q)(\omega + i\kappa\Gamma q^6/2) - \kappa q^4 m_{tr}^2/2, \tag{28}$$

where

$$m_{tr}^2 = \left(\delta_{\alpha\beta} - \frac{q_\alpha q_\beta}{q^2}\right) m_\alpha m_\beta,$$

and the notation $\langle\ldots\rangle_m$ means averaging over the statistics of $\mathbf{m}$. In calculating (26) and (27) we substituted $\mathbf{j}=\rho h\mathbf{v}$. Actually, the terms with ρh can be and we omit them below.

In averaging (26) and (27) we can assume that the statistics of $\mathbf{m}$ are Gaussian. The point is that only the simultaneous correlation functions of $\mathbf{m}$ enter the expressions. These functions are described by the harmonic energy (3). The pair correlation function of $\mathbf{m}$ is equal to (24). Therefore,

$$\langle m_{tr}^2\rangle = \frac{TL}{2\pi\kappa},$$

and we find from (26)

$$G(\omega,\mathbf{q}) = -\int_{-\infty}^{\infty} \frac{d\varsigma}{\sqrt{2\pi}} \exp(-\varsigma^2/2) \times\left(\omega + i\,\frac{\kappa\Gamma}{2}\, q^6 + i\,\frac{TL}{8\pi\eta}\, q^3\varsigma^2\right)^{-1}. \tag{29}$$

We see that G as a function of the frequency ω has the cut along the imaginary semiaxis, which starts from $\omega=-i\Gamma\kappa q^6/2$ and goes to $-i\infty$. The effective length of the cut can be estimated as Tq^3/η, which is the new characteristic frequency associated with the fluctuations. Let us compare the frequency with the position of the pole in the bare expression:

$$\frac{Tq^3/\eta}{\Gamma\kappa q^6}\sim\frac{g}{(qh_0)^3}. \tag{30}$$

We conclude that the fluctuation effects dominate in the region $g^{1/2}<qh_0<g^{1/3}$. We can now justify the disregard of $\rho h\omega$ in comparison with ηq in the above expressions. When $qh\sim 1$,

$$\rho h_0\omega/\eta q\sim\rho\kappa/\eta^2 h_0\sim a/h_0\ll 1,$$

and at $qh_0\sim\sqrt{g}$

$$\rho h_0\omega/\eta q\sim\rho\kappa g^2/\eta^2 h_0\ll 1.$$

Performing Fourier transform of (29) over frequencies, we obtain

$$G(t,\mathbf{q})=i\left(1+\frac{TL}{4\pi\eta}\,q^3 t\right)^{-1/2}\exp\left\{-\frac{\kappa}{2}\,\Gamma q^6 t.\right. \tag{31}$$

The expression (31) is correct for a positive time t. For negative times $G(t)=0$ due to the causality principle since G is the response function. We see from (31) that in the fluctuation region $g^{1/2}<qh<g^{1/3}$ there appears an intermediate power asymptotics $t^{-1/2}$, which at large times t is changed by the exponential decay. This means that the squeezing mode is described by a dynamic equation, which is nonlocal in time.

The above assertion is correct for the wave vectors $q\gtrsim\sqrt{g}/h_0$. In the limit $qh_0\ll\sqrt{g}$ we return to the local equation (11) but with the renormalized kinetic coefficient $\tilde{\Gamma}$. The quantity can be found by integrating the weight $\exp(i\mathscr{S})$ over the degrees of freedom with the wave vectors $q\gtrsim\sqrt{g}/h_0$. The main effect is attributed to the sweeping term in the effective action (17). Because of the integration over the degrees of freedom with the wave vectors $q\gtrsim\sqrt{g}/h$, the term $iT\Gamma(\nabla p)^2$ in (18) for the long-wavelength degrees of freedom is renormalized. We find for the renormalized value

$$\overline{\Gamma}-\Gamma=\frac{1}{4T}\int dt\; d^2r\langle\mathbf{v}(t,\mathbf{r})h(t,\mathbf{r})\mathbf{v}(0,\mathbf{0})h(0,\mathbf{0})\rangle, \tag{32}$$

where averaging is performed over the degrees of freedom with the wave vectors $q\gtrsim\sqrt{g}/h_0$. Using the renormalized expressions for the correlation functions, we obtain the estimate $\tilde{\Gamma}\sim g^{-1/2}\Gamma\gg\Gamma$.

4. CONCLUSIONS

We demonstrated that fluctuations essentially modify the character of the squeezing mode due to its nonlinear coupling with transverse shear hydrodynamic mode. The fluctuation effects lead to nonlocality of the equation for the mode; the corresponding Green's function is (31). The new characteristic frequency of the mode related to the fluctuations is $\omega\sim Tq^3/\eta$ (q is the wave vector); remarkably, it does not depend on the bending elasticity. It is important to distinguish the characteristic frequency from the attenuation of the membrane bending mode (8), which has the same q^3 dependence on the wave vector. We stress that the strong fluctuation effects are observed only for dynamics. The static characteristics are not influenced by fluctuations because of the smallness of the coupling constant (2). This is the reason why we need only the harmonic part of the energy (3).

Strong dynamic fluctuations of h occur for the wave vectors $q\gtrsim\sqrt{g}/h$. For smaller wave vectors the fluctuations of h are weak. Nevertheless, even for the wave vectors there is a memory of the region of strong fluctuations, which is the renormalized value of the kinetic coefficient Γ in Eq. (11): The bare value (12) is substituted by $\tilde{\Gamma}\sim g^{-1/2}\Gamma\gg\Gamma$. Note also that to analyze the dispersion relation of the squeezing mode in the limit $q\ll\sqrt{g}/h$ starting from (11) we should take into account in addition to the energy (5), the steric contribution (4) and the van der Waals (6) contribution to the energy. As a result, we find

$$\omega=-i\tilde{\Gamma}q^2\left(\frac{9\pi^2T^2}{64\kappa h^4}+\frac{10Ha^2}{\pi h^6}\right).$$

Let us discuss the possibility of checking our predictions experimentally. The membranes can be studied by a variety of experimental techniques. Lately, laser «tweezers» have become a useful tool for probing dynamical properties of membranes. This technique enables us to obtain direct information about amplitudes and characteristic times of dynamical fluctuations of different objects consisting of membranes. For details see the monography[27] and recent experiments.[29–31] We can also mention force apparatus measurements,[28] which make it possible to investigate dynamical response for two very thin lamellar systems confined between the walls, and the classical light-scattering experiments. Because of relaxation of the membrane fluctuations, the scattered light has a broadened spectral distribution compared to the incident light. Despite the small broadening, the modern technique of light beating (intensity fluctuation spectroscopy) allows one to obtain information about eigenmodes of the system.

The conclusions concerning the renormalization of the squeezing mode, in our opinion, are interesting, both in their own right and as a new test of the membrane fluctuations.

The research described in this publication was made possible in part by Russian Fund for Fundamental Research grants. One of the authors (E.K.) thanks Max Planck Institute for Physics of Complex Systems (Dresden) for supporting his stay at this institute.

APPENDIX

Elastic modes

Here we consider the elastic modes associated with the relaxation of the surface density of molecules that comprise the two membranes of a double film. To find the dispersion relation for the modes we should start from the elastic energy associated with the variations of the surface density of molecules n_s. In the harmonic approximation the elastic energy of a single membrane is[21]

$$\mathscr{H}_{el}=\frac{1}{2}\int dAB\varsigma^2. \tag{A1}$$

Here

$$\varsigma=(n_s-n_0)/n_0,$$

where n_s-n_0 is the deviation of the surface density of molecules n_s from its equilibrium value n_0, and the coefficient B has the meaning of the inverse compressibility of the membrane. The elastic energy is the sum of terms (A1) for both membranes that constitute the double film.

Let us consider the elastic modes in the linear approximation. We assume that at equilibrium the membranes lie in the planes $z=\pm h_0/2$. The deviations of the membranes from the positions can then be characterized by their displacements $u_{1,2}$ along the z axis. To find the dispersion relations for the modes one should solve conventional hydrodynamic bulk equations supplemented by boundary conditions at the membranes. As we will see, the frequencies of the elastic modes are small compared to sound frequency. Consequently, we can use the convential linearized equations for an incompressible liquid,[32]

$$\nabla_k v_k=0, \quad \left(\frac{\eta}{\rho}\nabla^2-\partial t\right)v_k=\frac{\nabla_k P}{\rho}, \tag{A2}$$

where $k=x,y,z$. Since the membranes are immersed into the liquid, they move with the velocity of the liquid which is continuous near the membranes. The boundary conditions for Eqs. (A2) for a membrane can be found in Refs. 20 and 21. In the linear approximation they are

$$\rho^s\partial_t v_{z\,1,2}+\kappa\nabla_\alpha^4 u_{1,2}=-\lfloor P\rfloor_{1,2}, \tag{A3}$$

$$\partial_t u_{1,2}=v_{z\,1,2}, \tag{A4}$$

$$\rho^s\partial_t v_{\alpha\,1,2}+B\nabla_\alpha\varsigma_{1,2}=\eta\lfloor\nabla_z v_\alpha\rfloor_{1,2}, \tag{A5}$$

$$\partial_t\varsigma_{1,2}+\nabla_\alpha v_{\alpha\,1,2}=0, \tag{A6}$$

where the ''floors'' designate a jump at the membranes, ρ^s is the two-dimensional mass density of amphiphilic molecules, and subscripts 1 and 2 numerate the membranes. The terms with ρ^s in Eqs. (A3) and (A5) are negligible.

Now, we will solve Eqs. (A2) with the boundary conditions (A3)–(A6) under the assumption that all variables are proportional to $\exp(-i\omega t+iqx)$, where ω is the frequency, and q is the wave vector.

The velocity of the liquid is divided into two parts: potential and solenoidal. The potential component is related to the pressure which obeys the equation

$$(\nabla_z^2-q^2)P=0.$$

The solenoidal component is described by the equation

$$-i\rho w v=\eta(\nabla_z^2-q^2)v.$$

Thus we can explicitly write the solutions of the equations inside and outside the film in terms of the velocity of the membranes. The expressions are slightly different for the symmetric ($\varsigma_1=\varsigma_2$) and the antisymmetric ($\varsigma_1=-\varsigma_2$) cases. Using the solutions in bulk, we can express the jumps in Eqs. (A3) and (A5) in terms of v_x and v_z on the membranes. As a result, we find a linear system for v_x and v_z. We can then write the condition for the existence of nontrivial solutions of the system, which for the symmetric case in the simplified form is

$$\begin{aligned}V^2&\left(\coth\frac{Vqh_0}{2}-\coth\frac{qh_0}{2}\right)^2-\left[V^2+(V+\beta)\right.\\&\left.\times\left(V\coth\frac{Vqh_0}{2}-\coth\frac{qh_0}{2}\right)\right]\\&\times\left[V\coth\frac{qh_0}{2}\coth\frac{Vqh_0}{2}\right.\\&\left.+\left(V\coth\frac{Vqh_0}{2}-\coth\frac{qh_0}{2}\right)\right]=0.\end{aligned} \tag{A7}$$

Here we introduce the notation

$$V=\sqrt{-i\frac{\omega\rho}{\eta q^2}}, \quad \beta=i\frac{Bq}{\eta\omega},$$

and suggest that in the case of elastic modes one deals with the frequency $\omega\gg\eta q^2/\rho$. For the antisymmetric case the condition can be obtained from (A7) by substituting coth $\rightarrow$tanh.

It is difficult to find the dispersion relations from (A7) and the analogous equation for the antisymmetric case in a general situation. Below we consider two different limiting cases and assume that

$$\frac{\eta^2}{Bph_0}\ll 1, \tag{A8}$$

which is natural since h_0 is much larger than the molecular length for real films. First, we consider the short-wavelength limit

$$qh_0\gg\left(\frac{\eta^2}{B\rho h_0}\right)^{1/2}. \tag{A9}$$

We then obtain the same dispersion relation as for the elastic mode of a single membrane[33,21]

$$\omega=\frac{\pm\sqrt{3}-i}{2}\left(\frac{B^2}{4\eta\rho}\right)^{1/3}q^{4/3}. \tag{A10}$$

This means that the thickness of layers near the membranes, where the hydrodynamic motion occurs, is much less than h_0 and, consequently, the membranes can be assumed to be nearly independent in this case. Note that due to (A8), the condition (A9) is compatible with $qh_0\ll 1$, where the membranes cannot be regarded as independent in considering, say, the squeezing mode. Therefore, one should be careful: under the condition (A9) the membranes can be treated as nearly independent only in examining the elastic modes. In the opposite long-wavelength limit,

$$qh_0\ll\left(\frac{\eta^2}{B\rho h_0}\right)^{1/2}, \tag{A11}$$

we deal with two different dispersion relations. In the symmetric case the dispersion relation is

$$\omega=\frac{\pm\sqrt{3}-1}{2}\left(\frac{B^2}{\eta\rho}\right)^{1/3}q^{4/3}. \tag{A12}$$

This is the same dispersion relation as (A10) but with the doubled membrane elasticity, which is natural for the double film. In the antisymmetric case the dispersion relation is

$$\omega=-i\,\frac{Bq^2h_0}{2\eta}. \tag{A13}$$

Thus we encounter a simple diffusion.

The dispersion laws (A10), (A12), and (A13) show that the frequencies of the elastic modes are small in comparison with the sound frequency cq (where c is the sound velocity), which justifies our using the incompressible hydrodynamic equations (A2). Note also that for the mode (A13) the condition (A8) ensures the inequality $\omega\gg\eta q^2/\rho$, which was suggested in the derivation of the relation (A7) (the inequality enables us to disregard the potential part of the velocity). Thus, the condition (A8) makes our scheme self-consistent.

In the long-wavelength limit the double membrane film can be treated as an effective single membrane. This effective membrane should be framed by hydrodynamic variables, which give the information about the ''microscopic'' construction of the double membrane film. In other words, one should incorporate into the set of ''macroscopic'' variables the surface densities (described by ς_1 and ς_2) of molecules, which comprise the two membranes, and the two-dimensional mass density ρh of the liquid between the membranes. The dynamic equation for the variable has been derived in the main text of this paper [see Eq. (11)]. The phenomenological dynamic equations for ς_1 and ς_2 can be derived in the same manner as the Poisson brackets for the density of any conserved scalar quantity, the same structure of the Poisson bracket as (10).[18] In terms of the variables $\varsigma_+=(\varsigma_1+\varsigma_2)/2$ and $\varsigma_-=\varsigma_1-\varsigma_2$ the equations are

$$\partial_t\varsigma_+=-\nabla_\alpha[(1+\varsigma_+)v_\alpha]+D_+\nabla^2\varsigma_+, \tag{A14}$$

$$\partial_t\varsigma_-=-\nabla_\alpha(\varsigma-v_\alpha)+D_-\nabla^2\varsigma_-. \tag{A15}$$

Here we discarded the bending motion. The system of equations (A14) and (A15) should be supplemented by Eq. (14) in the main text of the paper, where

$$P_s=2B\varsigma_+. \tag{A16}$$

In analyzing the elastic degrees of freedom we should use the following expression for the momentum flow from the bulk:

$$J_\alpha=-2\sqrt{i\eta\rho\hat{\omega}}v_\alpha, \tag{A17}$$

where the velocity v_α is implied to be longitudinal, since just the longitudinal component of the velocity is involved in the elastic motion. We stress that the expression (A17) is correct if $\omega\gg\eta q^2/\rho$, in contrast with the applicability condition of (15).

Linearizing the system of equations (14), (A14), and (A15) (and ignoring the squeezing degrees of freedom), we find the dispersion relations (A12) and $\omega=-iD_-q^2$. The term with D_+ in (A14) appears to be irrelevant. Comparing the dispersion relation $\omega=-iD_-q^2$ with (A13), we obtain

$$D_-=Bh_0/2\eta. \tag{A18}$$

It can then be verified that the linear coupling between the elastic and the squeezing degrees of freedom described by the term with κ in (14) is negligible. The nonlinear terms in Eqs. (14), (A14), and (A15) lead to the interaction of different modes. Explicit analysis shows that the fluctuation effects do not affect appreciably the linear dispersion relations (A10), (A12), and (A13) due to $g\ll1$. The same holds for a nonlinear interaction with the bending degree of freedom, which (because of the same inequality $g\ll1$) does not change the results obtained in the linear approximation.

*)E-mail: kats@landau.ac.ru

[1] *Physics of Amphiphilic Layers*, ed. by J. Meuner, D. Langevin, and N. Boccara, Springer Proceedings in Physics, Vol. 21, Springer-Verlag, Berlin (1987).
[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. Bellocq, J. Piaif, and P. Bothorel, Adv. Colloid Interface Sci. **20**, 167 (1984).
[5] G. Porte, J. Appel, P. Bassereau, J. Marignan, M. Scouri, I. Billard, and M. Delsanti, Physica A **176**, 168 (1991).
[6] G. Porte, J. Phys. II (Condens. Matter) **4**, 8649 (1992).
[7] P. B. Canham, J. Theor. Biol. **26**, 61 (1970).
[8] W. Helfrich, Z. Naturforsch. B **103**, 67 (1975).
[9] P. G. de Gennes and C. Taupin, J. Chem. Phys. **86**, 2294 (1982).
[10] W. Helfrich, J. Phys. (Paris) **46**, 1263 (1985).
[11] D. Förster, Phys. Lett. A **114**, 115 (1986).
[12] L. Peliti and S. Leibler, Phys. Rev. Lett. **54**, 1690 (1985).
[13] H. Kleinert, Phys. Lett. A **114**, 263 (1986).
[14] A. M. Polyakov, Nucl. Phys. B **268**, 406 (1986).
[15] W. Helfrich, Z. Naturforsch. A **33**, 305 (1978).
[16] J. N. Israelashvili, *Intermolecular and Surface forces*, Academic Press, Orlando (1985).
[17] F. Brochard and M. Lennon, J. Phys. Lett. (Paris) **36**, 1035 (1975).
[18] E. I. Kats and V. V. Lebedev, *Fluctuational Effects in the Dynamics of Liquid Crystals*, Springer-Verlag, New York (1993).
[19] I. E. Dzyaloshinskii and G. E. Volovik, Ann. Phys. **125**, 67 (1980).
[20] V. V. Lebedev and A. R. Muratov, Zh. Éksp. Teor. Fiz. **95**, 1751 (1989) [Sov. Phys. JETP **68**, 1011 (1989)].
[21] E. I. Kats and V. V. Lebedev, Phys. Rev. E **49**, 3003 (1994).
[22] H. W. Wyld, Ann. Phys. **14**, 143 (1961).
[23] P. C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).
[24] S. K. Ma, *Modern Theory of Critical Phenomena*, Benjamin, New York (1976).
[25] C. de Dominicis, J. Phys. (Paris), Colloq. **37**, C1-247 (1976).
[26] H. K. Janssen, Z. Phys. B **23**, 377 (1976).
[27] S. M. Block, *Noninvasive Techniques in Cell Biology*, Wiley-Liss, New York (1990).
[28] P. Richetti, L. Moreau, P. Barois, and P. Kekicheff, Phys. Rev. E **54**, 1749 (1996).
[29] R. Bar-Ziv and E. Moses, Phys. Rev. Lett. **73**, 1392 (1994).
[30] R. Bar-Ziv, R. Menes, E. Moses, and S. A. Safran, Phys. Rev. Lett. **75**, 3356 (1995).
[31] R. Bar-Ziv, T. Frisch, and E. Moses, Phys. Rev. Lett. **75**, 3481 (1995).
[32] L. D. Landau and E. M. Lifshitz, *Fluids Mechanics*, Pergamon, New York (1988).
[33] E. I. Kats and V. V. Lebedev, Zh. Éksp. Teor. Fiz. **94**(5), 134 (1988) [Sov. Phys. JETP **67**, 940 (1988)].

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

Fluid flow through three-dimensional fibrous porous media

A. L. Chernyakov*)

''Kurchatov Institute'' Russian Science Center, 123182 Moscow, Russia
(Submitted 22 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 2109–2128 (June 1998)

A system of self-consistent equations for determining the hydrodynamic resistance of dilute fibrous porous media in the case of arbitrary low Reynolds numbers and arbitrary random packing of the fibers in the media is derived on the basis of a multiple-scattering hydrodynamic theory. The equations obtained are applied to the case of isotropic packing of the fibers and to the anisotropic case when all the fibers are orthogonal to the direction of fluid flow. Equations are derived and analyzed for the velocity correlation function in a random fibrous medium. The longitudinal and transverse diffusion coefficients of a passive impurity embedded in the fluid are calculated. [S1063-7761(98)01506-6]

1. INTRODUCTION

An extensive theoretical literature has been devoted to various aspects of the hydrodynamics of granular media and suspensions. We call attention to recent papers and surveys, which offer detailed bibliographies on the subject.[1–9] In particular, the hydrodynamic resistance of a porous medium formed by rigid spheres has been calculated by the ensemble-average approach.[10,11] Howells[10] also determined the resistance of a porous medium consisting of randomly distributed parallel cylindrical fibers. However, very few theoretical and computational papers have been published on three-dimensional media formed by long fibers or, in particular, fibrous filters, despite the importance of this problem.[12,13] Apart from a phenomenological study[14] based on the Brinkman equation,[15] only a few recent papers have reported investigations of three-dimensional fluid flows through porous media on the basis of the Navier–Stokes equations. The resistance of various three-dimensional periodic arrays of infinitely long fibers as a function of the density of the medium was first investigated in Ref. 16, which included a calculation of the resistances of square and hexagonal arrays, as well as a review of theoretical papers on the determination of the resistance of two-dimensional periodic and random fibrous media. The use of the variational principle to calculate the hydrodynamic resistance of three-dimensional periodic arrays of fibers was discussed in Ref. 17. The influence of the relative position and orientation of fibers in a three-dimensional array on the hydrodynamic resistance has been investigated analytically.[18]

In this paper we investigate the averaged characteristics of the hydrodynamic flow of an incompressible fluid through a porous medium formed by randomly distributed, randomly oriented, infinitely long cylindrical fibers of various diameters. Proceeding from the Navier–Stokes equations, we derive self-consistent field equations, which lead to the Brinkman equation in the long-wavelength approximation. We investigate the viscosity renormalization properties, calculate the velocity correlation function, and determine the diffusion coefficients of a passive impurity in the fluid.

2. DESCRIPTION OF THE MODEL OF A FIBROUS POROUS MEDIUM

Our model of a porous medium is a three-dimensional space filled with randomly distributed, arbitrarily oriented, infinitely long, polydisperse cylindrical fibers. For a mathematical description of such a fibrous medium it is convenient to choose an arbitrary plane containing the average velocity vector, wherein the position of each fiber of radius a_α is specified by a vector $\boldsymbol{\rho}_\alpha$, which characterizes the point of intersection of the fiber with the plane, and by the unit tangent vector $\mathbf{n}_\alpha$, as shown in Fig. 1 for a single fiber. Below the x axis is directed along the average fluid flow velocity, and the z axis along a normal to the plane selected. The statistical ensemble of the model porous media described here, i.e., the ensemble over which averaging is performed, is defined by the combined N-particle distribution functions

$$f_N = f_N(\boldsymbol{\rho}_1, \mathbf{n}_1, a_1; \boldsymbol{\rho}_2, \mathbf{n}_2, a_2; \ldots; \boldsymbol{\rho}_N, \mathbf{n}_N, a_N).$$

In addition to the distribution functions, it is also necessary to specify in the space of straight lines a measure over which to perform averaging. If the unit vector is specified by the spherical angles θ and ϕ, the unique (to within a constant factor) measure which is invariant under a group of motions of the three-dimensional space E_3 is written in the form[19]

$$d\chi = \sin\theta \cos\theta \, d\theta \, d\phi \, dx \, dy. \tag{1}$$

All averaging will henceforth be performed over this invariant measure. Often it is useful to express distribution functions in terms of correlation functions.[20] In particular, for $N = 1,2,3$ we can write

$$f_1(t_1) = g_1(t_1), \quad f_2(t_1,t_2) = g_1(t_1)g_1(t_2) + g_2(t_1,t_2),$$

$$f_3(t_1,t_2,t_3) = g_1(t_1)g_1(t_2)g_1(t_3) + g_1(t_1)g_2(t_2,t_3) + g_1(t_2)g_2(t_1,t_3) + g_1(t_3)g_2(t_1,t_2) + g_3(t_1,t_2,t_3),$$

where t represents all the variables characterizing each fiber.

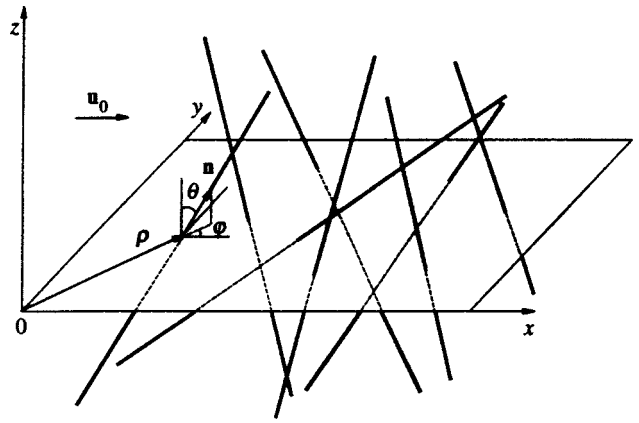


FIG. 1. Model of a porous fibrous medium.

3. THE SINGLE-FIBER SCATTERING OPERATOR

To determine the hydrodynamic resistance of the porous medium investigated, we use the Navier–Stokes equations for an incompressible fluid in the Oseen approximation[21]:

$$\rho(\mathbf{U}_0\cdot\nabla)\mathbf{V}=\mu\nabla^2\mathbf{V}-\nabla p,\quad \text{div } \mathbf{V}=0, \tag{2}$$

which is valid at the low Reynolds numbers $\mathrm{Re}=U_0a/\nu \ll 1$, where ρ is the density of the fluid, U_0 is the average velocity, a is the radius of the fiber, and μ and ν are the dynamic and kinematic viscosities, respectively. These equations must be solved with satisfaction of the boundary conditions. In particular, if slip can be disregarded, the velocity at each fiber must be equal to zero. Before solving Eqs. (2) for a random configuration of fibers, we proceed by analogy with Refs. 22–25 and define the scattering operator for a solitary fiber embedded in an arbitrary external flow, which is regular at this site.

Any fluid flow can be represented by the Fourier expansion

$$\mathbf{V}(\mathbf{r})=\int \mathbf{V}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}. \tag{3}$$

When a fiber is embedded in an assigned flow, forces appear on its surface and act on the fluid in such a way as to alter the flow field and satisfy the boundary conditions. The induced forces depend linearly on the freestream velocity. A linear relation for an arbitrary αth fiber can be written for the Fourier transforms of the resulting forces and the freestream flow field:

$$F_i^{\alpha}(k')=\mu\int T_{ij}^{\alpha}(k',k)V_j(k)\frac{d^3k}{(2\pi)^3}. \tag{4}$$

In papers on hydrodynamics the matrix $T_{ij}^{\alpha}(k',k)$ is usually called the friction coefficient matrix,[22] but in this paper we prefer, by analogy with wave problems, to call it the scattering operator for the αth fiber.

To calculate the scattering operator $\hat{T}$, we initially choose a coordinate system such that the z axis coincides with the axis of the fiber. We represent the freestream velocity on the surface of a fiber of radius a by the expansion

$$\mathbf{V}(z,\phi)=\sum_m\int \mathbf{V}_m(k_3)\exp(ik_3z+im\phi)\frac{dk_3}{2\pi}. \tag{5}$$

The Fourier surface harmonics $\mathbf{V}_m(k_3)$ are expressed in terms of the coefficients of the three-dimensional Fourier integral for the velocity:

$$\mathbf{V}_m(k_3')=\int \mathbf{V}(\mathbf{k})i^mJ_m(k_\perp a)e^{-im\psi}2\pi\delta(k_3'-k_3)\frac{d^3k}{(2\pi)^3}, \tag{6}$$

where J_m is the Bessel function of order m, and $(k_\perp,\psi,k_3)$ denotes the cylindrical components of the vector $\mathbf{k}$. Assigning the distribution of the surface forces in the form

$$\mathbf{F}(\mathbf{r})=\sum_n\int \mathbf{F}_n(k_3)\exp(ik_3z+in\phi)\delta(r-a)\frac{dk_3}{2\pi a}, \tag{7}$$

we can calculate the three-dimensional Fourier spectrum of forces exerted by the fiber on the fluid:

$$\mathbf{F}(\mathbf{k})=\sum_n(-i)^n2\pi J_n(k_\perp a)e^{in\phi}\mathbf{F}_n(k_3). \tag{8}$$

Knowing the spectrum of surface forces, we can find the flow field $\mathbf{V}_s$ induced by these forces:

$$\mathbf{V}_s(\mathbf{r})=\int \hat{G}_0(\mathbf{k})\mathbf{F}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}, \tag{9}$$

where $\hat{G}_0(\mathbf{k})$ denotes the bare Green's function for the free fluid:

$$\hat{G}_0(\mathbf{k})=\frac{\hat{\Pi}(\mathbf{k})}{k^2+\mathbf{k}\cdot\mathbf{U}_0/\nu}. \tag{10}$$

Here $\hat{\Pi}(\mathbf{k})$ is the projection operator, which can be written in the component form

$$\Pi_{ij}=\delta_{ij}-\frac{k_ik_j}{k^2}. \tag{11}$$

We assume here that the total velocity field is continued continuously inside the fiber, where it vanishes.[22,23] Expanding the velocity field $\mathbf{V}_s(\mathbf{r})$ on the fiber surface into a Fourier series, we obtain

$$V_{sm}(k_3)=i^{m-n}K_{ijmn}(k_3)F_{jn}(k_3), \tag{12}$$

where the matrix $K_{ijmn}(k_3)$ is given by the equality

$$K_{ijmn}(k_3)=\int G_{ij}(k_\perp,k_3)J_m(k_\perp a)J_n(k_\perp a)\times\exp[i(n-m)\psi]\frac{k_\perp dk_\perp d\psi}{(2\pi)^2}. \tag{13}$$

Setting the total fluid velocity $\mathbf{V}+\mathbf{V}_s$ on the fiber surface equal to zero, we obtain a system of equations for all the surface-force harmonics:

$$i^{m-n}K_{ijmn}(k_3)F_{jn}=-V_{im}. \tag{14}$$

If we introduce the reciprocal matrix

$$K_{ijmn}^{-1}K_{jknt}=\delta_{ik}\delta_{mt}, \tag{15}$$

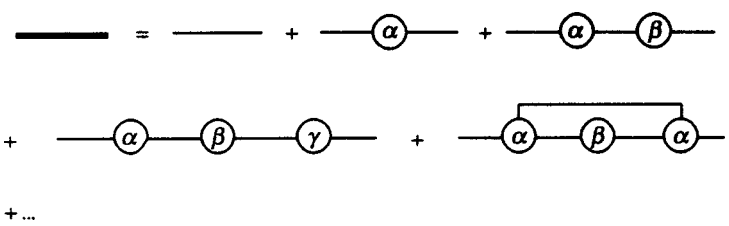


FIG. 2. Multiple-scattering series for the Green's function $\hat{G}$.

we can express the force harmonics in terms of the freestream velocity and derive an expression for the single-fiber scattering operator:

$$T_{ij}(\mathbf{k}',\mathbf{k}) = -2\pi\delta(k_3'-k_3)T_{ij}(k_3,k_\perp',k_\perp)$$

$$= -2\pi\delta(k_3'-k_3)\sum_{n,m} K_{ijnm}^{-1} J_n(k_\perp' a)J_m(k_\perp a)$$

$$\times \exp(in\psi' - im\psi). \tag{16}$$

In the general case where the αth fiber intersects the plane selected at the point specified by the vector $\boldsymbol{\rho}_\alpha$ and is directed along the vector $\mathbf{n}_\alpha$, it is readily shown that the scattering operator has the form

$$\hat{T}^\alpha(\mathbf{k}',\mathbf{k}) = -2\pi\delta[\mathbf{n}_\alpha\cdot(\mathbf{k}'-\mathbf{k})]\exp[i(\mathbf{k}'-\mathbf{k})\boldsymbol{\rho}_\alpha]$$

$$\times \sum_{n,m} \hat{K}_{nm}^{-1} J_n(k'^\alpha_\perp a_\alpha)J_m(k_\perp^\alpha a_\alpha)$$

$$\times \exp(in\psi' - im\psi), \tag{17}$$

where $k_\perp^\alpha$ is the transverse part of $\mathbf{k}$ with respect to the αth fiber:

$$k_\perp^\alpha = |\mathbf{k} - \mathbf{n}_\alpha(\mathbf{n}_\alpha\cdot\mathbf{k})|.$$

When the condition $k_\perp^2 a^2 \ll 1$ holds for small transverse harmonics, only the term with $n=m=0$ can be retained in the sum in Eqs. (16) and (17), and the Bessel function J_0 can be replaced by unity. In this case the expression (16) for the scattering operator assumes the simple form $\hat{T} = -(2\pi)\delta(k_3'-k_3)\hat{K}_{00}^{-1}$. Neglecting terms $\sim k_\perp^2 a^2$ in the integral used to calculate $\hat{K}_{00}$, we can also assume to the same accuracy that one of the Bessel functions is equal to unity. In this case the expression for $\hat{K}_{00}$ can be reduced to the form

$$\hat{K}_{00} = \int J_0(k_\perp a)\hat{G}(k_\perp,k_3)\frac{d^2k}{(2\pi)^2}$$

$$= \left\langle \int e^{i\mathbf{k}_\perp\cdot\mathbf{r}}\hat{G}(\mathbf{k})\frac{d^2k}{(2\pi)^2}\right\rangle_a, \tag{18}$$

where the angle brackets signify averaging over the surface of the fiber for $r=a$. We therefore see that under the condition $k_\perp^2 a^2 \ll 1$ the general equation for the scattering operator reduces to a result that can be obtained by the Burgers method.[18,26,27]

4. MULTIPLE-SCATTERING HYDRODYNAMIC THEORY

Knowing the hydrodynamic scattering operator, we can use power series in the scattering multiplicity, as is customary in problems concerning the multiple scattering of waves in randomly inhomogeneous media,[28] to calculate the averaged characteristics of fluid flow in a porous medium. Reference 29 is probably the first paper in which attention is called to the analogy between the hydrodynamics of granular media and the theory of the multiple scattering of waves. This analogy has since been used repeatedly in calculations of the viscosity of polymer solutions[30] and various hydrodynamic properties of suspensions consisting of rigid spherical particles[22–25] and rods.[31] For our investigation of the hydrodynamics in fibrous porous media, we use the diagram technique developed for multiple-scattering theory, which was described in detail in Frisch's review.[32] In particular, the Green's function of the Navier–Stokes equations for porous media, which specifies the flow velocity of a fluid in a medium when a delta-shaped force is applied to it, can be written as the sum of sequences of scattering processes on various fibers. A diagram can be associated with each sequence of scattering processes. We note that a diagram representation of a sequence of interaction processes was used to calculate the friction force by ''operator transfer'' in Ref. 33. The diagrams for the Green's function prior to averaging over the fiber positions and orientations are shown in Fig. 2. The total Green's function is represented by a heavy line, the bare Green's function by a thin horizontal line, and the operator for scattering by the αth fiber is represented by a circle containing α. The overhead line shown in the last diagram connects identically numbered fibers. All possible sequences of scattering processes contribute to the Green's function. After averaging with distribution functions expressed in terms of the correlation functions, we obtain the Dyson equation shown in Fig. 3 for the Greens' function. In the diagrams for the self-energy operator Σ the correlation function is denoted by a dashed line. After averaging, each independent fiber introduces into the diagram a multiplier n, which denotes the density of the points of intersection of the fiber axes with an arbitrarily selected plane. The series constructed usually lead to divergent expressions, because the bare Green's functions are used.[6,32] The divergence is usually removed in hydrodynamic problems by segregating the class

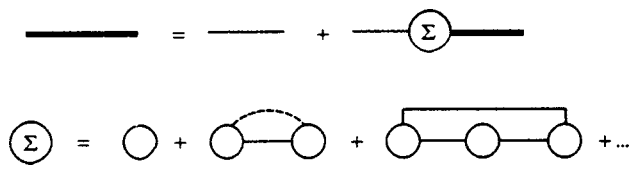


FIG. 3. Dyson equation for the Green's function $\hat{G}$ and series representation of the self-energy $\hat{\Sigma}$.

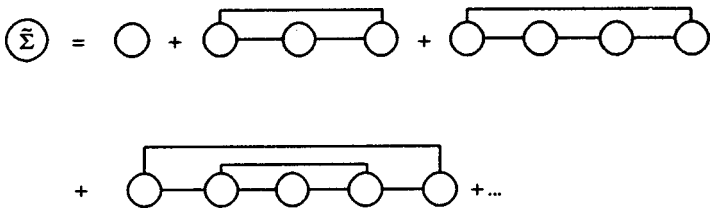


FIG. 4. Diagrams specifying the self-consistent self-energy operator.

of diagrams for the self-energy operator Σ and summing them.[31,34] We introduce the self-consistent operator $\tilde{\Sigma}$, to which contributions are made from all graphs without correlation functions and diagrams without any mutually intersecting overhead lines, which indicate identical fibers, as shown in Fig. 4. Diagrams of the type indicated in Fig. 5 do not appear in the definition of $\tilde{\Sigma}$. If we introduce a self-consistent Green's function satisfying the equation

$$\tilde{G}=G_0+G_0\tilde{\Sigma}\tilde{G}, \tag{19}$$

and the self-consistent operator for scattering on an individual fiber $\tilde{T}^\alpha$ defined by the equation

$$\tilde{T}^\alpha=T^\alpha+T^\alpha G_0\tilde{\Sigma}\tilde{G}\tilde{T}^\alpha, \tag{20}$$

we obtain a simple expression for the self-energy operator $\tilde{\Sigma}$ in terms of the self-consistent scattering operator:

$$\tilde{\Sigma}=\sum_\alpha \langle \tilde{T}^\alpha\rangle, \tag{21}$$

where the angle brackets denote averaging over the position and orientation of each fiber. The system of self-consistent equations (19)–(21) specifying the scattering operator and the Green's function is shown in graphical form in Fig. 6. All diagrams for the total self-energy operator Σ can be expressed in terms of the self-consistent quantities introduced above, as is shown in Fig. 7.

The Bethe–Salpeter equation for the averaged two-particle Green's function $\hat{K}$, which characterizes the velocity correlation equation[32] can be derived analogously. The Bethe–Salpeter equation and the series for its kernel are shown in graphical form in Fig. 8. The vertical lines in the diagrams for the kernel of the Bethe–Salpeter equation connect identically numbered fibers.

5. SELF-CONSISTENT THEORY OF FLUID FLOW IN FIBROUS MEDIA

We first investigate the system of self-consistent equations for a homogeneous and isotropic porous medium. The distribution function with respect to the directions of the fibers $f(\mathbf{n})$ is a constant equal to $1/\pi$ in this case. We confine the present study to the thin-fiber approximation, $ka\ll 1$. The bare and self-consistent scattering operators then depend only on the longitudinal component of $\mathbf{k}$ and can be written in the coordinate system associated with a fiber in the form

$$T_{ij}=2\pi\delta(k_3'-k_3)T_{ij}(k_3), \tag{22}$$

$$\tilde{T}_{ij}=2\pi\delta(k_3'-k_3)\tilde{T}_{ij}(k_3). \tag{23}$$

We note that upon averaging over the fiber positions in the plane selected, the scattering operator becomes proportional to $\delta(\mathbf{k}_\perp'-\mathbf{k}_\perp)$. It is evident from Eq. (17) that the definition of the scattering operator itself contains the delta function $\delta[\mathbf{n}\cdot(\mathbf{k}'-\mathbf{k})]$ of vector components which are longitudinal relative to the fiber. The product of these delta functions can be written in the form

$$\delta(\mathbf{k}_\perp'-\mathbf{k}_\perp)\,\delta[\mathbf{n}\cdot(\mathbf{k}'-\mathbf{k})]=\delta(\mathbf{k}'-\mathbf{k})\frac{1}{\cos\theta},$$

where θ is the spherical angle of the unit vector $\mathbf{n}$ in the coordinate system which we have chosen. Averaging over the measure (1) now reduces to ordinary spherical integration. Bearing this remark in mind, we can write the self-consistent field equations for the scattering operator and the Green's function in Fig. 6 in this approximation as follows:

$$\tilde{T}_{ij}(k_3)=T_{ij}(k_3)+T_{il}(k_3)\int(\tilde{G}_{lm}-G_{0lm})\times\frac{d^2k}{(2\pi)^2}\,\tilde{T}_{mj}(k_3), \tag{24}$$

$$\tilde{G}_{ij}^{-1}(\mathbf{k})=G_{0ij}^{-1}(\mathbf{k})-\tilde{\Sigma}_{ij}(\mathbf{k}), \tag{25}$$

$$\tilde{\Sigma}_{ij}(\mathbf{k})=n\langle\tilde{T}_{ij}(\mathbf{k}\cdot\mathbf{n})\rangle. \tag{26}$$

In the last equation the averaging is carried out over all fiber orientations with the distribution function $f(\mathbf{n})$. For the inverse of the scattering operator we obtain the following equation from (24):

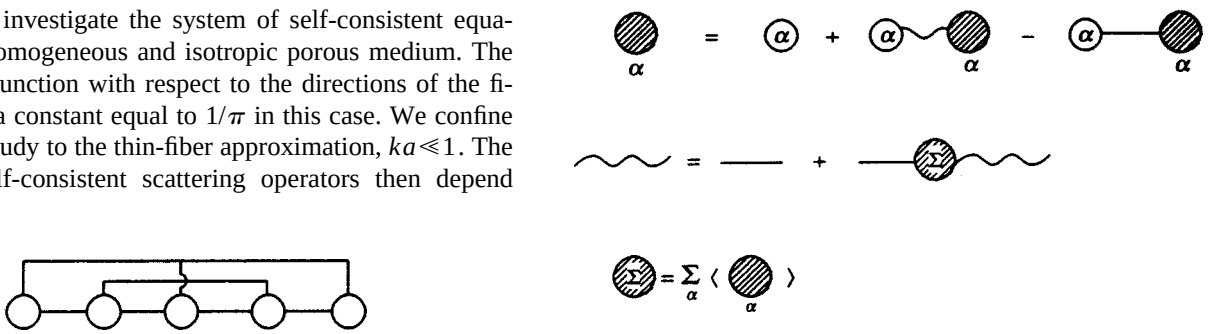


FIG. 6. System of self-consistent equations for the Green's function and the renormalized scattering operator.

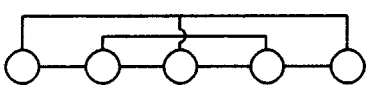


FIG. 5. Example of a diagram that does not contribute to the operator $\tilde{\Sigma}$.

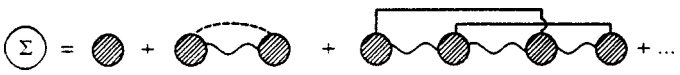


FIG. 7. Diagram representation of the self-energy operator in terms of the self-consistent Green's function and the renormalized scattering operator.

$$\tilde{T}_{ij}^{-1}(k_3)=T_{ij}^{-1}(k_3)-\int(\tilde{G}_{ij}-G_{0ij})\frac{d^2k}{(2\pi)^2}.$$

It is evident from this equation that the renormalized scattering operator is expressed in terms of the self-consistent Green's function in the same way as the bare operator is expressed in terms of the bare Green's function (18):

$$\tilde{T}_{ij}^{-1}(k_3)=-\left\langle\int\exp(i\mathbf{k}_\perp\cdot\mathbf{r})\tilde{G}_{ij}\frac{d^2k}{(2\pi)^2}\right\rangle_a. \tag{27}$$

Numerical methods must be used for the exact solution of these integral equations. We therefore consider several approximations that can be used to obtain simple analytical results. We first calculate the scattering operator, specifying the self-energy in the form

$$\tilde{\Sigma}_{ij}(\mathbf{k})=-\Pi_{ij}(\mathbf{k})\kappa^2(k^2). \tag{28}$$

If we restrict the discussion to the long-wavelength approximation and regard the function κ as a constant, this form of the self-energy operator corresponds to the situation of averaged fluid flows satisfying the Brinkman equations[15] with a tensor of Darcy friction coefficients proportional to the unit tensor. The Green's function then has the form

$$\tilde{G}_{ij}=\Pi_{ij}(\mathbf{k})\frac{1}{k^2+2i\mathbf{k}\cdot\mathbf{k}_*+\kappa^2}, \tag{29}$$

where we have introduced the notation $\mathbf{k}_*=\mathbf{U}_0/2\nu$. Substituting this expression for the Green's function into Eq. (27), we obtain an integral representation for the scattering operator. The scattering operator is calculated in the Appendix.

We confine the analysis of an isotropic medium to a high packing density, so that the inertial effects can be disregarded. As is shown in the Appendix, the scattering operator in an arbitrary coordinate system has the form (58) in this case:

$$\tilde{T}_{ij}=T_0\left[\delta_{ij}-\frac{1}{2}n_in_j\right],$$

where

$$T_0=4\pi\left[\ln\frac{\gamma a((\mathbf{k}\cdot\mathbf{n})^2+\kappa^2)^{1/2}}{2}-\frac{1}{2}\frac{(\mathbf{k}\cdot\mathbf{n})^2}{\kappa^2}\ln\frac{(\mathbf{k}\cdot\mathbf{n})^2+\kappa^2}{(\mathbf{k}\cdot\mathbf{n})^2}\right]^{-1}.$$

On large scales, where the condition $k^2/\kappa^2\ll1$ holds, we can write

$$T_0=\frac{4\pi}{\ln(\gamma a\kappa/2)}\left[1-\frac{1}{2}\frac{(\mathbf{k}\cdot\mathbf{n})^2}{\kappa^2\ln(\gamma a\kappa/2)}\ln\frac{(\mathbf{k}\cdot\mathbf{n})^2}{\kappa^2}-\frac{1}{2}\frac{(\mathbf{k}\cdot\mathbf{n})^2}{\kappa^2\ln(\gamma a\kappa/2)}\right].$$

Now, averaging the tensor $\tilde{T}_{ij}$ over all directions of the vector $\mathbf{n}$ and separating the transverse part, we obtain the self-energy equation

$$\tilde{\Sigma}_{ij}=\Pi_{ij}\frac{16\pi n}{\ln(\gamma a\kappa/2)}\left\{\frac{5}{6}-\frac{17}{300}\frac{k^2}{\kappa^2\ln(\gamma a\kappa/2)}-\frac{3}{20}\frac{k^2\ln(k^2/\kappa^2)}{\kappa^2\ln(\gamma a\kappa/2)}\right\}. \tag{30}$$

Comparing this equation with Eq. (28), we obtain the self-consistency condition

$$\kappa^2=-\frac{5}{6}n\frac{16\pi}{\ln(\gamma a\kappa/2)}. \tag{31}$$

If the material is polydisperse, the right-hand side of the latter equation must still be averaged over the fiber diameters. The wavenumber-dependent part of the self-energy can be interpreted as the decrease in the friction coefficient in the Brinkman equation as the characteristic flow dimensions decrease. The expression (31) for the self-energy is valid for infinitely long fibers or, as shown in the Appendix, for $kl\gg1$, if the finite fiber length $2l$ is taken into account. In the large-scale domain, $kl\ll1$, the expression derived in the Appendix for the zero-angle scattering operator of a finite cylinder can be used to obtain an equation for the self-energy in this case:

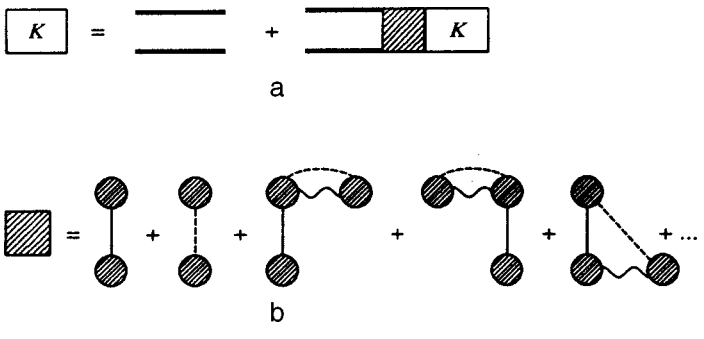


FIG. 8. a) Bethe–Salpeter equation for the two-particle Green's function; b) series representation of the kernel of the Bethe–Salpeter equation.

$$\tilde{\Sigma}_{ij}=\Pi_{ij}\frac{4\alpha}{a^2\ln(\gamma a\kappa/2)}\left\{\frac{5}{6}+\frac{3}{10}\frac{k^2\ln(2\kappa l/\pi)}{\kappa^2\ln(\gamma\kappa a/2)}\right\}, \quad (32)$$

where α denotes the volume fraction of the fibers in the medium. The quadratic term in this equation is often interpreted as renormalization of the viscosity. If this interpretation is applied to the porous media under investigation, as opposed to suspensions,[21,25,31] the viscosity decreases as the packing density increases.[35]

We now consider an anisotropic porous medium consisting of fibers oriented perpendicularly to a specific preferred direction. We assume that the fluid flows in this direction on the average. Such a medium is often treated as a model of fibrous filters.[14] The angular distribution function for this model is

$$f(\mathbf{n})=\frac{\pi}{2\ \sin\ \theta}\left[\delta\left(\phi-\frac{\pi}{2}\right)+\delta\left(\phi+\frac{\pi}{2}\right)\right], \quad (33)$$

where the angle ϕ is measured from the preferred direction. For a uniaxial anisotropic medium we seek the self-energy in the form

$$\tilde{\Sigma}_{ij}(\mathbf{k})=-\Pi_{ij}(\mathbf{k})\kappa_\perp^2-\Pi_{i1}\Pi_{1j}(\kappa_\parallel^2-\kappa_\perp^2). \quad (34)$$

Here we have taken into account that the porous medium in our model has a preferred direction coinciding with the x axis and that the tensor $\tilde{\Sigma}_{ij}$ must be transverse relative to the vector $\mathbf{k}$:

$$k_i\tilde{\Sigma}_{ij}=\tilde{\Sigma}_{ij}k_j=0.$$

The Green's function now has the form

$$\tilde{G}_{ij}=\left[\Pi_{ij}(\mathbf{k})-\frac{\Pi_{i1}\Pi_{1j}(\kappa_\parallel^2-\kappa_\perp^2)}{k^2+2i\mathbf{k}\cdot\mathbf{k}_*+(\kappa_\parallel^2-\kappa_\perp^2)\Pi_{11}}\right]\times\frac{1}{k^2+2i\mathbf{k}\cdot\mathbf{k}_*+\kappa_\perp^2}. \quad (35)$$

The Green's function contains the isotropic term (29) with κ replaced by $\kappa_\perp$ and an anisotropic term proportional to the difference $(\kappa_\parallel^2-\kappa_\perp^2)$. Substituting this expression for the Green's function into (27), we obtain an equation for the renormalized scattering operator in an anisotropic medium. The scattering operator for an isotropic medium has already been calculated. We analyze the contribution of the anisotropic part of the Green's function to the scattering operator for $k_3=0$ and the low fluid flow velocities $k_*\ll\kappa_\perp$. Evaluating the simple integrals, we obtain a diagonal scattering operator with the following matrix elements in the coordinate system associated with a fiber:

$$\tilde{T}_{11}=4\pi\left[\ln\frac{\gamma a}{4}(\kappa_\perp+\kappa_\parallel)+\frac{\kappa_\parallel}{\kappa_\perp+\kappa_\parallel}-\frac{1}{2}\right]^{-1},$$

$$\tilde{T}_{22}=4\pi\left[\ln\frac{\gamma a}{4}(\kappa_\perp+\kappa_\parallel)-\frac{\kappa_\parallel}{\kappa_\perp+\kappa_\parallel}+\frac{1}{2}\right]^{-1},$$

$$\tilde{T}_{33}=2\pi\left[\ln\frac{\gamma a}{2}\kappa_\perp\right]^{-1}.$$

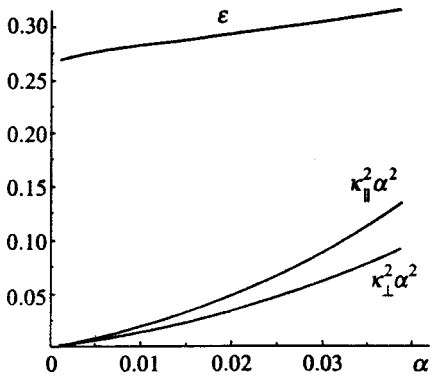


FIG. 9. Longitudinal ($\kappa_\parallel^2$) and transverse ($\kappa_\perp^2$) friction coefficients and their anisotropy ϵ versus the packing density α.

Averaging over the angles of inclination of the fibers and then isolating the part of the averaged scattering operator which is transverse to $\mathbf{k}$, we obtain an expression for $\tilde{\Sigma}_{ij}$:

$$\tilde{\Sigma}_{ij}=2n\Pi_{ij}(\tilde{T}_{22}+\tilde{T}_{33})+2n\Pi_{i1}\Pi_{1j}[2\tilde{T}_{11}-(\tilde{T}_{22}+\tilde{T}_{33})].$$

Comparing this expression with Eq. (34), we write equations for determining $\kappa_\perp$ and $\kappa_\parallel$:

$$\kappa_\parallel^2=-16\pi n\left[\ln\frac{\gamma a}{4}(\kappa_\perp+\kappa_\parallel)+\frac{\kappa_\parallel}{\kappa_\perp+\kappa_\parallel}-\frac{1}{2}\right]^{-1},$$

$$\kappa_\perp^2=-4\pi n\left\{\left[\ln\frac{\gamma a}{2}\kappa_\perp\right]^{-1}+2\left[\ln\frac{\gamma a}{4}(\kappa_\perp+\kappa_\parallel)-\frac{\kappa_\parallel}{\kappa_\perp+\kappa_\parallel}+\frac{1}{2}\right]^{-1}\right\}. \quad (36)$$

In the polydisperse case the right-hand sides of these equations must be averaged over the fiber radii.

The solutions of these equations are shown in Fig. 9 as functions of the packing density $\alpha=4\pi na^2$. Clearly, the anisotropy of the tensor of friction coefficients in the Brinkman equation

$$\epsilon=\frac{\kappa_\parallel^2-\kappa_\perp^2}{\kappa_\parallel^2}$$

increases slowly with increasing packing density, remaining at the level $\epsilon\sim0.3$. Figure 10 shows the results of calculations of the resistance force as a function of the packing density for an anisotropic medium with and without allowance for the anisotropy of the Green's function. It is evident that disregarding the anisotropy of the Green's function, as in Ref. 14, leads to a somewhat higher resistance force.

If the anisotropy is ignored, it is easy derive the self-consistency condition for the model under consideration:

$$\kappa^2=-16\pi n\left[\ln\frac{\gamma a}{2}(\kappa^2+k_*^2)^{1/2}-\frac{1}{2}+\frac{\kappa^2}{2k_*^2}\ln\left(1+\frac{k_*^2}{\kappa^2}\right)\right]^{-1}, \quad (37)$$

and the expression for the resistance force per unit fiber length

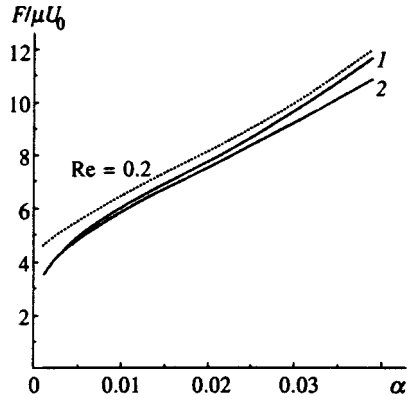


FIG. 10. Resistance force per unit length of fiber of the model filter versus packing density. *1)* Calculated with neglect of the anisotropy of the friction coefficients; *2)* calculated with allowance for the anisotropy; the dashed curve is plotted with allowance for the inertial effects at Re=0.2.

$$F=-4\pi\mu U_0\left[\ln\frac{\gamma a}{2}(\kappa^2+k_*^2)^{1/2}-\frac{1}{2}+\frac{\kappa^2}{2k_*^2}\ln\left(1+\frac{k_*^2}{\kappa^2}\right)\right]^{-1}, \tag{38}$$

in which the inertial effects are taken into account. At high velocities, $k_*^2\gg\kappa^2$, Eq. (38) goes over to the equation for the resistance force of an isolated fiber,[36] and in the opposite limit the self-consistency condition (37) coincides with the result obtained in Ref. 14. The dotted curve in Fig. 10 represents the resistance force as a function of the packing density for Re=0.2.

6. VELOCITY CORRELATION FUNCTION IN A FIBROUS POROUS MEDIUM

To investigate the velocity fluctuations in a random medium, we introduce the velocity correlation function

$$B_{ij}(\mathbf{r})=\int B_{ij}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}=\langle V_i(\mathbf{r})V_j(0)\rangle, \tag{39}$$

where $B_{ij}(\mathbf{k})=\langle V_i(\mathbf{k})V_j^*(\mathbf{k})\rangle$ satisfies the Bethe–Salpeter equation.[32] In the kernel of the Bethe–Salpeter equation for a dilute medium it is sufficient to retain only the term corresponding to the first diagram in Fig. 8. In this approximation we have the following equation for the correlation function:

$$B_{ij}(\mathbf{k})=U_0^2\delta_{i1}\delta_{j1}\delta(\mathbf{k})+\tilde{G}_{is}(\mathbf{k})\tilde{G}_{jp}(\mathbf{k})n\times\int\langle\tilde{T}_{sm}(\mathbf{k},\mathbf{k}_1)\tilde{T}_{pn}(\mathbf{k},\mathbf{k}_1)\rangle B_{mn}(\mathbf{k}_1)\frac{d^3k_1}{(2\pi)^3}. \tag{40}$$

We henceforth assume that the inertial effects and the nonzero value of ϵ can be disregarded. We are also interested in the solution in the region $k\ll\kappa$, where the scattering operator can be considered to be independent of k. We solve this equation by expanding in powers of the density n. In the zeroth approximation the solution corresponds to an unperturbed flow. In the linear approximation with respect to the density we obtain the correlation function in the form

$$B_{ij}=4nU_0^2\tilde{G}_{i1}\tilde{G}_{j1}\int\tilde{T}_{11}^2\delta(\mathbf{n}\cdot\mathbf{k})f(\mathbf{n})\frac{d\Omega}{4\pi}, \tag{41}$$

where the integration is carried out over the surface of a sphere. For an isotropic medium

$$\tilde{T}_{11}=4\pi\left[\ln\frac{\gamma a\kappa}{2}\right]^{-1}\left[1-\frac{1}{2}n_1n_1\right].$$

Integrating over angles, in this case we arrive at a correlation function of the form

$$B_{ij}=4nU_0^2\frac{(4\pi)^2}{[\ln(\gamma a\kappa/2)]^2}\Pi_{i1}\Pi_{j1}(k^2+\kappa^2)^2\frac{1}{4k}\times\left[\left(1-\frac{k_\perp^2}{4k^2}\right)^2+\frac{k_\perp^4}{32k^4}\right]. \tag{42}$$

where $k_\perp^2=k^2-k_1^2$. For a material in which all fibers are oriented orthogonal to the direction of flow, similar calculations give

$$B_{ij}=4nU_0^2\frac{(4\pi)^2}{[\ln(\gamma a\kappa/2)]^2}\frac{\Pi_{i1}\Pi_{j1}}{k_\perp(k^2+\kappa^2)^2}. \tag{43}$$

For this model the velocity correlation function in coordinate space is equal to

$$B_{ij}(\mathbf{r})=4nU_0^2\frac{(4\pi)^2}{[\ln(\gamma a\kappa/2)]^2}\int\frac{\Pi_{i1}\Pi_{j1}e^{i\mathbf{k}\cdot\mathbf{r}}}{k_\perp(k^2+\kappa^2)^2}\frac{d^3k}{(2\pi)^3}. \tag{44}$$

If we introduce the function

$$B_0(\kappa,\mathbf{r})=\int\frac{1}{k_\perp(k^2+\kappa^2)}e^{i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3}=\frac{1}{4\pi}I_0\left[\frac{1}{2}\kappa(r-x)\right]K_0\left[\frac{1}{2}\kappa(r+x)\right],$$

we can represent $B_{ij}(\mathbf{r})$ in the form

$$B_{ij}(\mathbf{r})=4nU_0^2\frac{(4\pi)^2}{[\ln(\gamma a\kappa/2)]^2}\left\{\delta_{i1}\delta_{j1}B_1+\left(\delta_{i1}\frac{\partial}{\partial x_j}+\delta_{j1}\frac{\partial}{\partial x_i}\right)B_2+\frac{\partial^2}{\partial x_i\partial x_j}B_3\right\}. \tag{45}$$

The B_i are expressed in terms of B_0 as follows:

$$B_1=-\frac{\partial}{\partial\kappa^2}B_0,$$

$$B_2=-\frac{\partial^2}{\partial x\partial\kappa^2}\left[\frac{1}{\kappa^2}\left(B_0-\frac{1}{4\pi}\ln\frac{4}{\gamma(r+x)}\right)\right],$$

$$B_3=-\frac{\partial^2}{\partial x^2\partial\kappa^2}\left[\frac{1}{\kappa^4}\left(B_0-\frac{1}{4\pi}\ln\frac{4}{\gamma(r+x)}\right)-\frac{r^2+x^2}{32\pi\kappa^2}\ln\frac{4}{\gamma(r+x)}\right].$$

In the region $\kappa x\gg1$ the exponentially small contribution of B_0 to the B_i can be neglected. The remaining logarithmic terms make the correlations decay as r^{-2} at large distances. For small values of the longitudinal coordinate, $\kappa x\leqslant1$, it is

necessary to add the contribution of the function B_0, which behaves as r^{-1} at $\kappa r \gg 1$. The correlation of the longitudinal velocities in this region decays as r^{-1}. The value of the correlation tensor at zero, $B_{ij}(0)$, is readily determined directly from Eq., (44). The tensor is, in fact, diagonal and has the components

$$B_{11}=6B_{22}=6B_{33}=\frac{3}{32}U_0^2\left(\ln\frac{2}{\gamma a\kappa}\right)^{-1}.$$

Clearly, the kinetic energy contained in the longitudinal velocity components is three times higher than that in the transverse components.

Knowing the velocity correlation function, we can find the longitudinal and transverse diffusion coefficients of a passive impurity embedded in a fluid. For a particle that is situated at the point $\mathbf{r}=0$ at the time $t=0$ and moves together with the fluid we have

$$\mathbf{r}(t)=\int_0^t \mathbf{V}(\mathbf{r}(t'))dt'.$$

For the mean-square component of the displacement relative to the unperturbed trajectory $x_i'=(x_i-U_{0i}t)$ we obtain

$$\langle x_i'^2\rangle=\int_0^t\int_0^t\int \langle \tilde V_i(\mathbf{k})\tilde V_i^*(\mathbf{k}')\exp[ik_1U_0t'-k_1'U_0t''+i\mathbf{k}\cdot\mathbf{r}_1(t')-i\mathbf{k}'\cdot\mathbf{r}_1(t'')]\rangle\frac{dk\,dk'}{(2\pi)^6}\,dt'\,dt''.$$

If we assume that $\mathbf{r}_1(t)$ is described by a stochastic Wiener process[20] with the transition probability density function

$$P(\mathbf{r}(t)|\mathbf{r}(t'))=\frac{1}{8D_\perp D_\parallel^{1/2}[\pi(t-t')]^{3/2}}\times\exp\left\{-\frac{[x(t)-x(t')]^2}{4D_\parallel(t-t')}-\frac{[\mathbf{r}_\perp(t)-\mathbf{r}_\perp(t')]^2}{4D_\perp(t-t')}\right\} \tag{46}$$

and correlates weakly with $\tilde V_i(\mathbf{k})$, we obtain the following expression for the displacement squared:

$$\langle x_i'^2(t)\rangle=\int_0^t\int_0^t\int \langle \tilde V_i(\mathbf{k})\tilde V_i^*(\mathbf{k}')\rangle\langle\exp[ik_1U_0(t'-t'')+i\mathbf{k}_\perp\cdot(\mathbf{r}_{1\perp}(t')-\mathbf{r}_{1\perp}(t''))]\rangle\frac{dk}{(2\pi)^3}dt'dt''.$$

Calculating the average of the exponential function with the transition density function (46), we find that at large times ($U_0\kappa t\gg 1$) the displacement squared obeys the diffusion law $\langle x_i'^2\rangle=2D_it$ with the following diffusion coefficients in the longitudinal and transverse directions to the flow, respectively:

$$2D_\parallel=\int\int_{-\infty}^{\infty}B_{11}(\mathbf{k})\exp(ik_1U_0\tau-k_1^2D_\parallel|\tau|-k_\perp D_\perp|\tau|)d\tau\frac{d^3k}{(2\pi)^3}, \tag{47}$$

$$2D_\perp=\int\int_{-\infty}^{\infty}B_{22}(\mathbf{k})\exp(ik_1U_0\tau-k_1^2D_\parallel|\tau|-k_\perp D_\perp|\tau|)d\tau\frac{d^3k}{(2\pi)^3}. \tag{48}$$

In calculating the longitudinal diffusion coefficient we can ignore the terms in the exponential function (47) that contain

$$D_\parallel=\frac{\pi}{U_0}\int B_{11}(\mathbf{k})\,\delta(k_1)\frac{d^3k}{(2\pi)^3}=-\frac{\pi}{4}\frac{U_0}{\kappa\,\ln(\gamma a\kappa/2)}.$$

In calculating the transverse diffusion coefficient we must retain the term in the exponential function (48) that contains

$$D_\perp=\int B_{22}(\mathbf{k})\frac{D_\parallel}{U_0^2+D_\parallel^2k_1^2}\frac{d^3k}{(2\pi)^3}.$$

Integrating and making use of the fact that $D_\parallel\kappa\ll U_0$, we obtain

$$D_\perp=-D_\parallel\frac{1}{8\,\ln(\gamma a\kappa/2)}.$$

The transverse diffusion coefficient in the model under consideration is found to be much smaller than the longitudinal diffusion coefficient. We note that the longitudinal diffusion coefficient for an anisotropic medium differs only by a numerical factor from the analogous coefficient obtained in Ref. 37 for an isotropic fibrous medium.

7. CONCLUSION

In this paper the diagram technique developed in multiple-scattering theory to investigate wave fields in random media has been used to study the hydrodynamics of three-dimensional fibrous porous media. As in problems concerning the hydrodynamics of suspensions, a system of self-consistent equations (24)–(26) has been derived for the Green's function in a porous medium and the renormalized friction coefficient matrix, which is analogous to the scattering operator in wave problems. It has been shown that the series of multiple-scattering theory can be partially summed in such a way that all the diagrams contain only renormalized self-consistent Green's functions and the scattering operator. Self-consistent field theory has been used to investigate the averaged equations of fluid flow in an isotropic medium and in an anisotropic fibrous filter. An expression has been derived for the renormalization of the viscosity in an isotropic fibrous medium as a function of the packing density and the geometric dimensions of the fiber (32). The velocity correlation function (45) has been calculated, and the longitudinal and transverse diffusion coefficients have been determined for a passive impurity moving together with a fluid.

I thank A. A. Vedenov, A. A. Kirsch, V. I. Roldutin, I. B. Stechkina, and S. V. Khudyakov for some helpful discussions.

APPENDIX

In this Appendix we calculate the scattering operator for a fiber oriented perpendicularly to the direction of the average velocity. To calculate the scattering operator using Eq. (27), we first need to evaluate the two integrals

$$I_1=\int \frac{e^{i\mathbf{k}_\perp\cdot\mathbf{r}}}{k^2+2i\mathbf{k}\cdot\mathbf{k}_*+\kappa_\perp^2}\frac{d^2k}{(2\pi)^2}, \tag{49}$$

$$I_2=\int \frac{e^{i\mathbf{k}_\perp\cdot\mathbf{r}}}{k^2(k^2+2i\mathbf{k}\cdot\mathbf{k}_*+\kappa_\perp^2)}\frac{d^2k}{(2\pi)^2}. \tag{50}$$

Invoking the representation

$$x^{-1}=\int_0^\infty e^{-\alpha x}d\alpha$$

for the denominator in I_1 and then taking the Gauss integral over the Fourier harmonics, we obtain

$$I_1=\frac{1}{4\pi}e^{\mathbf{r}\cdot\mathbf{k}_*}\int_0^\infty\frac{d\alpha}{\alpha}\exp\left\{-(k_3^2+\kappa_\perp^2+k_*^2)\times\left[\alpha+\frac{r^2}{4\alpha(k_3^2+\kappa_\perp^2+k_*^2)}\right]\right\}$$
$$=\frac{1}{2\pi}e^{\mathbf{r}\cdot\mathbf{k}_*}K_0(r\lambda), \tag{51}$$

where $\lambda^2=(k_3^2+\kappa_\perp^2+k_*^2)$, and we have used the integral representation for the Bessel functions of an imaginary argument[38]

$$\int_0^\infty\exp\left[-\frac{x}{2}\left(t+\frac{z^2}{t}\right)\right]t^{-\nu-1}dt=2z^{-\nu}K_\nu(xz). \tag{52}$$

Analogously, I_2 can be reduced to the form

$$I_2=\frac{1}{4\pi}\int_0^1 d\eta e^{\mathbf{r}\cdot\mathbf{k}_*\eta}\frac{r}{\lambda(\eta)}K_1[r\lambda(\eta)], \tag{53}$$

where $\lambda^2(\eta)=k_3^2+\kappa_\perp^2\eta+k_*^2\eta^2$. Invoking the expansions of the Bessel functions for small values of the argument[38]

$$K_0(x)\approx-\ln\frac{\gamma x}{2},\quad K_1(x)\approx\frac{1}{x}+\frac{x}{2}\ln\frac{\gamma x}{2}-\frac{x}{4}, \tag{54}$$

where $\gamma=1.781$ is the Euler constant, for the first integral we obtain

$$I_1=-\frac{1}{2\pi}\ln\frac{\gamma r\lambda}{2}, \tag{55}$$

and I_2 is reduced to the form

$$I_2=\frac{1}{4\pi}\left\{\frac{r^2}{2}\left(\ln\frac{\gamma r\lambda}{2}-\frac{1}{2}\right)+\int_0^1\frac{1+(k_*x-\kappa^2r^2/4)\eta-k_*^2y^2\eta^2/2}{k_3^2+\kappa^2\eta+k_*^2\eta^2}d\eta\right\}. \tag{56}$$

This integral is readily evaluated, but it is simpler to calculate various limiting cases directly from Eq. (56). The matrix elements sought are expressed in terms of I_1 and I_2 and the derivatives of I_2 with respect to the coordinates with subsequent averaging over the surface of a cylinder of radius $r=a$. As a result, we obtain expressions for $\tilde{T}_{ij}^{-1}$ in the thin-fiber approximation. In the case of a dilute medium or high harmonics, where the condition $\kappa^2\ll k_3k_*$ holds, the calculations yield matrix elements for the inverse operator, which coincides with the bare operator in this case:

$$T_{11}^{-1}=-\frac{1}{4\pi}\left[-\ln\frac{\gamma a(k_3^2+k_*^2)^{1/2}}{2}+\frac{1}{2}\right],$$
$$T_{22}^{-1}=T_{11}^{-1}-\frac{1}{4\pi}\left[-1+\frac{k_3}{k_*}\arctan\left(\frac{k_3}{k_*}\right)\right], \tag{57}$$
$$T_{33}^{-1}=2T_{11}^{-1}+\frac{1}{4\pi}\left[1+\frac{k_3}{k_*}\arctan\left(\frac{k_3}{k_*}\right)\right],$$
$$T_{13}^{-1}=T_{31}^{-1}=-\frac{i}{2}\frac{k_3}{k_*}\ln\frac{(k_3^2+k_*^2)}{k_3^2}.$$

In the opposite case, $\kappa^2\gg k_3k_*$, we encounter two different possibilities. We first disregard the inertial effects altogether, assuming that $k_*=0$. For the matrix elements of the inverse scattering operator, which is diagonal in this case, we obtain

$$\tilde{T}_{11}^{-1}=\tilde{T}_{22}^{-1}=\frac{1}{2}\tilde{T}_{33}^{-1}=-\frac{1}{4\pi}\left[-\ln\frac{\gamma a(k_3^2+\kappa^2)^{1/2}}{2}+\frac{1}{2}\frac{k_3^2}{\kappa^2}\ln\frac{(k_3^2+\kappa^2)}{k_3^2}\right]. \tag{58}$$

In the case of $k_3=0$ with allowance for the inertial terms, we can confirm that the operator is once again diagonal:

$$\tilde{T}_{33}^{-1}=\frac{1}{2\pi}\ln\frac{\gamma a(k_*^2+\kappa^2)^{1/2}}{2},$$
$$\tilde{T}_{11}^{-1}=\frac{1}{2}\tilde{T}_{33}^{-1}-\frac{1}{4\pi}\left[\frac{1}{2}-\frac{1}{2}\frac{\kappa^2}{k_*^2}\ln\frac{(k_*^2+\kappa^2)}{\kappa^2}\right],$$
$$\tilde{T}_{22}^{-1}=\frac{1}{2}\tilde{T}_{33}^{-1}-\frac{1}{4\pi}\left[-\frac{1}{2}+\frac{1}{2}\frac{\kappa^2}{k_*^2}\ln\frac{(k_*^2+\kappa^2)}{\kappa^2}\right]. \tag{59}$$

Inverting the matrices obtained, we find the scattering operator under the stated conditions.

The expressions obtained are valid for infinitely long fibers. To assess the influence of the finite length of the fibers, we consider a cylinder of length $2l$ aligned with the z axis. We assume that its lower base lies in the $z=0$ plane. Since the fiber now exerts a nonzero force on the fluid only at $0<z<2l$, the force can be expanded into a Fourier series in this interval:

$$\mathbf{F}(z)=\sum_n\mathbf{F}_n\sin k_nz,$$

where $k_n=\pi n/2l$. An analogous expansion can be written for the velocity of the external fluid flow on this segment of the z axis:

$$\mathbf{V}(z)=\sum_n\mathbf{V}_n\sin k_nz.$$

The coefficients of the Fourier series $\mathbf{F}_n$ and $\mathbf{V}_n$ are related to the coefficients of the integral Fourier transform:

$$\mathbf{F}_n(k_3)=\sum_n f^*(k_n,k_3)\mathbf{F}_n,$$

$$\mathbf{V}_n=\frac{1}{l}\int f(k_n,k_3)\mathbf{V}(\mathbf{k})\frac{d^3k}{(2\pi)^3},$$

where we have introduced the function

$$f(k_n,k_3)=\int_0^{2l}\sin k_n z e^{-ik_3 z}dz=ie^{ik_3 l}\left[e^{-ik_n l}\frac{\sin(k_3-k_n)l}{(k_3-k_n)}-e^{ik_n l}\frac{\sin(k_3+k_n)l}{(k_3+k_n)}\right].$$

The distribution of forces along a fiber must be determined from the condition of zero net velocity on its surface:

$$V_i(z)=\frac{1}{\mu}\int \tilde{T}_{ij}^{-1}F_j(k_3)\frac{dk_3}{2\pi}. \tag{60}$$

Substituting the expressions for the velocity and the force in the form of Fourier series into this equation on the interval $0<z<2l$, we obtain

$$V_{in}=\int \tilde{T}_{ij}^{-1}(k_3)Q_n(k_3,k_n)\frac{d^3k}{(2\pi)^3}\frac{F_{jn}}{\mu}, \tag{61}$$

where

$$Q_n(k_3,k_n)=\left[\frac{\sin(k_3-k_n)2l}{k_3-k_n}+\frac{\sin(k_3+k_n)2l}{k_3+k_n}\right].$$

The scattering operator of an infinitely long fiber is diagonal for an isotropic medium at low flow velocities. In this case we can use Eq. (61) to express the resistance force explicitly in terms of the flow field and to obtain an expression for the diagonal components of the scattering operator for a finite cylinder:

$$\tilde{T}_i(\mathbf{k},\mathbf{k}')=\frac{1}{l}\sum_n\frac{f^*(k_n,k_3)f(k_n,k_3')}{\int\tilde{T}_i^{-1}(k_3'')Q_n(k_3'',k_n)dk_3''/2\pi}. \tag{62}$$

Under the condition $\kappa l\gg l$ the function $Q_n(k_3,k_n)$ can be replaced by $2\pi\delta(k_3-k_n)$. For disturbances which are short in comparison with the wavelength, $(k_3 l,k_3' l)\gg 1$, the main contribution to the sum (62) is from large values of n, at which the sum can be replaced by the integral over k_n. Making use of the fact that the function $f(k_n,k_3)$ is proportional to $\pi\delta(k_n-k_3)$ when $k_n l\gg 1$, we obtain the scattering operator in this limit

$$\tilde{T}_i(\mathbf{k},\mathbf{k}')=2\frac{\sin[(k_3-k_3')l]}{(k_3-k_3')}\exp(i(k_3'-k_3)l)\tilde{T}_i(k_3).$$

In the long-wavelength range, $(k_3 l,k_3' l)\ll 1$, the main contribution to the sum (62) is from small values of n. In particular, for the zero-angle scattering operator we obtain the expression

$$\tilde{T}_i(\mathbf{k},\mathbf{k})=\frac{\tilde{T}_i(0)}{l}\sum_n|f(k_n,0)|^2+\frac{k_3^2}{2l}\sum_n\tilde{T}_i(k_n)\frac{\partial^2}{\partial k_3^2}|f(k_n,0)|^2. \tag{63}$$

Calculating the sum over n, we obtain an expression for the diagonal components of the scattering operator in the long-wavelength limit:

$$\tilde{T}_1(\mathbf{k},\mathbf{k})=\tilde{T}_2(\mathbf{k},\mathbf{k})=2\tilde{T}_3(\mathbf{k},\mathbf{k})=\frac{8\pi l}{\ln(\gamma\kappa a/2)}\left[1+\frac{k_3^2\ln(2\kappa l/\pi)}{\kappa^2\ln(\gamma\kappa a/2)}\right]. \tag{64}$$

*)E-mail: tburinsk@classic.iki.rssi.ru

[1] H. Cheng and G. Papanicolaou, J. Fluid Mech. **335**, 189 (1997).
[2] B. Cichocki and K. Hinsen, Phys. Fluids **2**, 285 (1995).
[3] B. Cichocki, B. U. Felderhof, K. Hinsen, *et al.*, J. Chem. Phys. **100**, 3780 (1994).
[4] I. L. Claeys and J. F. Brady, J. Fluid Mech. **251**, 411 (1993); **443**, 479 (1993).
[5] T. Tran-Cong and N. Phan-Thien, Phys. Fluids A **1**, 453 (1989).
[6] J. F. Brady and G. Bossis, Annu. Rev. Fluid Mech. **20**, 111 (1988).
[7] V. N. Pokrovskiĭ, *Statistical Mechanics of Dilute Suspensions* [in Russian] (Nauka, Moscow, 1978).
[8] D. J. Jeffrey and A. Acrivos, AIChE. J. **22**, 417 (1976).
[9] G. K. Batchelor, Annu. Rev. Fluid Mech. **6**, 227 (1974).
[10] I. D. Howells, J. Fluid Mech. **64**, 449 (1974).
[11] E. J. Hinch, J. Fluid Mech. **83**, 695 (1977).
[12] G. W. Jackson and D. F. James, Can. J. Chem. Eng. **64**, 364 (1986).
[13] A. A. Kirsch and I. R. Stechkina, in *Fundamentals of Aerosol Science*, edited by D. T. Shaw (Wiley, New York, 1978), p. 165.
[14] L. Spielman and S. L. Goren, Environ. Sci. Technol. **2**, 279 (1968).
[15] H. C. Brinkman, Appl. Sci. Res., Sect. A **1**, 27 (1947).
[16] J. J. L. Higdon and G. D. Ford, J. Fluid Mech. **308**, 341 (1996).
[17] R. C. Brown, Filtr. Sep. (3), 165 (1997).
[18] A. L. Chernyakov and A. A. Kirsch, Kolloidn. Zh. **59**, 698 (1997).
[19] L. A. Santalo, *Integral Geometry and Geometric Probability* (Addison-Wesley, Reading, Mass., 1976) [Russ. transl., Nauka, Moscow, 1983].
[20] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (North-Holland, Amsterdam-New York, 1992) [Russ. transl., 1st ed., Vysshaya Shkola, Moscow, 1990].
[21] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., rev., Pergamon Press, Oxford-New York (1987) [Russ. original, 3rd ed., Nauka, Moscow, 1973].
[22] B. U. Felderhof, Physica A **151**, 1 (1988).
[23] P. Mazur, Faraday Discuss. Chem. Soc. **83**, 33 (1987).
[24] K. F. Freed and M. Muthukumar, J. Chem. Phys. **68**, 2088 (1978).
[25] M. Muthukumar and K. F. Freed, J. Chem. Phys. **70**, 5875 (1979).
[26] J. M. Burgers, in *Second Report on Viscosity and Plasticity* (Eerste Sectie), DI XVI, No. 4 (K. Ned. Akad. Wet. Verhand., 1938), p. 113.
[27] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics*, 2nd rev. ed. (Noordhoff, Leiden, Netherlands, 1973) [Russ. transl., Mir, Moscow, 1976].
[28] A. Ishimaru, *Wave Propagation and Scattering in Random Media, Vol. 2: Multiple Scattering, Turbulence, Rough Surfaces, and Remote Sensing* (Academic Press, New York, 1978) [Russ. transl., Mir, Moscow, 1981].
[29] C. K. W. Tam, J. Fluid Mech. **38**, 537 (1969).
[30] K. F Freed and S. F. Edwards, J. Chem. Phys. **61**, 3626 (1974).
[31] E. S. G. Shaqfen and G. H. Fredrickson, Phys. Fluids A **2**, 7 (1990).
[32] U. Frisch, in *Probabilistic Methods in Applied Mathematics*, Vol. 1, edited by A. T. Bharucha-Reid (Academic Press, New York, 1968), p. 76.
[33] S. Childress, J. Chem. Phys. **56**, 2527 (1972).
[34] C. R. Gochnour, H. C. Anderson, and M. D. Fayer, J. Chem. Phys. **70**, 4254 (1979).

[35] J. Koplik and H. Levine, Phys. Fluids **26**, 2864 (1983).
[36] H. L. Lamb, *Hydrodynamics*, 6th ed. (Cambridge Univ. Press, Cambridge, 1932) [Russ. transl., Gostekhizdat, Moscow, 1947].
[37] D. I. Koch and J. F. Brady, AIChE. J. **32**, 575 (1986).
[38] I. S. Gradshteyn (Gradshteĭn) and I. M. Ryzhik, *Table of Integrals, Series, and Products* transl. of 4th ed. (Academic Press, New York, 1980) [Russ. original, 5th ed., Nauka, Moscow, 1971].

Translated by James S. Wood
Edited by P. Shelnitz

The method of exchange perturbation theory as applied to magnetic ordering in high-T_c materials

E. V. Orlenko*) and T. Yu. Latyshevskaya

St. Petersburg State Technical University, 195251 St. Petersburg, Russia
(Submitted 14 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 2129–2147 (June 1998)

We unify the method of exchange perturbation theory for multicenter systems. For the case of exchange degeneracy in the total spin of the system we give a secular equation that is more compact and convenient for calculations than those obtained earlier. On the basis of this formalism we develop an algorithm for calculating the Heisenberg parameter for magnetic materials. Finally, we calculate the characteristics of antiferromagnetic transitions for the high-T_c materials $La_{2-x}MeCuO_4$ and $YBa_2Cu_3O_6$. © *1998 American Institute of Physics.* [S1063-7761(98)01606-0]

INTRODUCTION

Usually the microscopic description of magnetic materials is either a statistical analysis of spin systems on the basis of the Heisenberg equation or a calculation and analysis of magnetization and magnetic susceptibility in the single-electron approximations in models of the Stoner[1] or Hubbard[2] type. But, in one way or another, the main parameter characterizing the spin system is still the exchange integral, which is chosen differently in different models; for instance, in Hubbard-type models it is the Coulomb exchange interaction of electrons strongly localized at the centers and, therefore, calculated in the Wannier-function representation. The Heisenberg parameter in spin models is estimated semiphenomenologically, by reduction to the simplest Heitler–London two-center problem.[3] Clearly, the intercenter interaction of electrons belonging to the inner shells of ions, which is responsible for the spontaneous orientation of the spins, is much more complicated than in the models. First, the overlap of the atomic wave functions of the inner electrons belonging to different centers is responsible for their spin correlation; thus, the use of a ''truncated'' Wannier-function basis artificially reduces the contribution of the intercenter exchange to the interaction energy and in this way essentially eliminates the intercenter correlation effects. Second, for many materials the number of ''active'' electrons of the inner shells of atoms participating in intercenter interaction exceeds unity, so that the wave function describing at least a two-center system is more than two-particle, and its spatial part is not reduced only to symmetric or antisymmetric form, as it is in the Heitler–London model. The use of the Slater determinant, which incorporates both coordinate and spin one-electron states simultaneously, makes it impossible to analyze the spin state emerging as a result of the interaction. Third, often the magnetic orientation of the spins is caused not simply by two-center exchange but by superexchange, in which the electrons of three or more centers participate.[4,5] The constants of such interaction are approximated variationally by combinations of pair integrals,[6] which actually means that the nonadditive part specific to multicenter interaction is discarded.

The situation is such that a meaningful description of spin systems requires not only effective summing over the states of the possible spin configurations (an enormous number of fine papers have been written on the subject, including those that use the renormalization method) but also developing an algorithm that would allow doing consistent calculations of the fundamental parameter present in any statistical scheme, the Heisenberg parameter.

The discovery of anomalous magnetic effects in high-T_c superconductors is vivid proof of the necessity of developing such an algorithm, since these effects are caused not so much by structural transformations in the crystal as by the change in the nature of the exchange interaction proper. The point is that such crystals as $YBa_2Cu_3Cu_3O_6$ and La_2CuO_4 and the materials Rb_2MnF_6 and Rb_2CoF_4 isostructural to the latter, in the pure or stoichiometric state are antiferromagnetic insulators with a fairly high transition temperature. Alloying, which is done by replacing the La^{2+} ion by an atom of a metal with valence 2^+ (such as Cu^{2+}, Ba^{2+}, and Sr^{2+}, so that we have the alloys $La_{2-x}Me_xCuO_4$, $Rb_2Mn_{1-x}Mg_xF_4$, and $Rb_2Co_{1-x}Mg_xF_4$, respectively) or by changing the oxygen content ($La_2CuO_{4-\delta}$ and $YBa_2Cu_3O_6$), lowers the Néel temperature so drastically that the antiferromagnetic state may be destroyed[7–9] and replaced by a weakly fluctuating 3D state of a spin liquid with preferentially parallel pair orientation of the spins.

A theoretical analysis of the behavior of these systems yields contradictory results. For example, according to calculations done with the one-electron band approximation, the La_2CuO_4 compound is a nonmagnetic metal.[10,11] At the same time, electronic-structure models used in studies of the Heisenberg Hamiltonian for two- and three-center systems in the representation of the spin eigenfunctions of the operator $J(\hat{s}_1+\hat{s}_2)\hat{s}_3$ provide a fairly realistic phase diagram that describes the transition of the system from the 2D antiferromagnetic state into the 3D state of a spin liquid and then into

the superconducting state.[12] However, in these papers the numerical value of J for different spin configurations is estimated by analyzing experimental data.

The Heisenberg parameter can in principle be calculated consistently by applying the formulas of exchange perturbation theory, which takes into account the effects of intercenter overlap of the wave function. There are many formal variants of this theory, which are classified according to the way in which the algorithm is constructed (a detailed description of this classification can be found in Kaplan's monograph,[13] which also analyzes the merits and drawback of the variants). There are two problems that must be dealt with in constructing the algorithm of exchange perturbation theory: the nonorthogonality of the base of the multicenter-system wave functions, which are antisymmetric in intercenter permutations (this problem is related to what is known as the overfilling catastrophe), and the asymmetry of the perturbation operator and the unperturbed part of the Hamiltonian with respect to intercenter permutations of electrons. In other words, if $[H,A]=0$, where $H=H^0+V$ is the total system Hamiltonian, H^0 is its unperturbed part, and A is the antisymmetrization operator, then $[H^0,A]\neq 0$ and $[V,A]\neq 0$. The zeroth wave function antisymmetrized in intercenter permutations is not an eigenfunction of H^0, and the corrections calculated in the perturbation V contain nonphysical contributions.

All variants of exchange perturbation theory can be divided into two groups.[13] The first consists of theories nonsymmetric in the Hamiltonian. The second consists of approaches that make it possible to use the common Rayleigh–Schrödinger perturbation theory by setting up a special zeroth symmetric Hamiltonian for which the antisymmetric functions are eigenfunctions. The first group uses the fundamental basis of zeroth functions nonsymmetric in intercenter permutations, functions that are the eigenfunctions of the nonsymmetric Hamiltonian H^0. Antisymmetrization is done *post factum* at each interpolation step, which in the final analysis requires using a variational procedure in the perturbation theory formalism (as, say, is done in Ref. 14). An attempt to modify the Hamiltonian so that the perturbation operator becomes symmetric (this is known as the Sternheimer procedure[15,16]) leads to a non-Hermitian total Hamiltonian and actually limits the use of the method to two-electron systems.

The second group of variants of exchange perturbation theory can be assumed to include the work of Ritchie.[17] In this paper special projection operators are employed whose action on an antisymmetric function is reduced to selecting a term with a specific permutation. Since the explicit form of these operators was not given, it was assumed that they make the Hamiltonian non-Hermitian. Rumyantsev[18] demonstrated the effectiveness of using such symmetrization of the Hamiltonian. Despite a conceptual difficulty, a variant of exchange perturbation theory the Rayleigh–Schrödinger form was constructed, and on the basis of this theory the spectral characteristics of the hydrogen–helium system were calculated with high accuracy. Only in Ref. 19 were the projection operators derived explicitly. These operators symmetrize the perturbation operator and the unperturbed Hamiltonian and retain the hermiticity of the total Hamiltonian in the sense that its eigenvalues are real numbers. In the same paper the variant of exchange perturbation theory was generalized to the case where the system is degenerate in total spin. A detailed description of the variant and a broad range of applications used in calculating specific system, including spin system, can be found in Ref. 5. Unfortunately, the organization of the material in Refs. 5 and 19 makes it impossible to explicitly analyze all the small parameters of the theory in which the power series expansions are done, and this makes the use of the formulas difficult. We also note that in Ref. 19 the corrections to the energy when degeneracy in total spin is lifted were calculated only for the case where there is another degeneracy, in orbital momentum.

In the present paper we use the idea of the method of symmetrizing the Hamiltonian[17–19] and construct a more compact algorithm of exchange perturbation theory. This allowed us to estimate the smallness of the terms discarded at each iteration step, terms that emerge because of the overfilling of the nonorthogonal base of antisymmetric functions. Due to a change in the normalization condition for the antisymmetric functions, all the projections in this variant, including Ritchie-type operators, are simpler. This has made it possible to obtain a solution of the secular equation when the system is degenerate in total spin in a more general form, more suitable for calculations. To show the possibilities of our version of exchange perturbation theory with degeneracy, we examined the high-T_c materials $La_{2-x}Sr_xCuO_4$ and $Ba_2Cu_3O_{6+x}$, for which we calculated the Heisenberg parameter in the stoichiometric and alloyed states. We show that alloying these materials dramatically changes the magnitude and sign of the exchange and superexchange integrals. These changes lead, in accordance with the models of Ref. 12, to destruction of the 2D antiferromagnetic state and emergence of a 3D ferromagnetic state of spin liquid.

1. EXCHANGE PERTURBATION THEORY

A system of noninteracting atoms in the adiabatic approximation can be specified solely by the electron part of the wave function, which is simply the product of atomic wave functions. We write the spatial part as

$$\Phi(r_1,\ldots,r_N)=\prod_{\alpha}\ \psi_\alpha(r_1,\ldots,r_j), \tag{1.1}$$

where α is the number of a center, or atom, and $i,\ldots,j$ are the numbers of the electrons belonging to an atom.

The Hamiltonian describing such a system consists of the kinetic energy of all the electrons, the potential energy of the interaction of electrons and the ''parent'' center, and the interaction of the electrons belonging to one center with each other.

If the Hamiltonian does explicitly contain spin operators, then

$$H^0\Phi_n(r_1,\ldots,r_N)=E_n^0\Phi_n(r_1,\ldots,r_N),$$

where $\{E_n^0\}$ is the set of the eigenfunctions of the energy of the noninteracting system, and $\Phi(r_1,\ldots,r_N)$ are the eigenfunctions corresponding to this spectrum.

The distances between the centers are such that atomic wave functions may strongly overlap. Then, in accordance with the Pauli exclusion principle, even for a noninteracting system the complete wave function must be antisymmetrized. So that the spin part can easily be separated later, leaving only the spatial part, Young tableaux can be used to antisymmetrize the complete wave function.[20] Then the spatial part of the complete wave function of the noninteracting system is

$$\Psi_n^0(r_1,\ldots,r_N)=A\Phi_n(r_1,\ldots,r_N), \tag{1.2}$$

where A is the antisymmetrization operator corresponding only to the spatial part of the Young tableau or, in greater detail,

$$\Psi_n^0(r_1,\ldots,r_N)=\frac{1}{f_n^P}\sum_p^P(-1)^{g_p}\Phi_n^p(r_1,\ldots,r_N), \tag{1.3}$$

where p is the number of permutations, g_p is the parity of that permutation, P is the total number of possible intercenter permutations, $1/f_n^p$ is the normalization constant, and $\Phi_n^p(r_1,\ldots,r_N)$ is the wave function of the form (1.1) containing the permutation p that corresponds to the Young tableau. We find the normalization constant (1.3) from the condition

$$\langle\Phi_n^0|\Psi_n^0(r_1,\ldots,r_N)\rangle=1. \tag{1.4}$$

Then

$$f_n^P=\sum_{p=0}^P(-1)^{g_p}\langle\Phi_n^0|\Phi_n^p\rangle \tag{1.5}$$

differs by a factor $\sqrt{P}$ from the same constant in the normalization $\langle\Psi^0|\Psi^0\rangle=1$ commonly used in exchange perturbation theory.

We introduce a projection operator Λ_n^π that separates the term with the πth permutation in an antisymmetrized function of type (1.3):

$$\Lambda_n^\pi=|\Phi_n^\pi\rangle\langle\Phi_n^\pi|. \tag{1.6}$$

Then $\Lambda_n^\pi|\Psi_n^0\rangle=(-1)^\pi|\Phi_n^\pi\rangle$.

Now we can write the system Hamiltonian without a perturbation in a form invariant under intercenter permutations:

$$H_0=\frac{1}{f_n^P}\sum_{p=0}^P H_p^0\Lambda_n^p,\quad V=\frac{1}{f_n^P}\sum_{p=0}^P V_p\Lambda_n^p, \tag{1.7}$$

where H_p^0 and V_p are the unperturbed Hamiltonian and the perturbation corresponding to the pth intercenter permutation of the electrons.

As usual, the perturbation operator incorporates the interaction of the nuclei, the interaction of electrons with ''foreign'' nuclei, and the interaction of the electrons belonging to different centers. Now the zeroth antisymmetric wave function (1.3) is the eigenfunction of the invariant unperturbed Hamiltonian H_0 (1.7):

$$H^0\Psi_n^0=E_n^0\Psi_n^0. \tag{1.8}$$

The Hamiltonian of the interacting system is always invariant under electron perturbations, so that the solution of the Schrödinger equation can be antisymmetric under any electron permutation:

$$H\Psi=E\Psi. \tag{1.9}$$

Solving Eq. (1.9) by the method of successive approximations, we seek the perturbative corrections to the zeroth energy and wave function taken from (1.8). For instance, for the initial wave function and its corrections to have the proper symmetry, we use the perturbation operator and the unperturbed part of the Hamiltonian in the form of (1.7). Then instead of (1.9) we have

$$(H^0+V)\Psi_i=E_i\Psi_i, \tag{1.10}$$

where

$$\Psi_i=\Psi_i^{(0)}+\Psi_i^{(1)}+\cdots,\quad E_i=E_i^{(0)}+E_i^{(1)}+\cdots.$$

At the beginning we keep only the zeroth- and first-order terms in (1.10). Then, allowing for (1.8), we have

$$H^0\Psi_i^{(1)}+V\Psi_i^{(0)}=E_i^{(1)}\Psi^{(0)}+E_i^{(0)}\Psi_i^{(1)}. \tag{1.11}$$

We impose the intermediate normalization condition

$$\langle\Phi_i|\Psi_i\rangle=\langle\Phi_i|\Psi_i^0\rangle, \tag{1.12}$$

i.e., $\langle\Phi_i|\Psi_i-\Psi_i^0\rangle=0$. This means that all corrections to the wave function of the zeroth approximation lie in the subspace of the state vectors orthogonal to $|\Psi_i^0\rangle$.

Let us introduce the projector on the subspace of vectors parallel to $|\Psi_i^0\rangle$

$$P_i=|\Psi_i^0\rangle\langle\Phi_i|, \tag{1.13}$$

where $P_i|\Psi_i^0\rangle\equiv|\Psi_i^0\rangle$. Since

$$\begin{aligned}P_iH^0|\Psi_i^{(1)}\rangle&=\frac{|\Psi_i^0\rangle}{f_i^P}\sum_{p=0}^P\langle\Phi_i^0|H_p^0|\Phi_i^p\rangle\langle\Phi_i^p|\Psi_i^{(1)}\rangle\\&=E_i^0\frac{|\Psi_i^0\rangle}{f_i^P}\sum_{p=0}^P\langle\Phi_i^0|\Phi_i^p\rangle\langle\Phi_i^p|\Psi_i^{(1)}\rangle\\&=E_i^0|\Psi_i^0\rangle\langle\Phi_i^0|\Psi_i^{(1)}\rangle=E_i^0P_i\Psi_i^{(1)},\end{aligned} \tag{1.14}$$

after the operator (1.13) is applied to Eq. (1.11) we get $\langle\Phi_i|V|\Psi_i^0\rangle|\Psi_i^0\rangle=E_i^{(1)}|\Psi_i^0\rangle$. This leads to an expression for the first-order correction to the energy,

$$E_i^{(1)}=\langle\Phi_i|V|\Psi_i^0\rangle. \tag{1.15}$$

Now let us introduce the projector on the subspace of vectors orthogonal to $|\Psi_i^0\rangle$ in accordance with the property of the double vector product:

$$O_i=1-P_i,\quad \text{where } O_i|\Psi_i^0\rangle\equiv0 \tag{1.16}$$

or, to put it differently,

$$O_i|\Psi\rangle=\langle\Phi_i|\times|\Psi\rangle\times|\Psi_i^0\rangle, \tag{1.17}$$

i.e., $|\Psi\rangle=|\Psi_i^0\rangle\langle\Phi|\Psi\rangle+\langle\Phi_i|\times|\Psi\rangle\times|\Psi_i^0\rangle$.

From (1.13) and (1.17) we see that the antisymmetric basis of zeroth wave functions is actually only weakly nonorthogonal,

$$P_i|\Psi_n^{(0)}\rangle = |\Psi_i^0\rangle\langle\Phi_i^0|\Psi_n^0\rangle$$

$$= |\Psi_i^0\rangle \frac{1}{f_i^P} \sum_{p=0}^{P} \langle\Phi_i^0|\Phi_n^p\rangle(-1)^{g_p} \approx 0, \qquad (1.18)$$

because $\langle\Phi_i^0|\Phi_n^0\rangle = 0$, and $\langle\Phi_i^0|\Phi_i^p\rangle \approx 0$, since the overlap of the wave functions of the ground and excited belonging to different centers is insignificant (this resembles the situation in diffraction theory for wave optics). Accordingly,

$$\langle\Phi_i| \times |\Psi_n^0\rangle \times |\Psi_i^0\rangle \approx |\Psi_n^0\rangle. \qquad (1.19)$$

We seek, in accordance with (1.12), the first-order correction to the antisymmetrized function of the zeroth approximation in the form of an expansion:

$$\Psi_i^{(1)} = \sum_n C_n \Psi_n^0, \qquad (1.20)$$

where $n \neq i$. Inserting the expansion (1.20) in (1.11) and applying the operator (1.16) to the result, we get

$$O_i V|\Psi_i^0\rangle = \sum_n C_n (E_i^0 - E_i^n)|\Psi_n^0\rangle, \qquad (1.21)$$

where we have allowed for (1.18) and (1.19).

Using the property of completeness of the orthogonal basis of the nonsymmetric zeroth functions,

$$\sum_n |\Phi_n^p\rangle\langle\Phi_n^p| = 1,$$

we can write the left-hand side of Eq. (1.21) as follows:

$$\frac{1}{P} \sum_{p=0}^{P} \sum_n |\Phi_n^p\rangle\langle\Phi_n^p|O_i V|\Psi_i^0\rangle$$

$$= \frac{1}{P} \sum_{p=0}^{P} \sum_n |\Phi_n^p\rangle(-1)^p\langle\Phi_n^0|O_i V|\Psi_i^0\rangle$$

$$= \frac{1}{P} \sum_n f_n^P |\Psi_n^0\rangle\langle\Phi_n^0|O_i V|\Psi_i^0\rangle. \qquad (1.22)$$

Since

$$\langle\Phi_i|O_i V|\Psi_i^0\rangle = \langle\Phi_i|1 - |\Psi_i\rangle\langle\Phi_i|V|\Psi_i^0\rangle \equiv 0,$$

we must drop the term with $n = i$ from (1.22).

Finally, Eq. (1.21) becomes

$$\frac{1}{P} \sum_n f_n^P |\Psi_n^0\rangle\langle\Phi_n^0|O_i V|\Psi_i^0\rangle = \sum_n C_n (E_i^0 - E_n^0)|\Psi_n^0\rangle, \qquad (1.23)$$

from which we find the first-order correction to the wave function:

$$\Psi_i^{(1)} = \frac{1}{P} \sum_n f_n^P \frac{\langle\Phi_n^0|O_i V|\Psi_i^0\rangle}{(E_i^0 - E_n^0)} |\Psi_n^0\rangle. \qquad (1.24)$$

Higher-order corrections can be found similarly.

2. THE CASE OF DEGENERACY

The zeroth wave function for a multicenter system can be antisymmetrized by various Young tableaux, which differ for different values of the total spin of the system. In other words, a multicenter system of noninteracting electrons is degenerate in total spin, with the degeneracy lifted by allowing for ordinary intercenter interaction. Thus,

$$\Psi_{n\alpha}^0 = A_\alpha \Phi_n^0,$$

$$H_\alpha^0 \Psi_{n\alpha}^0 = E_n^0 \Psi_{n\alpha}^0, \quad H_\alpha^0 = \sum_{p=0}^{P} \frac{1}{f_n^\alpha} H_p^0 \Lambda_n^p, \qquad (2.1)$$

where $\{\Psi_{n\alpha}^0\}$ is the set of wave functions antisymmetrized by different Young tableaux α and corresponding to the same energy level E_n^0 of the system.

We seek the wave function of an interacting multicenter system of electrons in the form

$$\Psi_i = \sum_\beta C_\beta^0 \Psi_{\beta i}^0 + \varphi. \qquad (2.2)$$

If we substitute (2.2) in the complete Schrödinger equation (1.9), we obtain

$$H \sum_\beta C_\beta^0 \Psi_\beta^0 + H\varphi = (E_i^0 + \varepsilon) \sum_\beta C_\beta^0 \Psi_\beta^0 + (E_i^0 + \varepsilon)\varphi.$$

Since the total Hamiltonian is invariant under all permutations, it can be taken outside the summation sign in accordance with the Young tableau:

$$\sum_\beta (H_\beta^0 + V_\beta) C_\beta^0 \Psi_\beta^0 + H\varphi$$

$$= (E_i^0 + \varepsilon) \sum_\beta C_\beta^0 \Psi_\beta^0 + (E_i^0 + \varepsilon)\varphi. \qquad (2.3)$$

Using (2.1), we can shift all the terms containing φ to the left-hand side of Eq. (2.3) and all the other terms to the right-hand side. The result is

$$H\varphi - (E_i + \varepsilon) = \sum_\beta (\varepsilon - V_\beta) C_\beta^0 \Psi_\beta^0. \qquad (2.4)$$

In (2.4) we drop all terms whose order is higher than the first. This means that in the total-energy operator acting on φ we must leave only the unperturbed part symmetrized by an arbitrary Young tableau γ and drop the term $\varepsilon\varphi$. Then

$$(H_\gamma^0 - E_i^0)\varphi = \sum_\beta (\varepsilon - V_\beta) C_\beta^0 \Psi_\beta^0. \qquad (2.5)$$

The solution of the homogeneous analog of Eq. (2.5) for φ is $\varphi = \Psi_\gamma^0$. But then, according to the Fredholm alternative,[21] the nonhomogeneous problem (2.5) has a solution only if the vector $\langle\Psi_\gamma^0|$ is orthogonal to the entire right-hand side:

$$\sum_\beta (\varepsilon\langle\Psi_\gamma^0|\Psi_\beta^0\rangle - \langle\Psi_\gamma^0|V_\beta|\Psi_\beta^0\rangle) C_\beta^0 = 0. \qquad (2.6)$$

Thus, we have a system of equations for determining the coefficients C_β^0 of the regular zeroth wave function. The system has a solution only if

$$|\varepsilon\Delta_{\gamma\beta} - \langle\Psi_\gamma^0|V|\Psi_\beta^0\rangle| = 0. \qquad (2.7)$$

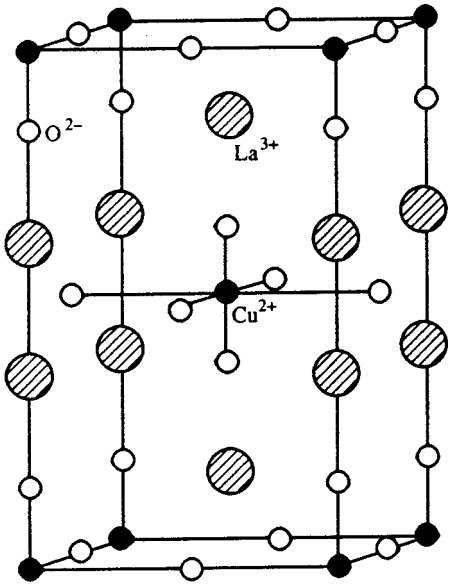


FIG. 1. Crystalline and magnetic structures of La_2CuO_4.

This is the secular equation for determining the corrections to the energy. If

$$\Delta_{\gamma\beta}=\langle\Psi^0_\gamma|\Psi^0_\beta\rangle=\frac{1}{f^\gamma_0}\sum_{p=0}^{P}(-1)^{g_p}\langle\Phi^{p\gamma}|\Psi^0_\beta\rangle$$
$$=\sum_{p=0}^{P}(-1)^{g_{p\gamma}}(-1)^{g_{p\beta}}\langle\Phi^0|\Psi^0_\beta\rangle$$
$$=\frac{1}{f^\gamma_0}\sum_{p=0}^{P}(-1)^{g_{p\gamma}+g_{p\beta}},$$

$$\langle\Psi^0_\gamma|V_\beta|\Psi^0_\beta\rangle=\frac{1}{f^\gamma_0}\sum_{p\gamma=0}^{P}(-1)^{g_{p\gamma}}\langle\Phi^{p\gamma}|V_\beta|\Psi^0_\beta\rangle$$
$$=\frac{1}{f^\gamma_0}\sum_{p\gamma=0}^{P}(-1)^{g_{p\gamma}+g_{p\beta}}\langle\Phi^0|V_\beta|\Psi^0_\beta\rangle$$
$$=\Delta_{\gamma\beta}\langle\Phi^0|V_\beta|\Psi^0_\beta\rangle, \tag{2.8}$$

the secular equation becomes

$$\prod_{\beta=1}(\varepsilon-\langle\Phi^0|V_\beta|\Psi^0_\beta\rangle)|\Delta_{\gamma\beta}|=0. \tag{2.9}$$

The corrections ε to the energy have definite values, $\varepsilon=\langle\Phi^0|V_\alpha|\Psi^0_\alpha\rangle$, only if

$$|\Delta_{\alpha\beta}|\neq 0. \tag{2.10}$$

Then the set of zeroth wave functions antisymmetrized by Young tableaux is regular.

3. THE HEISENBERG PARAMETER FOR THE HIGH-T_c MATERIALS $La_{2-x}Me_xCuO_4$ AND $Ba_2Cu_3O_{6+x}$

The crystalline structure of the compound La_2CuO_4 in the stoichiometric state is depicted in Fig. 1. The crystal of copper dioxide alloyed with, say, strontium ($La_{2-x}Sr_xCuO_4$) has a similar lattice: a body-centered tetragonal structure whose space group is $14/mmm$. NMR and muon-precession experiments[22–24] have shown that the antiferromagnetic state occurs in this material due to the interaction of Cu^{2+} ions lying in a single plane, while the interplanar magnetic interaction is weak. The magnetic form factor of the Cu^{2+} ion measured in the antiferromagnetic state[25] corresponds to the $3d^9$ state. The O^{2-} ion occupying a position between interacting copper ions does not affect this interaction because the electronic shell is filled.

The wave function of the electrons of a pair of interacting Cu^{2+} ions in the zeroth approximation corresponds to states with total spin $S=1$ or $S=0$, i.e., its spatial part is antisymmetric or symmetric, respectively. Then, from the secular equation (2.9), for the singlet and triplet states we have

$$\varepsilon_{\text{sing,tr}}=\frac{K\pm A}{1\pm I^2},$$

where

$$\Psi_{s,a}(r_1,r_2)=\frac{1}{1\pm I^2}[\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)+\psi_{\rm I}(r_2)\psi_{\rm II}(r_1)],$$

$$\Delta_{ss}=2,\quad \Delta_{aa}=2,\quad \Delta_{sa}=\Delta_{as}=0\rightarrow\begin{vmatrix}2&0\\0&2\end{vmatrix}\neq 0,$$

$$K=\langle\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)|V_1|\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)\rangle,$$

$$A=\langle\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)|V_2|\psi_{\rm I}(r_2)\psi_{\rm II}(r_1)\rangle, \tag{3.1}$$

$$V_1=\frac{z^2e^2}{|R_{\rm I}-R_{\rm II}|}-\frac{ze^2}{|r_1-R_{\rm II}|}-\frac{ze^2}{|r_2-R_{\rm I}|}+\frac{e^2}{|r_1-r_2|},$$

$$V_2=\frac{ze^2}{|R_{\rm I}-R_{\rm II}|}-\frac{ze^2}{|r_1-R_{\rm I}|}-\frac{ze^2}{|r_2-R_{\rm II}|}+\frac{e^2}{|r_1-r_2|},$$

where $R_{\rm I,II}$ are the radii of the interacting ions, z stands for the ion charges, $r_{1,2}$ are the radius vectors of the electrons, and I is the one-electron exchange density.

The wave function of an electron belonging to the Cu^{2+} ion corresponds to the $3d^9_{x^2-y^2}$ orbital and is chosen in the hydrogenlike form

$$\psi(r)=\sqrt{\frac{15}{16\pi}\frac{2^7}{6!}}\,r^2e^{-r}\sin^2\vartheta\cos 2\varphi \tag{3.2}$$

(in Bohr units). Then the Heisenberg parameter for the lattice constant $R=7.3346a_B$ is

$$j=\varepsilon_{\rm sing}-\varepsilon_{\rm tr}=-0.1043278\ \text{eV}. \tag{3.3}$$

In the given case, j is negative, which corresponds to antiparallel orientation of the spins at the neighboring lattice sites as being an energetically preferable configuration.

Similar calculations of j that use Eqs. (3.1) and (3.2) for the Cu^{2+} ions in adjacent layers ($R=12.3818a_B$) yield the value

$$j=-9.673\times10^{-5}\ \text{eV}, \tag{3.4}$$

which also corresponds to antiparallel orientation of the spins, but the coupling constant is very small.

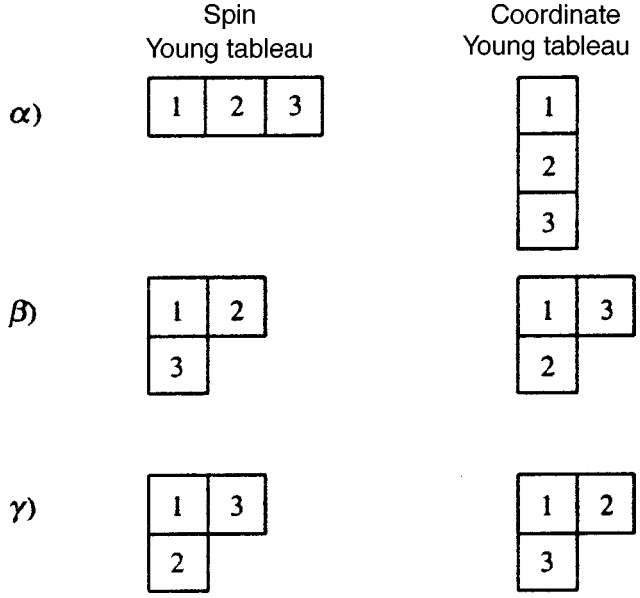
Thus, the assumption that the antiferromagnetic interaction between the planes is small, which was made in the spin models of Refs. 12 and 13, is justified, and we indeed are dealing with a 2D antiferromagnetic system.

The same experiments[22–24] show that alloying with, say, strontium ($La_{2-x}Sr_xCuO_4$), causes a rapid decrease in the Néel temperature in proportion to the alloying degree x. The point is that alloying with a doubly charged ion of a metal activates the oxygen ion O^{2-} positioned between copper ions, which becomes a single charged ion O^{1-} with an unpaired electron. Now, in studying the interaction of two copper ions, we must allow for the presence of an additional electron belonging to the O^{1-} ion, whose state, being highly delocalized, strongly overlaps with the electronic states of the Cu^{2+} ions.

The complete electron wave function of the ion chain Cu^{2+}–O^{1-}–Cu^{2+} (for the sake of brevity we denote this chain by I–II–III) can be antisymmetrized in three different ways corresponding to the following spin configurations:

(I↑II↑III↑), (I↑II↑III↓), (I↑II↓III↑).

The corresponding Young tableaux are given below:



(3.5)

At the beginning, the electrons with numbers 1, 2, and 3 belong to the ions I, II, and III, respectively. In this case the coordinate parts of the wave functions corresponding to the Young tableaux (α), (β), and (γ) are

$$\Psi^0_\alpha(r_1,r_2,r_3)=\frac{1}{1-2I_1^2+2I_2^2I_1^2-I_2^2}\begin{vmatrix}\psi_{\rm I}(r_1) & \psi_{\rm I}(r_2) & \psi_{\rm I}(r_3)\\ \psi_{\rm II}(r_1) & \psi_{\rm II}(r_2) & \psi_{\rm II}(r_3)\\ \psi_{\rm III}(r_1) & \psi_{\rm III}(r_2) & \psi_{\rm III}(r_3)\end{vmatrix},$$

$$\Psi^0_\beta(r_1,r_2,r_3)=\frac{1}{1-2I_2^2-I_1^2-I_2I_1^2}\begin{vmatrix}\psi_{\rm I}(r_1)\psi_{\rm III}(r_3)+\psi_{\rm III}(r_1)\psi_{\rm I}(r_3) & \psi_{\rm I}(r_1)\psi_{\rm II}(r_2)+\psi_{\rm II}(r_2)\psi_{\rm I}(r_1)\\ \psi_{\rm III}(r_3) & \psi_{\rm II}(r_2)\end{vmatrix},$$

$$\Psi^0_\gamma(r_1,r_2,r_3)=\frac{1}{1-I_2I_1^2}\begin{vmatrix}\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)+\psi_{\rm I}(r_2)\psi_{\rm II}(r_1) & \psi_{\rm I}(r_1)\psi_{\rm III}(r_3)+\psi_{\rm I}(r_3)\psi_{\rm III}(r_1)\\ \psi_{\rm II}(r_2) & \psi_{\rm III}(r_3)\end{vmatrix}, \tag{3.6}$$

where we have allowed for the normalization condition (1.4) and have introduced the following notation for the integrals corresponding to the exchange densities:

$$I_1=\int\psi^*_{\rm I}(r)\psi_{\rm II}(r)dr=\int\psi^*_{\rm II}(r)\psi_{\rm III}(r)dr,$$

$$I_2=\int\psi^*_{\rm I}(r)\psi_{\rm III}(r)dr. \tag{3.7}$$

For the initial distribution of the numbered electrons over the centers, the unsymmetrized perturbation operator is

$$V=\frac{z_1z_2e^2}{|R_{\rm I}-R_{\rm II}|}+\frac{z_1^2e^2}{|R_{\rm I}-R_{\rm III}|}+\frac{z_1z_2e^2}{|R_{\rm II}-R_{\rm III}|}-\frac{z_2e^2}{|r_1-R_{\rm II}|}-\frac{z_1e^2}{|r_1-R_{\rm III}|}-\frac{z_1e^2}{|r_2-R_{\rm I}|}-\frac{z_1e^2}{|r_2-R_{\rm III}|}-\frac{z_1e^2}{|r_3-R_{\rm I}|}-\frac{z_2e^2}{|r_3-R_{\rm II}|}+\frac{e^2}{|r_1-r_2|}+\frac{e^2}{|r_2-r_3|}+\frac{e^2}{|r_1-r_3|}. \tag{3.8}$$

Direct calculation of the parameter in (2.8) yields the following values:

$$\Delta_{\alpha\beta}=\Delta_{\alpha\gamma}=0,\quad \Delta_{\beta\gamma}=\frac{2}{1-I_2I_1^2},$$

$$\Delta_{\gamma\gamma}=\frac{4}{1-I_2I_1^2},\quad \Delta_{\alpha\alpha}=\frac{6}{1-2I_1^2(I_1-I_2)-I_2^2},$$

$$\Delta_{\beta\beta}=\frac{4}{1-I_1^2+I_2^2-I_2I_1^2}.$$

The determinant condition (2.10) has the form

$$\begin{vmatrix}\Delta_{\alpha\alpha} & 0 & 0\\ 0 & \Delta_{\beta\beta} & \Delta_{\beta\gamma}\\ 0 & \Delta_{\gamma\beta} & \Delta_{\gamma\gamma}\end{vmatrix}\neq 0. \tag{3.9}$$

Thus, condition (2.10) is met. In this case and in accordance with (2.9), the corrections to the energy are

$$\varepsilon_1=\langle\Phi^0(r_1r_2r_3)|\hat{V}_\alpha|\Psi^0_\alpha\rangle,$$

$$\varepsilon_2=\langle\Phi^0(r_1r_2r_3)|\hat{V}_\beta|\Psi^0_\beta\rangle,$$

$$\varepsilon_3=\langle\Phi^0(r_1r_2r_3)|\hat{V}_\gamma|\Psi^0_\gamma\rangle, \tag{3.10}$$

where

$$|\Phi^0(r_1r_2r_3)\rangle=|\psi_{\rm I}(r_1)\psi_{\rm II}(r_2)\psi_{\rm III}(r_3)\rangle.$$

Note that in contrast to the two-center case, superexchange three-center integrals of the form

$$K_{\rm I\to II,II\to III}=\int\int\frac{\psi^*_{\rm I}(\mathbf{r}_1)\psi^*_{\rm II}(\mathbf{r}_2)\psi_{\rm II}(\mathbf{r}_1)\psi_{\rm III}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\,d^3r_1d^3r_2 \tag{3.11}$$

contribute substantially to all the expressions in (3.10). Such integrals ensure three-center correlation of the spins, since they can enter into in the general expressions (3.10) for the energy with different signs. The sign sequence of these superexchange terms is determined by the Young diagrams used in antisymmetrizing the wave function in (3.10). Due to this superexchange interaction, long-range order may set in the system without the conduction electrons participating, as is the case in the Kondo and RKKY models.

Calculations of the matrix elements in (3.10) for the given lattice constant $R=3.88$ Å with allowance for the superexchange contributions (see Appendix B) yield

$$\varepsilon_1(\uparrow\uparrow\uparrow)=-0.7628577E_B,$$

$$\varepsilon_2(\uparrow\downarrow\downarrow)=-0.6705097E_B, \tag{3.12}$$

$$\varepsilon_3(\uparrow\downarrow\uparrow)=-0.6391877E_B.$$

In the given case of Cu^{2+} ions, the orientation of the spins is ferromagnetic, with the Heisenberg parameter being

$$J=\varepsilon_2-\varepsilon_1\approx 0.092348E_B=2.51288\ \text{eV}. \tag{3.13}$$

Comparison with (3.3) yields the following value of the parameter ratio:

$$\frac{J}{|j|}=25. \tag{3.14}$$

Thus, alloying the material, which activates the oxygen ions O^{1-}, does indeed reorient the electron spins in Cu^{2+} and leads to strong ferromagnetism. Hence, as Birgeneau and Schirane[25] pointed out by analyzing the experimental facts, a strong ferromagnetic bond in the CuO_2 plane destroys the local antiferromagnetic order. In the case of strong localization, the concentration of Φ-bonds would be equal to x. As x grows, the localization length l_0 of each hole increases, which leads to an increase in the effective concentration of Φ-bonds. Birgeneau and Schirane[25] found that a large value of $J/|j|$ reduces the threshold values of x at which antiferromagnetism in alloyed La_2CuO_4 disappears even if x is small.[26]

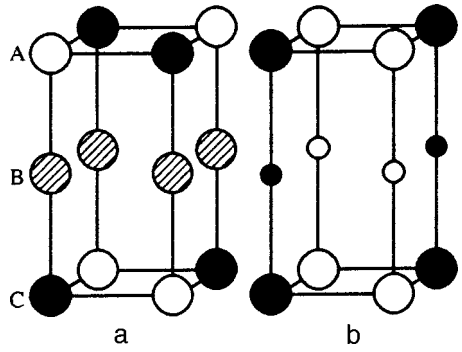


FIG. 2. (a) Magnetic spin structure of $YBa_2Cu_3O_{6+x}$ with $x=0$. Only copper atoms are depicted. The hatched circles stand for the nonmagnetic Cu^{1+} ions, the dark and light circles stand for the antiparallel spins in the Cu^{2+} positions, and the solid straight lines denote double bonds with oxygen atoms. (b) The second type of spin structure observed for large values of x. The average spin in layer B is the fraction ε of the spin in a CuO_2 layer.

The experimental data on another copper dioxide, $YBa_2Cu_3O_{6+x}$ suggests a remarkable resemblance between the two systems.[4,27] The sublattice in $YBa_2Cu_3O_6$ is depicted in Fig. 2. Each chemical cell contains two CuO_2 layers, denoted by A and C. The copper atoms in layer B have no bonds with oxygen. X-ray absorption measurements have clearly shown that the valence of the copper atoms in layer B is $1+$, so that the atoms are nonmagnetic.

To represent the spin structure of the Cu^{2+} ions in layers A and C, the antiparallel spins are depicted by dark and light circles. The calculation of the corresponding parameter j is done by (3.1)–(3.3) with a lattice constant $R=7.225a_B=3.822$ Å:

$$j_{\rm AA}=\varepsilon_{\rm sing}-\varepsilon_{\rm tr}=-0.0935\ \text{eV}$$

inside a layer, and

$$j_{\rm AC}=\varepsilon_{\rm sing}-\varepsilon_{\rm tr}=-0.0713\ \text{eV}$$

between layers A and C. An estimate of the same parameter done by analyzing the experimental data on the Néel temperature yielded $j\approx 0.086$ eV. X-ray absorption measurements have shown that adding oxygen facilitates the transition of Cu^{1+} into Cu^{2+}. Now the Cu^{2+} in layer B facilitates the destruction of the antiferromagnetic order between neighboring Cu^{2+} ions in layers A and C.

Consider a system of three Cu^{2+} ions, with the layers A, B, and C each having one ion. Then, using Eqs. (3.5)–(3.10), we can calculate the energy corrections to the spin configuration in the Young diagram:

$$\varepsilon_1(\uparrow\uparrow\uparrow) = -0.33847E_B,$$

$$\varepsilon_2(\uparrow\downarrow\downarrow) = -0.33027E_B,$$

$$\varepsilon_3(\uparrow\downarrow\uparrow) = -0.34666E_B.$$

Thus, the states with ferromagnetic orientation of the spins in layers A and B are the most probable, the spin in layer B may be assumed fluctuating, and the complete ferromagnetic state has a Heisenberg parameter

$$J_1 = \varepsilon_2 - \varepsilon_1 = 8.2\times 10^{-3}E_g = 0.2230 \text{ eV},$$

while the state of the oppositely directed spin in layer B has a Heisenberg parameter

$$J_2 = 16.4\times 10^{-3}E_B = 0.446 \text{ eV}.$$

The average value of the magnetic moment per magnetic atom is

$$m = \frac{3\mu_0 + \mu_0}{2\times 3} = \frac{2}{3}\mu_0 = 0.6(6)\mu_0.$$

The experiment of Birgeneau *et al.*[28] gives the same value for the maximally ordered moment.

CONCLUSION

We have developed a variant of exchange perturbation theory that allows for degeneracy in total spin, with the applicability criteria (1.18) and (2.10) added. Using it we have derived a procedure for *ab initio* calculations of the Heisenberg parameter for high-T_c materials, a procedure that is based on first principles and avoids computer simulations, so that the necessary relationships are obtained in analytical form. Numerical estimates of the energy values for given lattice parameters yield results that are in good agreement with those of experimental and phenomenological approaches. For instance, for La_2CuO_4 the experiment of Peters *et al.*[29] yields the following values of the Heisenberg parameter.

For the antiferromagnetic interaction of the ions in Cu^{2+}–Cu^{2+} lying in a single plane, $|J_1| = 0.116$ eV; between the planes, $|J_2| \approx 2\times 10^{-6}$ eV.

For the ferromagnetic Cu^{2+}–O^-–Cu^{2+}, interaction $|J_1| \approx 1$ eV; our values of the corresponding parameters are $|J_1| = 0.104$ eV, $|J_2| \approx 9\times 10^{-5}$ eV, and $|J| \approx 2.5$ eV.

For $YBa_2Cu_3O_6$, the Heisenberg energy estimated using an analysis of the Néel temperature is $|J| \approx 0.086$ eV (Refs. 24 and 27); our values are $|J| \approx 0.0935$ eV inside a layer and $|J| \approx 0.07$ eV between layers.

Note that the following simplifications were made in specific calculations:

1) Intratomic interaction is incorporated only indirectly by using one-electron states, whose parameters are taken from the Gombás and Szondy tables.[30]

2) The effect of the mean crystalline field on the intercenter interaction of *d*-electrons is ignored; the latter, however, incorporates the interaction with the nearest-neighbor ions.

3) The influence of conduction electrons on the magnetic effects due to the orientation of spins of *d*-electrons is ignored.

4) Finally, spin–orbit coupling effects is ignored.

In conclusion we note that the analytical expressions for the energy of a three-center chain of atoms contains (as the leading contribution) the nonadditive contribution of the three-center interaction (see Appendix B). The terms are K_{1321}, K_{1232}, K_{2321}, K_{1323}, K_{2313}, K_{1231}, K_{2312}, and K_{1332}. Such integrals determine the structure, since they depend not only on the intercenter distance but also on the angles between the straight lines connecting these centers. We believe that they are responsible for exchange-correlation effects in solids, including spin systems.

We are grateful to V. V. Rumyantsev for his attention to our work and to his critical remarks, which enabled us to significantly improve the presentation in this paper. The work was made possible by the financial support of the Russian Academy of Sciences (Young Scientist Stipend of the Russian Academy of Sciences) given to one of the coauthors.

APPENDIX A: COMPLETENESS PROPERTY OF THE NONORTHOGONAL BASIS OF ANTISYMMETRIC FUNCTIONS

States that are antisymmetric under particle permutations are nonorthogonal. Nevertheless, they constitute complete system. To verify this, we act with the operator $\Sigma_n|\Phi_n\rangle\langle\Psi_n|$ on an arbitrary function antisymmetric according to the same Young diagram as the states $|\Psi_n\rangle$:

$$\sum_n |\Phi_n\rangle\langle\Psi_n|\Psi\rangle = \sum_n \sum_{p=0}^{P} |\Phi_n\rangle\langle\Phi_n^p|\Psi\rangle(-1)^{g_p}\frac{1}{f_n}, \tag{A1}$$

where

$$f_n = \sum_{p=0}^{P} \langle\Phi_n|\Phi_n^p\rangle(-1)^{g_p}.$$

Using the antisymmetry of the vector $|\Psi\rangle$, we can write

$$\sum_n \sum_{p=0}^{P} \frac{1}{f_n}|\Phi_n\rangle\langle\Phi_n^p|\Psi\rangle = \sum_n \sum_{p=0}^{P} \frac{1}{f_n}|\Phi_n\rangle\langle\Phi_n|\Psi\rangle(-1)^{g_p}$$

$$= \sum_{p=0}^{P} \frac{1}{f_0}|\Psi\rangle = \frac{P}{f_0}|\Psi\rangle, \tag{A2}$$

where we have used the fact that $f_n = f_0 = \Sigma\langle\Phi_0|\Phi_0^p\rangle(-1)^{g_p}$ and the completeness property of an orthogonal basis of nonsymmetric functions, $\Sigma_n|\Phi_n\rangle\langle\Phi_n| = 1$.

Thus, we have

$$\frac{f_0}{P}\sum_n |\Phi_n\rangle\langle\Psi_n| = \sum_n |\Phi_n\rangle\langle\Phi_n|, \tag{A3}$$

which is simply the completeness property of the system. Using (A3), we can decompose an arbitrary antisymmetrized state in the same antisymmetric states.

In deriving the completeness property we assumed that $f_0/f_n \approx 1$. Let us now estimate the smallness of the terms discarded. Clearly, the above ratio can be written as

$$\frac{f_0}{f_n}=1+\sum_{p=0}^{P}(-1)^{g_p}[\langle\Phi_0|\Phi_0^p\rangle-\langle\Phi_n|\Phi_n^p\rangle]+O(I^{2P}), \tag{A4}$$

where

$$\mathrm{Supr}\left(\sum_{p}^{P}(-1)^{g_p}[\langle\Phi_0|\Phi_0^p\rangle-\langle\Phi_n|\Phi_n^p\rangle]\right) \approx\frac{N!}{g_P!(N-g_P)!}\,\frac{I_0-I_n}{1+I_0+I_n},$$

I_0 and I_n are still the one-electron exchange densities of the ground and excited states, and N and g_P are the respective total number of electrons and the number of electrons participating in the intercenter permutation.

Since the overlap integrals are, in general, the product of a polynomial, $P(R/a_B)$, and an exponential, $\exp(-R/a_B)$, the discarded terms are

$$\mathrm{Supr}\left[\frac{N!}{(N-g_P)!g_P!}\,P\left(\frac{R}{a_B}\right)\exp\left(-\frac{R}{a_B}\right)\right]=\mathrm{const}<1. \tag{A5}$$

This constant can, in principle, be accounted for in (A3).

APPENDIX B: EXPRESSIONS FOR OBTAINING THE CORRECTIONS TO THE ENERGY

1. The energies of a two-center interaction are

$$E_{\mathrm{sing}}=\frac{e^2}{1+I^2}\left[\frac{z_1^2}{R}(1+I^2)+2z_1C_{11}-2z_2C_{12}+K_{1212}+K_{1221}\right], \tag{B1}$$

$$E_{\mathrm{tr}}=\frac{e^2}{1-I^2}\left[\frac{z_1^2}{R}(1-I^2)-2z_1C_{11}+2z_2C_{12}+K_{1212}-K_{1221}\right], \tag{B2}$$

where we have introduced the following notation:

$$C_{ij}=\int\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\frac{d^3r}{|\mathbf{r}-\mathbf{R}|}$$

is the direct or exchange interaction of an electron with a ''foreign'' nucleus, i and j label the nucleus,

$$K_{ijkl}=\int\frac{\psi_i^*(\mathbf{r}_1)\psi_k^*(\mathbf{r}_2)\psi_j(\mathbf{r}_1)\psi_l(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}\,d^3r_1d^3r_2$$

is the exchange interaction of two electrons distributed between the nuclei i,j,k,l, and

$$K_{ijij}=\int\frac{|\psi_i(\mathbf{r}_1)|^2|\psi_j(\mathbf{r}_2)|^2}{|\mathbf{r}_1-\mathbf{r}_2|}d^3r_1d^3r_2$$

is the direct Coulomb interaction of two electrons centered at different nuclei i and j.

2. The energies of the three-center interaction are

$$E_1=\frac{1}{6}\Delta_{\alpha\alpha}(K_0-K_{2-3}-K_{1-2}+K_{1-2,2-3}+K_{1-3,1-2}-K_{1-3}), \tag{B3}$$

$$E_2=\frac{1}{4}\Delta_{\beta\beta}(K_0+K_{1-3}-K_{2-3}-K_{1-2,2-3}), \tag{B4}$$

$$E_3=\frac{1}{4}\Delta_{\gamma\gamma}(K_0+K_{1-2}-K_{2-3}-K_{1-2,1-3}), \tag{B5}$$

where

$$K_0=\frac{z_1z_2}{R}+\frac{z_1^2}{R}-z_2(B_{11}-B_{33})-z_1(C_{11}-C_{22}+S_{22}+S_{33}), \tag{B6}$$

$$K_{1-2}=\left(\frac{z_1^2}{R}+\frac{z_1z_2}{R}\right)I_1^2-(z_2I_1B_{21}+z_1I_1C_{21}+z_1I_1S_{12} +I_1z_1C_{12}+z_1|I_1|^2S_{33}+I_1^2z_2B_{33})z_1+K_{1221} +K_{2313}I_1+K_{1323}I_1, \tag{B7}$$

and K_{2-3} can be obtained from (B7) by cyclic permutation of the subscripts 1, 2, and 3.

Similarly, for K_{1-3} we have

$$K_{1-2,2-3}=\left(\frac{z_1^2}{R}+\frac{z_1z_2}{R}\right)I_1^2I_2-(z_2I_1I_2B_{23}+z_1I_1I_2C_{32} +z_1I_1^2S_{13}+I_1^2z_1C_{13}+z_1I_1I_2S_{21} +I_1I_2z_2B_{21})z_1+K_{1231}I_1+K_{2312}I_2+K_{1332}I_1. \tag{B8}$$

Here we have used the integrals

$$B_{ij}=\int\frac{\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{R}|}d^3r,\quad S_{ij}=\int\frac{\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})}{|\mathbf{r}|}d^3r$$

complementary to (B1) and (B2).

*)E-mail: quark@stu.neva.ru

[1] E. C. Stoner, Proc. R. Soc. London, Ser. A **165**, 372 (1938).
[2] J. Hubbard, Proc. R. Soc. London, Ser. A **276**, 238 (1963).
[3] W. Heitler and F. London, Z. Phys. **44**, 455 (1927).
[4] J. M. Tranquada, A. H. Moudden, A. I. Goldman, P. Zolliker, D. E. Cox, G. Shirane, S. K. Sinha, D. Vaknin, D. C. Johnston, M. S. Alvarez, A. J. Jacobson, J. T. Lewandowski, and J. M. Newsam, Phys. Rev. B **38**, 2477 (1988).
[5] A. V. Kulakov, E. V. Orlenko, and A. A. Rumyantsev, *Quantum Exchange Forces in Condensed Media* [in Russian], Nauka, Moscow (1989).
[6] P. J. Hay, J. C. Thibeault, and R. Hoffmann, J. Am. Chem. Soc. **97**, 4884 (1975).
[7] K. Yamada, E. Kudo, Y. Endoh *et al.*, Solid State Commun. **64**, 753 (1987).
[8] T. Freltoft, J. E. Fisher, G. Shirane *et al.*, Phys. Rev. B **36**, 826 (1987).
[9] D. C. Johnston, J. P. Stokes, D. P. Goshorn, and J. T. Lewandowski, Phys. Rev. B **36**, 4007 (1987).
[10] L. F. Mattheiss, Phys. Rev. Lett. **58**, 1028 (1987).
[11] J. Yu, A. J. Freeman, and J.-H. Xu, Phys. Rev. Lett. **58**, 1035 (1987).
[12] P. W. Anderson, G. Baskaran, Z. Zou, and T. Hsu, Phys. Rev. Lett. **58**, 2790 (1987).
[13] I. G. Kaplan, *Theory of Intermolecular Interactions* Elsevier, New York (1986) [Russ. orig., Nauka, Moscow (1982)]; P. W. Anderson, G. Baskaran, Z. Zou, and T. Hsu, Phys. Rev. Lett. **58**, 2790 (1987).
[14] A. Van der Avoird, J. Chem. Phys. **47**, 3649 (1967).
[15] R. M. Sternheimer, Phys. Rev. **96**, 951 (1954).
[16] L. Jansen, Phys. Rev. **162**, 63 (1967).
[17] A. B. Ritchie, Phys. Rev. **171**, 125 (1968).

[18] A. A. Rumyantsev, Zh. Éksp. Teor. Fiz. **65**, 926 (1973) [Sov. Phys. JETP **38**, 459 (1974)].
[19] E. V. Orlenko and A. A. Rumyantsev, Zh. Éksp. Teor. Fiz. **97**, 439 (1990) [Sov. Phys. JETP **70**, 244 (1990)].
[20] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[21] G. Korn and T. Korn, *Mathematical Handbook*, 2nd ed., McGraw-Hill, New York (1968).
[22] K. Kumagai, I. Watanabe, H. Aoke *et al.*, Physica B **148**, 460 (1987).
[23] T. Takahashi, F. Maeda, H. Arai *et al.*, Physica B **148**, 476 (1987).
[24] Y. Budnic *et al.*, eul **5**, 647 (1988).
[25] R. J. Birgeneau and G. Schirane, in *Physical Properties of High-Temperature Superconductors*, Vol. 1, D. M. Ginsberg (Ed.), World Scientific, Singapore (1989).
[26] J. M. Tranquada, P. E. Cox, W. Kunnmann, H. Moudden, G. Shirane, M. Suenaga, P. Zolliker, D. Vaknin, S. K. Sinha, M. S. Alvarez, A. J. Jacobson, and D. C. Johnston, Phys. Rev. Lett. **60**, 156 (1988).
[27] W.-H. Li, J. W. Lynn, H. A. Mook, B. C. Sales, and Z. Fisk, Phys. Rev. B **37**, 9844 (1988).
[28] R. J. Birgeneau, J. Skayo, and G. Schirane, Phys. Rev. B **3**, 1736 (1971); R. J. Birgeneau, J. Als-Nielsen, and G. Schirane, Phys. Rev. B **16**, 280 (1977).
[29] C. J. Peters, R. J. Birgeneau, M. A. Kastner, H. Yoshizawa, Y. Endoh, J. Tranquada, G. Shirane, Y. Hidaka, M. Oda, M. Suzuki, and T. Murakami, Phys. Rev. B **37**, 9761 (1988).
[30] P. Gombás and T. Szondy, *Solution of the Simple Self-Consistent Field for Atoms*, Acad. Kiodo, Budapest (1970).

Translated by Eugene Yankovsky

On the current-voltage characteristic of a Josephson junction with a high critical current density

V. P. Silin*) and A. V. Studenov

P. N. Lebedev Physical Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 16 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 2148–2161 (June 1998)

A theory is constructed for the current-voltage characteristic of a Josephson junction with a high critical current density described by nonlocal vortex electrodynamics in the approximation of small dissipation due to the single-particle conductivity of the Josephson junction. The role of the resonant excitation of generalized short-wavelength Swihart waves is determined. New quantitative properties of the current-voltage characteristic, which is an envelope of wave excitation resonances, are established. © *1998 American Institute of Physics.* [S1063-7761(98)01706-5]

1. INTRODUCTION

The theory of the current-voltage characteristics (IVC's) of Josephson junctions is of great interest specifically because the experimental study of IVC's permits the comparatively simple investigation of the manifestations of vortex structures in such junctions. The theory of the IVC's of Josephson junctions is ordinarily based on the sine-Gordon equation with dissipation. In Ref. 1 a theory of IVC's was formulated for a vortex structure in the form of a traveling wave put into motion by a spatially uniform current and decelerated by dissipation due to the single-particle conductivity of a Josephson junction. The results of the theory of IVC's needed for annular junctions, which have recently attracted the attention of experimentalists, are presented in Ref. 2. Annular junctions are of interest, because there are no edge effects complicating the vortex pattern in the case of such Josephson junctions. This is the reason why the present work is devoted to annular junctions. At the same time, in contrast to Refs. 1–3, our analysis is devoted to the theory of Josephson junctions with a high critical current density, where[4]

$$j_c \gg j_0 = \frac{\hbar c^2}{16\pi |e| \lambda^3} = 1.24 \cdot 10^4 (\lambda[\mu\mathrm{m}])^{-3}\ \mathrm{A/cm^2} \quad (1.1)$$

and the standard theory based on the sine-Gordon equation with dissipation is inapplicable. In Eq. (1.1) λ is the London penetration depth. When the inequality (1.1) holds, the London depth is greater than the Josephson penetration depth

$$\lambda \gg \lambda_j = \lambda \sqrt{\frac{j_0}{j_c}} \gg \frac{\lambda_j^2}{\lambda} = l, \quad (1.2)$$

where e is the electron charge and c is the speed of light.

Josephson electrodynamics is strongly nonlocal in this limit. The annular junction geometry which we have chosen is illustrated in Fig. 1 (compare Ref. 3). The tunnel junction width $2d$ is assumed to be small compared to the London depth λ, while the latter is assumed to be much less than the thickness a of the annular superconducting cylinders. Assuming the radius R of the tunnel junction to be much greater than the London depth λ ($0 < z < 2\pi R$, where z is the coordinate along the annular tunnel junction) and continuing the electric and magnetic fields periodically along the z axis onto the entire number axis $(-\infty, +\infty)$, for the phase difference φ between the wave functions of the Cooper pairs on different sides of the tunnel junction the we employ equation[4–6]

$$\frac{1}{\omega_j^2}\frac{\partial^2 \varphi}{\partial t^2} + \frac{\beta}{\omega_j^2}\frac{\partial \varphi}{\partial t} + \sin\varphi - \frac{l}{\pi}\int_{-\infty}^{+\infty} \frac{dz'}{z'-z}\frac{\partial \varphi}{\partial z'} + \gamma = 0. \quad (1.3)$$

Here $\gamma = j/j_c$ is the normalized short-circuit current density through the Josephson junction, $\omega_j = (16\pi |e| d j_c/\hbar\varepsilon)^{1/2}$ is the Josephson plasma frequency, $\beta = 4\pi\sigma/\varepsilon$ characterizes the dissipation, and ε and σ are the permittivity and conductivity of the tunnel junction material. Our analysis is based on Eq. (1.3).

The existing results for the IVC of a Josephson junction with a high critical current density satisfying the inequality (1.2) were obtained in the resistive model, where the second derivative with respect to time in Eq. (1.3) can be neglected.[5–9] The exact solution describing an infinite chain of vortices having a period $2\pi L$ and moving with a constant velocity v in accordance with the resistive model in Ref. 7 makes it easier to compare the results of those works. In this case

$$\varphi(z,t) = -\theta + \pi + 2\tan^{-1}\left(\frac{\tan[(z-vt)/2L]}{\tanh[\alpha_0/2]}\right), \quad (1.4)$$

where

$$\frac{L^2}{l^2}\cos^2\theta = \sinh^2\alpha_0 = \left(\frac{1}{4}\left[\gamma^2 + \frac{l^2}{L^2} - 1\right]^2 + \frac{l^2}{L^2}\right)^{1/2} + \frac{1}{2}\left(\gamma^2 + \frac{l^2}{L^2} - 1\right), \quad (1.5)$$

and the IVC is given by the expression

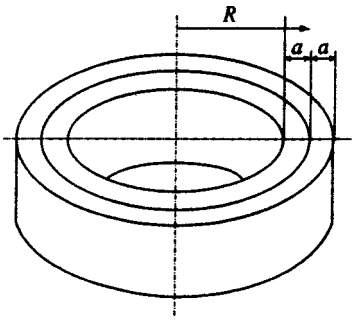


FIG. 1. Two concentric superconducting cylinders separated by a thin tunnel junction.

$$\frac{V}{j_c R_s}=\frac{\beta v}{\omega_j^2 L}=\left\{\left[\frac{1}{4}\left(\frac{j^2}{j_c^2}+\frac{l^2}{L^2}-1\right)^2+\frac{l^2}{L^2}\right]^{1/2}+\frac{1}{2}\left(\frac{j^2}{j_c^2}+\frac{l^2}{L^2}-1\right)\right\}^{1/2}. \tag{1.6}$$

Here $R_s=2d/\sigma$ is the resistance per unit area of the tunnel contact, and V is the time-averaged potential difference across the junction. The IVC in the form (1.6) corresponds to the case when the potential difference is determined by an assigned current density. This formula can also be expressed in a form[8] in which the current is determined by an assigned potential:

$$j^2=j_r^2+\frac{j_r^2 j_c^2 L^2}{L^2 j_r^2+l^2 j_c^2},\quad \text{where}\quad j_r=\frac{V}{R_s}. \tag{1.7}$$

The latter form, in contrast to Eq. (1.6), corresponds less directly to the experimental data, but it does facilitate comparisons with the results obtained by our predecessors. Let us examine Ref. 5 first. The $I-V$ curve obtained there as a result of a numerical investigation is qualitatively similar to the plot constructed in Ref. 8 on the basis of Eq. (1.7). Further, in Ref. 5 the relation

$$j=\frac{L}{l}\,j_r \tag{1.8}$$

is indicated for weak currents. According to Eq. (1.7) the relation (1.8) holds for

$$\frac{V}{R_s j_c}\ll\frac{l}{L}\ll 1. \tag{1.9}$$

This corresponds to a narrower region than that indicated in Ref. 5. The second asymptotic formula of Ref. 5 describes convergence to Ohm's law:

$$j=j_r\left\{1+\frac{L^2 j_c^2}{2(L^2 j_r^2+l^2 j_c^2)}\right\}. \tag{1.10}$$

This formula holds for $V\gg j_c R_s$, as follows from Eq. (1.7) and as was indicated in Ref. 5.

The IVC of Ref. 6 is based on the use of the exact solution of the resistive model for a single vortex moving with constant velocity:

$$\varphi(z,t)=-\theta+\pi+2\tan^{-1}\left(\frac{z-vt}{\rho}\right), \tag{1.11}$$

where $\theta=\arcsin\gamma$, $\rho=l/(1-\gamma^2)^{1/2}$, and

$$v=\frac{l\omega_j^2}{\beta}\,\frac{\gamma}{\sqrt{1-\gamma^2}}. \tag{1.12}$$

Assuming that the solution of the resistive model can be written in an approximation in the form of a chain of periodically arranged vortices (1.11), Gurevich[6] obtained an IVC of the form (1.7), but without the first term on the right-hand side. This means that the result obtained in Ref. 6 in the case $L\gg l$ is realized, according to Eq. (1.7), for the low voltages $V\ll j_c R_s$, which correspond to currents much higher than the ohmic current. At the same time, according to Eq. (31) of Ref. 6 the current density should be less than the critical Josephson current density [compare Eq. (1.12)]. Our discussion makes it possible to see the heretofore not discussed consistency of the results obtained in the resistive model, especially for weak currents, which is important for what follows. We note here that to obtain the IVC of an annular Josephson junction with radius R from Eq. (1.7), the parameter L in Eq. (1.7) should be replaced by R/m, where m is a positive integer corresponding to the number of Abrikosov–Josephson vortices in the ring (see Refs. 8 and 9).

In the case of small dissipation of interest to us, i.e., in the limit opposite to that obtained in the resistive model, the dissipation is determined by the emission of generalized Swihart waves. The deceleration arising in the process for a single Abrikosov–Josephson vortex was studied in Ref. 10. In the case when such a vortex moves with a low velocity

$$v\ll l\omega_j \tag{1.13}$$

short waves with wavelengths less than the London depth are excited, and the spectrum of generalized Swihart waves has the form (compare Ref. 11)

$$\omega^2(k)=\omega_j^2(1+|k|l), \tag{1.14}$$

where k is the wave vector. According to the condition for Cherenkov radiation, the wave vector of the disturbances excited can be determined from the equation

$$\omega^2(k)=k^2v^2. \tag{1.15}$$

In contrast to Ref. 10, to construct a theory for IVC's, below we shall employ the exact stationary solution which was obtained in Ref. 12 for Eq. (1.3) with $\gamma=0$ and describes a chain of Abrikosov–Josephson vortices:

$$\varphi(z)=\pi+2\tan^{-1}\left(\frac{\tan[z/2L]}{\tanh[\alpha/2]}\right), \tag{1.16}$$

$$\sinh\alpha=(l/L). \tag{1.17}$$

We assume, as in Ref. 10, that the structure (1.16) moves with a low velocity v. The wave disturbance produced by such a motion of a single vortex in a ring when $L=R$ is obtained in Sec. 2. The friction force acting on the moving vortex structure together with its disturbance field due to the dissipation caused by the finite conductivity σ of the quasiparticles passing through the tunnel junction is determined in

Sec. 3. The balancing of the friction force by the Lorentz force due to the current flowing through the Josephson junction is studied in Sec. 4. This makes it possible to determine the relationship between the velocity of the vortex structure and the current. A generalization to the case of m Abrikosov–Josephson vortices in an annular junction is given in Sec. 5. The IVC is obtained in Sec. 6. Section 7 is devoted to a discussion of the results.

2. MOVING VORTEX STRUCTURE

In this section we assume that a single vortex ($L=R$) moving with a low velocity satisfying the inequality (1.13) is present in an annular Josephson junction with radius R. At the same time, we assume that the dissipation is small, whence it follows that the rate of such motion should not be very low. Thus,

$$R\beta \ll v \ll l\omega_j . \tag{2.1}$$

We assume that the ring is large compared with an Abrikosov–Josephson vortex:

$$R \gg l. \tag{2.2}$$

We seek the perturbed state arising during the motion of the chain of vortices (1.16), in the form

$$\varphi(z,t)=\varphi_0(z-vt)+\varphi_1(z,t), \tag{2.3}$$

where φ_0 is specified by formula (1.16) and

$$\varphi_1 \ll \varphi_0 . \tag{2.4}$$

The substitution of (2.3) into Eq. (1.3) with $\gamma=0$ followed by linearization gives

$$\frac{1}{\omega_j^2}\frac{\partial^2\varphi_1}{\partial t^2}+\frac{\beta}{\omega_j^2}\frac{\partial\varphi_1}{\partial t}+\varphi_1\cos\varphi_0-\frac{l}{\pi}\int_{-\infty}^{+\infty}\frac{dz'}{z'-z}\frac{\partial\varphi_1}{\partial z'} = -\frac{v^2}{\omega_j^2}\frac{d^2\varphi_0}{ds^2}+\frac{\beta v}{\omega_j^2}\frac{d\varphi_0}{ds}, \tag{2.5}$$

where $s=z-vt$. The term $\varphi_1\cos\varphi_0$ on the left-hand side of Eq. (2.5) is comparatively small, and its retention is important only for simplifying the mathematical operations. For this reason, just as in Ref. 10, we replace it by φ_1. It is easy to see that the following stationary solution in the form of a structure traveling with velocity v can then be written:

$$\varphi_1(z,t)=\varphi_1(s)=\sum_{n=1}^{\infty}\exp\left(-\frac{nl}{R}\right)\left[\left(\frac{n\beta v}{R}\right)^2+\left(\omega_n^2-\frac{n^2v^2}{R^2}\right)^2\right]^{-1}\left\{\frac{2v^2}{R^2}\left[\frac{n\beta v}{R}\cos\left(\frac{ns}{R}\right)+\left(\omega_n^2-\frac{n^2v^2}{R^2}\right)\sin\left(\frac{ns}{R}\right)\right]+\frac{2\beta v}{R}\times\left[\left(\omega_n^2-\frac{n^2v^2}{R^2}\right)\cos\left(\frac{ns}{R}\right)-\frac{n\beta v}{R}\sin\left(\frac{ns}{R}\right)\right]\right\}+\frac{\beta v}{R\omega_j^2}, \tag{2.6}$$

where

$$\omega_n=\omega_j\sqrt{1+\frac{nl}{R}} \tag{2.7}$$

corresponds to Eq. (1.14) for the spectrum of generalized Swihart waves if $k=n/R$. The last term in Eq. (2.6) is due to the contribution

$$\left\langle\frac{d\varphi_0}{ds}\frac{\beta v}{\omega_j^2}\right\rangle=\frac{\beta v}{R\omega_j^2} \tag{2.8}$$

to the right-hand side of Eq. (2.5). Here the angle brackets denote averaging over a period. By virtue of the left-hand inequality in Eq. (2.1) we neglect the second term in the curly brackets in Eq. (2.6) in comparison with the first term.

Since β is small, it is obvious that Eq. (2.6) describes a resonance dependence on the velocity. The number n_r of the resonance term in the sum (2.6) and the resonance value v_r of the velocity are related as

$$\omega_{n_r}^2=\omega_j^2\left(1+\frac{n_rl}{R}\right)=\left(\frac{n_rv_r}{R}\right)^2. \tag{2.9}$$

Accordingly,

$$n_r=\frac{lR\omega_j^2}{v_r^2} \tag{2.10}$$

and according to Eq. (2.2) and the right-hand side of Eq. (2.1) we have

$$n_r \gg \frac{R}{l} \gg 1. \tag{2.11}$$

The latter justifies replacing $\varphi_1\cos\varphi_0$ by φ_1 when resonance effects are examined. As a result, the resonance term with $v=v_r$ can be written as

$$\varphi_r(z-v_rt)=\frac{2v_r}{\beta R}\exp\left(-\frac{l^2\omega_j^2}{v_r^2}\right)\cos\left[\frac{l\omega_j^2}{v_r^2}(z-v_rt)\right]. \tag{2.12}$$

This expression is distinguished from that arising in the theory of the Cherenkov radiation of a single Abrikosov–Josephson vortex[10] only by the preexponential factor. This difference is due to the fact that Eq. (2.12) takes into account the dissipation in an annular Josephson junction.

Next, we note that, owing to the left-hand inequality in (2.1), the condition

$$\left(\frac{n\beta v}{R}\right)^2 \ll \left(\omega_n^2-\frac{n^2v^2}{R^2}\right)^2. \tag{2.13}$$

is satisfied for $v\approx v_r$ and $n\neq n_r$. This allows us to neglect β in all the nonresonance terms in Eq. (2.6) and thereby leads to the following nonresonance contribution to $\varphi_1(s)$:

$$\varphi_{nr}(s)=\frac{2v^2}{R^2}\sum_{\substack{n=1\\ n\neq n_r}}^{\infty} n\left\{\omega_n^2-\left(\frac{nv}{R}\right)^2\right\}^{-1}\times\exp\left\{-\frac{nl}{R}\right\}\sin\left(\frac{ns}{R}\right). \tag{2.14}$$

The last expression can be simplified using the inequalities (2.11) and neglecting the contribution that is obviously small compared with the resonance contribution. Then

$$\varphi_{nr}\approx\frac{2}{n_r}\sum_{n=1}^{\infty}\frac{n}{n+(R/l)}\exp\left(-\frac{nl}{R}\right)\sin\left(\frac{ns}{R}\right). \tag{2.15}$$

Here we also call attention to the relative smallness of the distance between neighboring resonance values of the velocity. Indeed, according to Eq. (2.9), we have

$$\Delta v_r=\frac{v_r}{2n_r}=v_r\frac{l}{2R}\left(\frac{v_r^2}{l^2\omega_j^2}\right)\ll v_r. \tag{2.16}$$

Then, away from resonance,

$$2|v-v_r|<\Delta v_r\ll v_r, \tag{2.17}$$

it can be shown that Eq. (2.13) is satisfied, and we can write

$$\varphi_1(s)=\frac{v_r}{R}\left[\frac{\beta^2}{4}+\frac{n_r^2(v-v_r)^2}{R^2}\right]^{-1}\exp\left(-\frac{l^2\omega_j^2}{v_r^2}\right)\times\left\{\frac{\beta}{2}\cos\left(\frac{n_rs}{R}\right)-\frac{n_r(v-v_r)}{R}\sin\left(\frac{n_rs}{R}\right)\right\}+\varphi_{nr}(s)+\frac{\beta v}{R\omega_j^2}. \tag{2.18}$$

Let us now focus on the realization of the approximation (2.4). First, the right-hand inequality in Eq. (2.1) gives the required smallness of φ_{nr}. Second, the inequality

$$\exp\left(\frac{l^2\omega_j^2}{v_r^2}\right)\gg\frac{2v_r}{\beta R} \tag{2.19}$$

ensures that the resonance contribution is small compared to φ_0. We shall also utilize this inequality below to satisfy the condition (2.4).

3. FRICTION FORCE

Dissipation ($\beta\neq 0$) results in energy loss. The rate of decrease in the energy of a vortex structure in an annular Josephson junction is given by the expression (compare, for example, Ref. 13)

$$\frac{dE}{dt}=-\frac{\hbar j_c}{2|e|}\frac{\beta}{\omega_j^2}\int_0^{2\pi R}dz\left(\frac{\partial\varphi}{\partial t}\right)^2. \tag{3.1}$$

Here E is the energy per unit length of the Josephson junction.

Using the expression obtained for $\varphi_0+\varphi_1$ in the preceding section, we can express the energy dissipation rate as

$$\frac{dE}{dt}=-\frac{v^2\beta}{\omega_j^2}\frac{\pi\hbar j_c}{|e|l}\left\{1+\frac{l^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v_r]^2)}{(\beta/2)^2+[n_r(v-v_r)/R]^2}\right\}. \tag{3.2}$$

The first term in curly brackets on the right-hand side corresponds to the energy loss by the moving periodic structure φ_0, and the second term corresponds to the energy loss by the resonantly excited wave $\varphi_r(s)$ (2.18).

The energy loss by a vortex structure is often attributed to the friction force $f_{\rm fr}$ (see, for example, Ref. 14), which is specified by the relation

$$f_{\rm fr}=\frac{1}{v}\frac{dE}{dt}. \tag{3.3}$$

Accordingly, it follows from Eq. (3.2) that

$$f_{\rm fr}=-\frac{v\beta}{\omega_j^2}\frac{\pi\hbar j_c}{|e|l}\left\{1+\frac{l^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v_r]^2)}{(\beta/2)^2+[n_r(v-v_r)/R]^2}\right\}. \tag{3.4}$$

The friction force and relatively small energy losses by a vortex structure can be discussed when

$$\beta\tau\ll 1, \tag{3.5}$$

where $\tau=2\pi R/v$ is the rotation period of the vortex structure in the ring. This condition is satisfied according to Eq. (2.1).

4. LORENTZ FORCE AND THE EQUILIBRIUM CONDITION

To solve the problem of the IVC of a Josephson junction we now assume that a constant spatially uniform current with density j flows through the junction. As a result, the current acts on the vortex structure, exerting the following a Lorentz force[13] per unit length:

$$f_L=\frac{1}{c}j\Phi, \tag{4.1}$$

where Φ is the magnetic flux through the Josephson junction. More specifically (see Ref. 13),

$$\Phi=\frac{\Phi_0}{2\pi}[\varphi(z=2\pi R,t)-\varphi(z=0,\,t)], \tag{4.2}$$

where $\Phi_0=\pi\hbar c/|e|$ is the magnetic flux quantum. Since according to Eq. (2.6) the disturbance φ_1 is a periodic function, the flux (4.2) is determined by φ_0. Therefore, according to Eq. (1.16) with $L=R$, as was assumed in Sec. 2, we have $\Phi=\Phi_0$. In this case the Lorentz force equals

$$f_L=\frac{1}{c}j\Phi_0. \tag{4.3}$$

In the case of stationary motion of our vortex structure, the constant value of v is determined from the condition that the accelerating Lorentz force (4.3) is balanced by the decelerating friction force (3.4). This condition leads to the relation

$$\frac{j}{j_c}=\frac{\beta v}{\omega_j^2 l}\left\{1+\frac{l^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v_r]^2)}{(\beta/2)^2+[n_r(v-v_r)/R]^2}\right\} \tag{4.4}$$

between the current density j and the velocity v of the vortex structure. We stress that, according to the conditions (2.1) and (2.19), the right-hand side of Eq. (4.4) is small compared to 1, i.e., $j\ll j_c$.

We note that Eq. (4.4) can be obtained on the basis of the energy relation following from Eq. (1.3)

$$\frac{dE}{dt}=\frac{\hbar j_c}{2|e|}\int_0^{2\pi R}dz\left(-\frac{\beta}{\omega_j^2}\left(\frac{\partial\varphi}{\partial t}\right)^2-\gamma\frac{\partial\varphi}{\partial t}\right), \tag{4.5}$$

where E is the energy of the Josephson junction. Equation (4.5) describes the variation of the energy in time due to dissipation (friction) and the accelerating action of the current ($\gamma = j/j_c$). Substituting the expression (2.3) from Sec. 2 into (4.5) likewise leads to Eq. (4.4) provided the energy remains constant ($dE/dt = 0$).

5. MULTIVORTEX STRUCTURE

In this section we give a generalization made before this analysis to the case of several vortices in an annular Josephson junction. Generalization to the case of m vortices is accomplished by replacing L by R/m in Eqs. (1.16) and (1.17). Since it was assumed in Secs. 2–4 that $L = R$, the required generalization is made by replacing R by R/m in the equations of these sections. It should be noted that the conditions (2.1) and (2.2) for the applicability of our analysis become

$$(\beta R/m) \ll v \ll l\omega_j \quad \text{and} \quad ml \ll R. \tag{5.1}$$

Accordingly, the number of the Cherenkov resonance is given by the formula

$$n_r = \frac{Rl\omega_j^2}{mv_r^2} \gg \frac{R}{ml} \gg 1. \tag{5.2}$$

The analog of the condition (2.19) has the form

$$\exp\left(\frac{l^2\omega_j^2}{v_r^2}\right) \gg \frac{2mv_r}{\beta R}. \tag{5.3}$$

These conditions make it possible to generalize Eq. (3.4) so as to describe the friction force acting on a chain of m Abrikosov–Josephson vortices, whose motion also causes the excitation of generalized Swihart waves. For the friction force we have

$$f_{\text{fr}} = -\frac{mv\beta}{\omega_j^2}\frac{\pi\hbar j_c}{|e|l} \times\left\{1+\frac{ml^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v_r]^2)}{(\beta/2)^2+[mn_r(v-v_r)/R]^2}\right\}. \tag{5.4}$$

Since we now have $f_L = (m/c)j\Phi_0$ for the Lorentz force, as opposed to Eq. (4.3), we obtain the following generalization of the relation (4.4) from the equilibrium condition:

$$\frac{j}{j_c} = \frac{\beta v}{\omega_j^2 l}\left\{1+\frac{ml^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v_r]^2)}{(\beta/2)^2+[mn_r(v-v_r)/R]^2}\right\}. \tag{5.5}$$

In this case, in analogy to Eq. (2.17), the following condition is satisfied:

$$2|v-v_r| < \Delta v_r = \frac{v_r}{2n_r} = v_r\frac{ml}{2R}\frac{v_r^2}{l^2\omega_j^2} \ll v_r, \tag{5.6}$$

where Δv_r is the distance between neighboring resonances. Fulfillment of the conditions (5.6) and (5.2) permits the replacement of v_r in the exponential function in Eq. (5.5) by v. Indeed,

$$\frac{|\exp[-2(l\omega_j/v)^2]-\exp[-2(l\omega_j/v_r)^2]|}{\exp[-2(l\omega_j/v_r)^2]} = |\exp[4(l\omega_j/v_r)^2(v-v_r)/v_r]-1| < \exp[2(l\omega_j/v_r)^2\Delta v_r/v_r]-1 = (ml/R) \ll 1.$$

Thus, the relation (5.5) can be written in the form

$$\frac{j}{j_c} = \frac{\beta v}{\omega_j^2 l}\left\{1+\frac{ml^3\omega_j^4}{2v^2R}\frac{\exp(-2[l\omega_j/v]^2)}{(\beta/2)^2+[mn_r(v-v_r)/R]^2}\right\}. \tag{5.7}$$

This relation between the current density and the velocity of the vortex structure allows us to write down an expression for the IVC sought, which we shall do in the next section.

6. CURRENT-VOLTAGE CHARACTERISTIC

To obtain the IVC, we employ the definition of a static potential difference

$$V = -\frac{\hbar}{2|e|}\left\langle\frac{\partial\varphi}{\partial t}\right\rangle, \tag{6.1}$$

where the averaging extends over the length of the ring, which in our case corresponds to averaging over time. In our case only $\varphi_0(z-vt)$ makes a nonzero contribution on averaging. It turns out that

$$V = \hbar mv/2|e|R.$$

The latter formula makes it possible to rewrite the relation (5.7) in the form of an IVC

$$\frac{j}{j_c} = \frac{V}{R_s j_c}\frac{R}{lm}\left\{1+\frac{2m^3(R_s j_1/V)^2\exp(-2m^2V_0^2/V^2)}{1+[(V/V_s)^2(V-V_r)^2/(\Delta V_r)^2]}\right\}, \tag{6.2}$$

where the notation

$$\Delta V_r = V_r/2n_r \tag{6.3}$$

has been used for the distance between neighboring resonance voltage values V_r on the IVC:

$$V_r[\mu\text{V}] = \frac{\hbar v_r m}{2|e|R} = \frac{\hbar c}{2|e|\lambda}\sqrt{\frac{md}{R\varepsilon n_r}} = \frac{100}{R}\sqrt{\frac{mdR}{\lambda^2\varepsilon n_r}}, \tag{6.4}$$

$$V_0[\mu\text{V}] = \frac{\hbar l\omega_j}{2|e|R} = \frac{\hbar c}{2|e|R}\sqrt{\frac{d}{\varepsilon\lambda}\frac{j_0}{j_c}} = \frac{100}{R}\sqrt{\frac{d}{\lambda\varepsilon}\frac{j_0}{j_c}}. \tag{6.5}$$

In the latter numerical expressions R is measured in millimeters. Finally,

$$j_1 = j_c\left(\frac{\lambda}{R}\frac{j_0}{j_c}\right)^{3/2}, \quad V_s = \frac{2\pi\hbar\sigma}{|e|\varepsilon} = \frac{\hbar}{2|e|R_sC_s}. \tag{6.6}$$

We note that $j_1/V_0 = (c/8\pi\lambda)\sqrt{\varepsilon/dR}$ and does not depend on j_c. In the resonance denominator of Eq. (6.2) $|V-V_r| < (1/2)\Delta V_r$. In accordance with Eqs. (5.1) and (5.3) the IVC (6.2) holds for values of V that satisfy the inequalities

$$V_s \ll V \ll mV_0, \tag{6.7}$$

$$\exp\left(\frac{m^2V_0^2}{V^2}\right) \gg 2\frac{V}{V_s}. \tag{6.8}$$

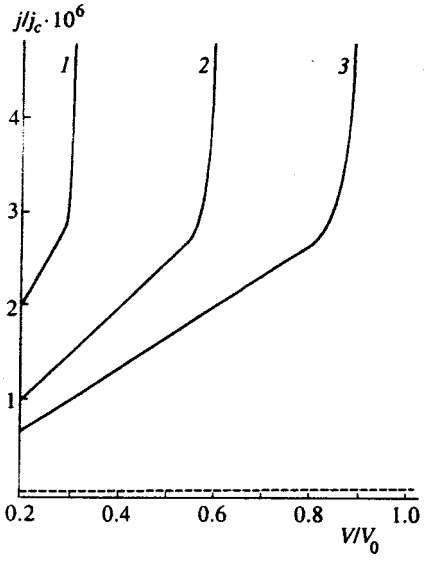


FIG. 2. Current-voltage characteristic of a Josephson junction with $m=1,2,3$ and $l/R=10^{-2}$ $(j_1R_s/V_0)=10^4$.

According to Eq. (6.2), the width δV_r of a resonance at half-height is given by

$$\delta V_r=\frac{V_s}{V}\,\Delta V_r. \tag{6.9}$$

In accordance with the inequality (6.7), this means that the width of a resonance is much less than the distance between neighboring resonances. This suggests that individual resonances are distinguishable. On the other hand, according to the definition (6.3) and the inequality (5.2), the resonances are very close together. Therefore, if the resolution accuracy ΔV_r in the experiment is inadequate, the envelope of the curve (6.2), which is described by the formula

$$\frac{j}{j_c}=\frac{V}{j_cR_s}\,\frac{R}{lm}\left\{1+2m^3\,\frac{R_s^2j_1^2}{V^2}\exp\left(-2m^2\,\frac{V_0^2}{V^2}\right)\right\}, \tag{6.10}$$

will be observed.

7. DISCUSSION

Figure 2 shows the IVC (6.10) for several values of the number of vortices in the Josephson junction $(m=1,2,3)$. This figure serves to illustrate our discussion. We begin our discussion with low voltages, for which the IVC's correspond to linear segments of the curves. Such linear segments correspond to a balance between the acceleration by the current and the deceleration by the dissipation of the main periodic vortex structure φ_0, when the occurrence of resonance losses is hardly observed. Therefore, the lower bound on the voltage in Eq. (6.7) is not important for such linear segments of the IVC. These linear segments correspond to the asymptotic result of Ref. 5, according to which the superconducting current through a Josephson junction is (R/lm) times greater than the ohmic current $j_r=V/R_s$ [see Eq. (1.8)]. In Fig. 2 the dashed line corresponds to the ohmic current. In the resistive approximation of Refs. 5 and 7, as the voltage increases, the rate of growth of the current slows according to Eq. (1.7), and the IVC approaches Ohm's law according to Eq. (1.10), which was obtained in Ref. 5.

In contrast to the resistive approximation, our analysis leads to increasingly rapid growth of the current with increasing voltage, rather than to slowing. This qualitative dependence is similar to the known dependence for a Josephson junction with a low critical density and small dissipation, for which theories are being devised on the basis of the standard sine-Gordon equation with friction. We make the required comparison using the result obtained in Ref. 1 for the IVC of a Josephson junction with small dissipation under the condition, just as in our case, that the size of a single vortex is much smaller than the period of the vortex structure in the Josephson junction:

$$\frac{j}{j_c}=\frac{4V}{\pi j_cR_s}\,\frac{R}{lm}\,\frac{\sqrt{j_0/j_c}}{\sqrt{1-(V/mV_0)^2(j_0/j_c)}}. \tag{7.1}$$

We note that on the linear segment of the IVC at low voltages this formula gives a current which differs by a factor of $(2/\pi)\sqrt{j_0/j_c}$ from the result obtained in the theory of Josephson junctions whose critical current density satisfies the condition (1.1). However, another property of (7.1) is extremely important to us. According to this property, as the voltage approaches the value

$$V\to mV_0\sqrt{j_c/j_0} \tag{7.2}$$

the standard IVC describes a sharp increase in current. For our IVC (6.10) there is likewise a sharp increase in current corresponding to a deviation of the curves in Fig. 2 from the linear segments in the low-voltage region. However, the analogy between the IVC's (6.10) and (7.1) ends with this qualitative similarity. Therefore, we should now discuss the differences.

First, the functional dependence of the upward deviation of the IVC from the continuation of the linear segment differs. In contrast to the comparatively smooth approach of (7.1) to the square-root singularity, which is not observed experimentally, because the IVC is cut off at a state with a large number of vortices in the Josephson junction, in our case Eq. (6.10) predicts a sharper exponential growth of the current with increasing voltage. This qualitative difference can be observed experimentally (compare Ref. 15).

The second difference between Eq. (6.10) and Eq. (7.1) is that for a Josephson junction with a high critical current density (1.1) rapid growth is predicted when

$$V\sim\frac{mV_0}{\sqrt{\ln[mR_s^2j_1^2/V_0^2]}}\equiv\frac{mV_0}{\sqrt{\ln[m\varepsilon(d/R)(c/4\pi\lambda\sigma)^2]}} \tag{7.3}$$

rather than when the voltage approaches the value specified by (7.2). In accordance with the right-hand inequality in Eq. (6.7), this value should be less than mV_0. This is possible if

$$\sqrt{m\,\frac{\varepsilon d}{R}}\gg\frac{4\pi\sigma\lambda}{c}\sim 4\cdot 10^{-14}(\lambda[\mu\mathrm{m}])(\sigma[\mathrm{s}^{-1}]), \tag{7.4}$$

which is readily realized. On the other hand, our approach to finding the resonance correction φ_r requires satisfaction of

the left-hand inequality in Eq. (6.7), where the expression (7.3) should now be used instead of the voltage. The following inequality then arises:

$$m\sqrt{\frac{\varepsilon d\lambda}{R^2}\frac{j_0}{j_c}}\,\frac{1}{\sqrt{\ln[m\varepsilon(d/R)(c/4\pi\lambda\sigma)^2]}}\gg\frac{4\pi\sigma\lambda}{c}. \tag{7.5}$$

When m is of order unity, the left-hand side of this inequality is much smaller than the left-hand side of Eq. (7.4). Therefore, the inequality (7.5) is the main constraint on the conductivity of a Josephson junction. For example, for $\varepsilon=10$, $d/R=10^{-6}$, and $j_c=5j_0$ the inequality (7.5) takes the form

$$\frac{m\sqrt{\lambda/R}\cdot 10^{11}}{\sqrt{\ln[\sqrt{m}\cdot 10^{11}/\lambda[\mu\mathrm{m}]\sigma[\mathrm{s}^{-1}]]}}\gg 4\lambda[\mu\mathrm{m}]\sigma[\mathrm{s}^{-1}],$$

which does not place excessively severe constraints on the conductivity if we assume that $\lambda/R\sim 10^{-2}-10^{-3}$.

We note that since the ring-averaged intensity of the constant magnetic field is given by the relation

$$\overline{H}=m\Phi_0/4\pi\lambda R,$$

Eqs. (6.2) and (6.10) can be used to describe the dependence of the IVC on the magnetic field.

8. CONCLUSIONS

We have constructed the first theory for the IVC of an annular Josephson junction with a high critical current density[1)] (1.1). This theory describes the exponential voltage dependence for an envelope of segments corresponding to the excitation of resonant states, as well as the fine structure due to such resonances.

To provide a better understanding of the importance of the material presented in this paper, we note that the exponential dependence of the current on the velocity of a solitary Josephson junction in the theory of Josephson junctions with a low critical current density was pointed out in Ref. 17. Our analysis leads to a substantial quantitative difference, and even the physical essence of the phenomenon is different. First, the exponential effect described in Ref. 17 corresponds to vortex velocities close to the ordinary Swihart velocity $c_s=\lambda_j\omega_j$, while the effect discussed in the present paper corresponds to the velocity $l\omega_j$, which is $\sqrt{j_0/j_c}$ times smaller than the Swihart velocity. Accordingly, the exponential increase in the current in Fig. 2 corresponds to a voltage $\sqrt{j_0/j_c}$ times lower. This difference is due to our condition (1.1), which is new compared with Ref. 7. Second, in Ref. 17 the dissipation is due to the Cherenkov excitation of generalized Swihart waves, whose energy escapes to infinity from a vortex, i.e., it is emitted from the Josephson junction, whereas in our case a resonantly excited Cherenkov wave does not escape in any direction from the Josephson junction, but is absorbed in the junction as a result of dissipation. This causes the preexponential factor in our IVC (6.10) to depend on the resistance of the Josephson junction, which, of course, cannot happen in the approach described in Ref. 17. Finally, we stress that no IVC was obtained in Ref. 17.

This work was supported by the Scientific Council on High-T_c Superconductivity (Project AD No. 95008) and as part of Project No. 96-02-17303 of the Russian Fund for Fundamental Research.

*)E-mail: silin@sci.lpi.ac.ru

1)Reference 16 reports a critical current density equal to 2×10^4 A/cm^2 for Josephson junctions.

[1] P. M. Marcus and Y. Imry, Solid State Commun. **33**, 345 (1980).
[2] A. Shnirman, Z. Hermon, A. V. Ustinov, B. A. Malomed, and E. Ben-Jacob, Phys. Rev. B **50**, 12 795 (1994).
[3] Z. Hermon, A. Stern, and E. Ben-Jacob, Phys. Rev. B **49**, 9757 (1994).
[4] Yu. M. Aliev, G. L. Alfimov, K. N. Ovchinnikov, V. P. Silin, and S. A. Uryupin, Fiz. Nizk. Temp. **22**, 626 (1996) [Low Temp. Phys. **22**, 477 (1996)].
[5] M. Yu. Kupriyanov, K. K. Likharev, and A. K. Semenov, Fiz. Nizk. Temp. **2**, 706 (1976) [Sov. J. Low Temp. Phys. **2**, 346 (1976)].
[6] A. Gurevich, Phys. Rev. B **48**, 12 857 (1993).
[7] V. P. Silin, JETP Lett. **60**, 460 (1994).
[8] V. P. Silin and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **108**, 2163 (1995) [JETP **81**, 1179 (1995)].
[9] G. L. Alfimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **108**, 1668 (1995) [JETP **81**, 915 (1995)].
[10] V. P. Silin and A. V. Studenov, Fiz. Tverd. Tela (St. Petersburg) **39**, 444 (1997) [Phys. Solid State **39**, 384 (1997)].
[11] Yu. M. Aliev, V. P. Silin, and S. A. Uryupin, Sverkhprovodimost: Fiz., Khim., Tekh. **5**, 228 (1992) [Supercond., Phys. Chem. Technol. **5**, 230 (1992)].
[12] G. L. Alfimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **106**, 671 (1994) [JETP **79**, 369 (1994)].
[13] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect*, Wiley, N.Y., 1982.
[14] A. Gurevich, Phys. Rev. B **46**, 3187 (1992).
[15] A. V. Ustinov, T. Doderer, R. P. Huebener, N. F. Pederson, B. Mayer, and V. A. Oboznov, Phys. Rev. Lett. **69**, 1815 (1992).
[16] Y. M. Zhang, D. Winkler, P.-A. Nilsson, and T. Claeson, Phys. Rev. B **51**, 8684 (1995).
[17] R. G. Mints and I. B. Snapiro, Phys. Rev. B **52**, 9691 (1995).

Translated by M. E. Alferieff
Edited by P. Shelnitz

Optical anisotropy as a technique for studying ultrafast phase transformations at solid surfaces

M. B. Agranat,*) S. I. Anisimov, S. I. Ashitkov, A. V. Kirillin, P. S. Kondratenko, A. V. Kostanovskiĭ, and V. E. Fortov

Research Center for the Thermophysics of Pulsed Processes, Russian Academy of Sciences, 127412 Moscow, Russia

(Submitted 23 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 2162–2173 (June 1998)

A new technique for testing long-range order in high-absorption anisotropic crystals has been developed using conversion of an incident p-(s-)wave to an s-(p-)wave due to optical anisotropy. The technique yields time-resolved measurements of parameters related to phase transformations in thin (10^{-6}–10^{-5} cm) layers with a high resolution (10^{-12} s). Using picosecond laser pulses and an ''Agat'' streak camera, the technique has been applied to an experimental investigation of melting and recrystallization kinetics at zinc and graphite surfaces . It was found that the process of melting takes less than 3 ps and the recrystallization time is about 100 ps. © *1998 American Institute of Physics.* [S1063-7761(98)01806-X]

1. INTRODUCTION

The progress in picosecond and femtosecond techniques has afforded extraordinary possibilities for studying dynamics of ultrafast phase transformations in condensed media.[1–5] The most important question in planning experiments in this field is selection of the technique for observing the dynamics of the processes of phase transformation. Various signals generated in materials and carrying information about phase transitions are used, such as optical reflection, luminescence, nonlinear optical effects, Raman spectra, etc. These techniques, however, yield mostly information on changes in the short-range order of crystal lattices, which do not change as a result of melting or the transition to an amorphous state.[6] Such techniques commonly used in studies of long-range order in crystals as electronography, x-ray diffraction, etc., cannot be operated with a high time resolution, although attempts have been made to perform time-resolved measurements of this kind. For example, ultrashort laser pulses were used to generate an electron beam for testing the dynamics of changes in the crystal structure of a metal, but the minimum time resolution was about 50 ps.[7]

In the present work, we have proposed and tested experimentally a new method of detecting ultrafast phase transformations at metal surfaces based on the effect of optical anisotropy. The technique can be applied to metals in which one of the phases is anisotropic, including, for example, zinc, cadmium, titanium, and also graphite. The essence of the technique is as follows.

When a light beam incident on a surface of an isotropic medium is polarized either in the incidence plane or perpendicular to this plane, the polarization of the reflected beam is the same. If the medium is an anisotropic crystal whose axis is neither in the incidence plane nor perpendicular to this plane, a component with polarization different from that of the incident light is detected in the reflected beam. This effect (rotation of polarization plane due to reflection from a surface), combined with pico- and femtosecond laser techniques, allows time-resolved measurements of the dynamics of ultrafast phase transformation when one of the phases is anisotropic, for example, when a crystal with a hexagonal close-packed (HCP) lattice is melted. Although the effect of optical anisotropy has been widely known, it has never been, to the best of our knowledge, applied to studies of phase transitions in solids.

The rest of this paper is organized as follows. Section 2 presents theoretical analysis of the characteristics of the polarization effect. Section 3 describes experimental techniques. Section 4 presents an experimental investigation of ultrafast melting of zinc and graphite exposed to picosecond laser pulses. The results are discussed in Sec. 5.

2. LIGHT REFLECTION FROM THE SURFACE OF AN ANISOTROPIC METAL: THEORETICAL ANALYSIS

Let us consider reflection of a monochromatic narrow light beam from a plane surface of an anisotropic metal. To this end, we use the impedance boundary condition[8]

$$E_\alpha = \xi_{\alpha\beta}(\mathbf{H}\times\mathbf{n})_\beta, \tag{1}$$

where **E** and **H** are the electric and magnetic fields, respectively, **n** is the unit vector of the normal to the surface, greek subscripts in Eq. (1) and the following equations take two values corresponding to the Cartesian coordinates in the surface plane, and $\xi_{\alpha\beta}$ is the two-dimensional surface impedance tensor, which is related in nonmagnetic metals to the three-dimensional permittivity tensor ε_{ik}:

$$(\xi^2)_{\alpha\beta} = (\varepsilon^{-1})_{\alpha\beta}. \tag{2}$$

In a uniaxial crystal, which is the only object of our analysis, the tensor ε_{ik} can be expressed as

$$\varepsilon_{ik} = \varepsilon(\delta_{ik} + s_i s_k \Delta), \tag{3}$$

where δ_{ik} is the three-dimensional unit tensor, **s** is the unit vector along the anisotropy axis, ε is the isotropic compo-

nent of the permittivity tensor, and Δ is the optical anisotropy constant. In what follows, we will assume that the following conditions are satisfied:

$$\varepsilon \gg 1, \quad \Delta \ll 1. \tag{4}$$

Substitution of Eq. (3) in Eq. (2) with account of Eq. (4) yields the following expression for the surface impedance tensor of a uniaxial metallic crystal:

$$\xi_{\alpha\beta} = \xi\left(\delta_{\alpha\beta} - \frac{1}{2}\nu_\alpha \nu_\beta \Delta \sin^2\psi \right), \tag{5}$$

where

$$\xi = 1/\sqrt{\varepsilon} \tag{6}$$

is the isotropic component of the surface impedance, ψ is the angle between the anisotropy axis and normal to the surface, and $\boldsymbol{\nu}$ is the two-dimensional unit vector along the projection of the anisotropy axis on the metal surface. In a single-crystal sample the angle ψ and vector $\boldsymbol{\nu}$ are constant, whereas in a polycrystalline sample they are functions of coordinates. In our calculation of reflection characteristics we begin with the case of a single crystal.

We denote by E_p and E_s the electric field amplitudes in the incident beam corresponding to polarizations in the incidence plane (p-wave) and perpendicular to this plane (s-wave). The respective amplitudes in the specular reflected beam will be denoted by E_p' and E_s'. When the incident beam is p-polarized ($E_s = 0$), the tangential components of the electric and magnetic fields on the metal surface are expressed as

$$\mathbf{E}_t = (E_p - E_p')\mathbf{l}\cos\theta + E_s'\mathbf{n}\times\mathbf{l}, \tag{7}$$

$$\mathbf{H}_t = (E_p + E_p')\mathbf{n}\times\mathbf{l} + E_s'\mathbf{l}\cos\theta, \tag{8}$$

where $\mathbf{l} = \mathbf{k}_t/|\mathbf{k}_t|$ is the unit vector directed along the crossing line between the light incidence plane and metal surface plane, $\mathbf{k}_t = \mathbf{k} - \mathbf{n}(\mathbf{n}\cdot\mathbf{k})$ is the tangential component of the wave vector $\mathbf{k}$, $k = 2\pi/\lambda$, and λ is the light wavelength.

By substituting Eqs. (7) and (8) in the boundary condition (1) and taking into account Eq. (5), we obtain a system of equations for the amplitudes E_p' and E_s'. By virtue of Eq. (4), the solution for amplitude E_p' is almost identical, to within some corrections, to the corresponding expression in the Fresnel formulas.[8] The amplitude E_s' describes the effect under consideration, namely the change in the light polarization due to the metal anisotropy. The quantitative measure of the effect is the ratio between the s-component intensity in the reflected beam and the intensity of the incident p-polarized light, $R_{p\to s}$. The calculation of this parameter yields

$$R_{p\to s} = \frac{1}{4}\frac{|\xi|^2|\Delta|^2\cos^2\theta}{|\cos\theta + \xi|^2}\sin^4\psi\sin^2 2\phi. \tag{9}$$

Here ϕ is the angle between vectors $\mathbf{l}$ and $\boldsymbol{\nu}$, in other words, the angle between the two planes perpendicular to the metal surface, one of which contains the incident light wave vector plane of (plane of incidence), the other the metal anisotropy axis (plane of anisotropy).

A similar calculation for the ratio between the intensities of the reflected p-polarized component and incident s-polarized component ($R_{s\to p}$) in the approximation determined by Eq. (4) yields

$$R_{s\to p} = R_{p\to s}. \tag{10}$$

Note two important implications of Eqs. (9) and (10). The first is that the intensity of the reflected component with the alternative polarization as a function of the angle ϕ has four maxima and four minima as the crystal is rotated around the surface normal through an angle of 2π, and the minima correspond to the absence of the effect. The maxima take place when the angle ϕ between the plane of incidence and the plane of anisotropy is an odd multiple of $\pi/4$, and no effect is detected when the two planes either coincide or are perpendicular with respect to one another.

The second consequence of Eqs. (9) and (10) is the strong dependence of the effect on the angle between the anisotropy axis and the reflecting surface normal described by the factor $\sin^4\psi$. When the anisotropy axis coincides with the normal, the effect vanishes.

Now let us proceed to the analysis of reflection from a polycrystalline sample of an anisotropic metal with a plane surface. Since the surface impedance is not constant in the surface plane, the projection of the wave vector on the interface is not conserved when the anisotropic part of the reflection occurs, so the corresponding reflection component is not specular but diffuse. Let us assume that the crystal grain size a satisfies the condition $\lambda/2\pi \ll a \ll D$, where D is the laser spot size. In addition, we assume that the distribution of grain parameters is uniform and isotropic. Denote by $dR(p,\mathbf{k}_t \to s,\mathbf{k}_t')$ and $dR(s,\mathbf{k}_t \to p,\mathbf{k}_t')$ the ratios of the intensities of the alternatively polarized component reflected into the solid angle element $d\Omega$ in the direction defined by the tangential component $\mathbf{k}_t'$ of the reflected light wave vector and the incident light intensity.

The calculation performed in this section for single crystals can be easily generalized to the case of a polycrystal with a plane surface if the second terms on the right-hand side of Eqs. (7) and (8) are replaced by the corresponding Fourier integrals with respect to the tangential component of the reflected light wave vector, $\mathbf{k}_t$. The solution of the equation system resulting from the boundary condition (1) yields the following expression for the relative intensities of the alternatively polarized reflected components:

$$dR(p,\mathbf{k}_t \to s,\mathbf{k}_t') = dR(s,\mathbf{k}_t \to p,\mathbf{k}_t') = \frac{1}{4}\frac{|\xi|^2|\Delta|^2\cos^2\theta}{|\cos\theta + \xi|^2}K(\mathbf{k}_t' - \mathbf{k}_t)d\Omega. \tag{11}$$

Here

$$K(\mathbf{k}_t' - \mathbf{k}_t) = \frac{k^2\cos\theta}{4\pi^2}\int d\boldsymbol{\eta}\exp\{-i(\mathbf{k}_t' - \mathbf{k}_t)\cdot\boldsymbol{\eta}\} \times \langle \sin^2\psi(\mathbf{r}+\boldsymbol{\eta})\sin 2\phi(\mathbf{r}+\boldsymbol{\eta}) \times \sin^2\psi(\mathbf{r})\sin 2\phi(\mathbf{r})\rangle_{\mathbf{r}}.$$

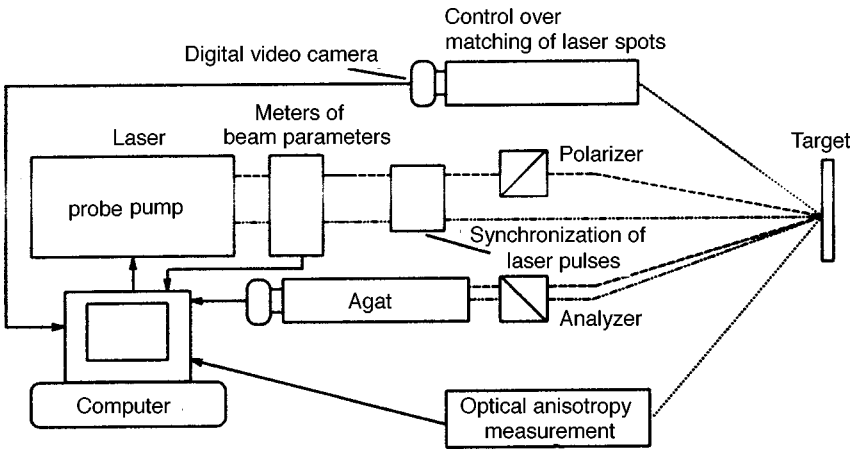


FIG. 1. Optical diagram of the experimental facility.

The angular brackets $\langle \ldots \rangle_{\mathbf{r}}$ denote averaging over coordinate $\mathbf{r}$ in the laser spot plane. It follows from Eq. (11) that the alternatively polarized reflected beam has an additional contribution to its divergence $\delta\theta$ relative to the incident beam:

$$\delta\theta \sim \lambda/2\pi a. \tag{12}$$

The expression for the relative intensity of alternatively polarized reflected light component integrated over the solid angle, which can be derived from Eq. (11), differs from Eq. (9) by the replacement

$$\sin^4\psi \sin^2 2\phi \rightarrow 1/2\langle \sin^4\psi \rangle. \tag{13}$$

Here we omit the complictaed expression for a small-grain polycrystal satisfying the condition

$$a \ll \lambda/2\pi.$$

Note only that in this case the diffusely reflected light due to the optical anisotropy occupies the entire half-space corresponding to the total solid angle 2π. In addition to the alternative polarization, it also contains a component with the polarization of the incident light. The total intensity of the alternatively polarized component, in comparison with that for a large-grain crystal, has an additional small factor $\sim(2\pi a/\lambda)^2$.

3. EXPERIMENTAL TECHNIQUES

An optical diagram of the experimental facility is given in Fig. 1.

The experiments were performed using a LIT-5 laser system. Its operation is based on the multistep (SBS and SRS) time compression of initial nanosecond pulses from a Nd:YAG laser, and it generates four time-synchronized optical pulses with parameters $\tau_1 \sim 200$ fs, $\lambda_1 = 1000$ nm; $\tau_2 \sim 1$ ps, $\lambda_2 = 780$ nm; $\tau_3 \sim 20$ ps, $\lambda_3 = 630$ nm; $\tau_4 \sim 500$ ps, $\lambda_4 = 530$ nm. The repetition rate is up to 25 Hz and the pulse energy (for τ_1 and τ_2) up to 1 mJ. The pulse parameters are monitored using the measuring devices of the facility, including optical pulse energy detectors , an ''Agat'' streak camera with a time resolution of about 3 ps, an autocorrelator with a resolution of about 50 fs, and a device measuring pulse spectra. All measurements, including those of monitoring devices, were fed to a multichannel digital device collecting and processing input data, which includes several highly sensitive CCD video-cameras (built around CCD detector arrays), several analogue inputs, a synchronization unit to control data inputs, a laser controller, and a computer loaded with instructions to make the experimental facility follow a predetermined program, to read and process simultaneously data fed from the electronic streak camera, oscilloscope, autocorrelator, pulse energy meters, light detectors, and spectral devices.

The dynamics of the reflected light was measured by the ''Agat'' electronic streak camera operated at time resolutions of 3 and 10 ps. The sample was pumped by a laser pulse with parameters $\tau_2 \sim 1$ ps and $\lambda_2 = 780$ nm. The probe pulse had parameters $\tau_4 \sim 500$ ps and $\lambda_4 = 530$ nm. The pump pulse was focused into a spot with a diameter of 100–300 μm, the probe pulse into a spot of 50 μm. The minimal size of the probe laser spot was controlled by the sensitivity of the electronic streak camera and the total pulse energy, which should be lower than the surface damage threshold, and this condition was checked by exposing the sample to a series of pulses. The coincidence of the pump and probe laser spots was checked using the CCD camera and a microscope with a magnification of 70×. The probe pulse was polarized in the incidence plane (p-wave), and its angle of incidence was less then 15°. We fed to the input slit of the streak camera the reflected probe pulse passing through a crossed analyzer (s-polarization), the reflected probe pulse (p-polarization), a fraction of the incident pump pulse, and the reflected pump pulse, which served as a time reference. The camera output was an arbitrary combination of these pulses or their sum. The polarization contrast of the optical package with the crossed polarizer and analyzer was about 10^5, which is much higher than the value required for accuracy compatible with the other measurements in our experiments. All experiments were performed in air.

The optical anisotropy of the tested area on the crystal surface was tested without pump pulses using radiation generated by a helium–neon laser. Photomultiplier tubes were used in measurements of intensities of s- and p-polarized components of reflected light as functions of the angle through which the target was rotated around the normal to its surface. Thus, the factor $R_{p\rightarrow s}$ of energy transfer from the p-

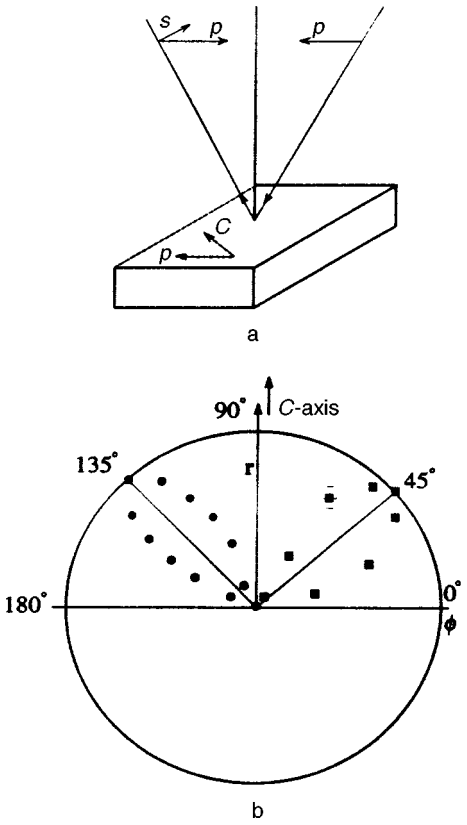


FIG. 2. Measurement of optical anisotropy: (a) beam reflection from surface of anisotropic metal; (b) $R_{p\to s}(\phi)$: squares are experimental data (Zn); circles represent the calculations ($\sin^2 2\phi$).

to s-wave as a function of angle ϕ between the incidence and anisotropy planes was measured.

4. EXPERIMENTAL RESULTS

4.1. Zinc

Experiments were performed with a zinc crystal whose surface was polished, then etched in a solution of nitric acid. The probe and pump laser pulses were focused on the surface whose plane contained the C-axis.

In order to identify the polarization effect, we first measured changes in probe light reflected from a cold zinc sample (not exposed to a pump laser pulse). Figure 2 shows the transformation coefficient $R_{p\to s}$ normalized to unity at the maximum as a function of the angle ϕ in polar coordinates. This graph also shows the theoretical dependence $R_{p\to s}(\phi)$, which is given by the function $\sin^2 2\phi$. Both these functions are plotted in one quadrant.

These measurements demonstrated the presence of the polarization effect in light reflection from the surface of an HCP crystal of Zn. It was found that, in accordance with Eq. (9), the normal to the sample surface is a four-fold symmetry axis for the transformation factor $R_{p\to s}$, i.e., $R_{p\to s}(\phi+\pi/4)=R_{p\to s}(\phi)$.

One can see in Fig. 2 that the experimental and theoretical data are virtually identical. The coefficient $R_{p\to s}$ of the p-wave to s-wave intensity transformation is about 10^{-2} at $\phi=\pi/4$. The minimum $R_{p\to s}=10^{-3}$ at $\phi=0$, $\pi/2$ is due to the surface inhomogeneities or error in measurements of ϕ, since the contrast of the optical system allows more accurate measurements.

Time-resolved measurements of light reflected from the zinc surface were performed at $\phi=\pi/4$ (at a peak of $R_{p\to s}$). The damage threshold fluence F_0 of the pump laser pulse was defined as the fluence producing a damaged spot with a diameter of about 60 μm on the surface. The distribution of fluence over the focal spot was approximately Gaussian, $F=F_m\exp(-d^2/d_0^2)$, where F_m is the maximum fluence at the spot center. When the fluence at the spot center exceeded F_0 by a factor of two to three, the damaged spot diameter was up to $2d_0\approx 200$ μm and approximately constant at higher pulse energies. The absolute value of the damage threshold fluence $F_0\approx 0.05$ J/cm^2. Note that precise measurements of the threshold fluence were not intended in our experiments. Relative changes in the damage threshold were measured to within 10%.

Time-resolved measurements of s- and p-polarized components of reflected probe beam were performed at the pump pulse fluence in the range $F_0<F<10F_0$, and in most cases a site on the surface was exposed to a pump pulse only once. Statistically processed measurements (about five hundred streak photos), which demonstrated reasonable reproducibility, have led us to the following conclusions.

1. At pumping laser fluences in the range $F_0<F<5F_0$, the intensity of the reflected s-polarized component drops to the noise level in less than 3 ps after the pump pulse. After a time ~100–300 ps, its intensity recovers completely or partially, depending on the fluence of the pump pulse. The pulse shape of the p-polarized reflected component is not affected by the pump pulse. Typical streak photos are given in Fig. 3a. Figure 4 shows the intensity of the s-polarized component versus time obtained by statistically processing our measurements with due account of the pump pulse shape.

2. When the pump pulse fluence is three to five times F_0, a drop in the p-polarized component with a duration of ~10–50 ps, decreasing with F, is observed, along with the changes in the s-polarized component. The intensities of the s- and p-waves do not recover to their initial values. Typical streak photos are shown in Fig. 3b.

3. The optical anisotropy parameter ($R_{p\to s}(\phi)$) in the region exposed to a single pump pulse with a fluence close to the damage threshold value recovers to its initial value. After multiple exposure $R_{p\to s}$ increases to $\sim 10^{-1}$ from the initial value $\sim 10^{-2}$, but this parameter no longer depends on angle ϕ, and the number N of pulses needed to attain such an effect decreases with the fluence of the pump pulse down to $N=1$ at $F>5F_0$. The surface structure after single and multiple exposures to pulses with $F\sim 3F_0$ has been studied using the electron microscope. After exposure to a single pulse, microscopic regions with dimensions of about 10 μm could be seen, and after exposure to repeated pulses a granular structure with a grain size of about 1 μm was formed. The intensity of the reflected probe pulse (integrated over time and measured without a pump pulse) notably dropped as a result of this process, which is caused by the diffuse nature of reflection from the modified surface.

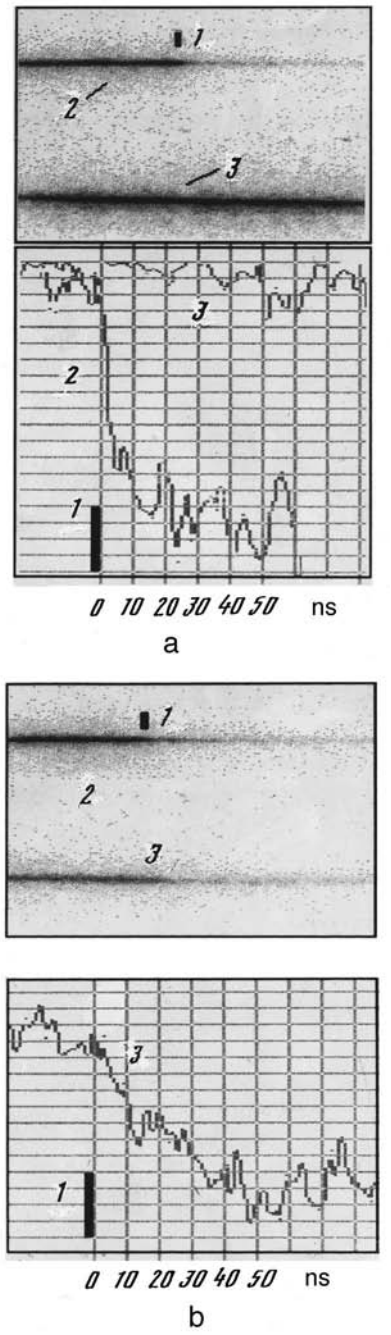


FIG. 3. Typical streak photos of (*1*) pump pulse, (*2*) *s*-, and (*3*) *p*-polarized components of the reflected probe beam at the pump pulse fluence (a) $F<5F_0$ and (b) $F>5F_0$ for a zinc single crystal.

4.2. Graphite

Experiments have been performed with a pyrolithically manufactured quasi-single crystal of graphite. The tested surface, whose plane contained the symmetry axis, underwent preliminary cleaning with laser radiation, namely, by exposure to multiple pump pulses with fluences higher than the damage threshold value.

Our measurements of optical anisotropy indicate that the maximum $R_{p\to s}$ is about 3×10^{-2} at $\phi=\pi/4$, i.e., it is higher than in the case of zinc, but the angular dependence is weaker: $R_{p\to s}(\phi=\pi/4)/R_{p\to s}(\phi=0)=3-4$. The shape of the $R_{p\to s}(\phi)$ curve is not affected by multiple exposure, unlike the case of zinc.

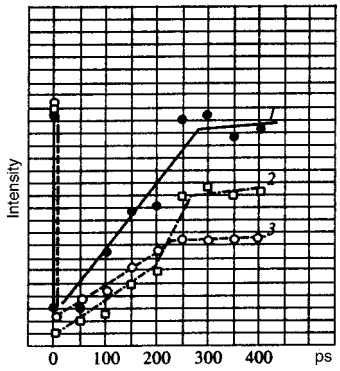


FIG. 4. Intensity of reflected *s*-wave versus the delay of the probe pulse with respect to the pump pulse at several fluence values of the pump pulse: (*1*) $F\sim2F_0$; (*2*) $F\sim4F_0$; (*3*) $F\sim5F_0$.

Time-resolved measurements of the intensities of reflected *s*- and *p*-polarized components after a pump laser pulse were performed similarly, but on the same surface area at different fluences of pump pulses. Our measurements have led us to the following conclusions.

1. The measurements are less reproducible than in the case of zinc, probably because of inhomogeneity of the structure and the various physical crystal parameters in the plane containing the symmetry axis.

2. The damage threshold of the pump pulse fluence is $F_0\approx0.1$ J/cm^2, which is close to the value given by Reitze *et al.*[3]

3. In the narrow range of fluences of the pump pulse, $F_0<F<2F_0$, the intensity of the *s*-polarized reflected component drops rapidly in about 3 ps after the pump pulse and then recovers almost to its initial value in a time ~100–150 ps. The pulse shape of the reflected *p*-polarized component is not affected. Typical streak photos are shown in Fig. 5.

4. At the pump pulse fluence $F>2F_0$, the intensity of the reflected *p*-polarized component also drops rapidly together with the drop in the probe beam *s*-polarized component. The typical decay time of the *s*-polarized component is again less than 3 ps, whereas the *p*-polarized component decays more slowly, as in Refs. 3 and 4.

5. DISCUSSION OF EXPERIMENTAL RESULTS

The theoretical analysis of experimental results described in the previous section is based on the model developed by Anisimov et al.[9,10] The incident pump laser pulse is absorbed by the degenerate electron gas, whose temperature during the pulse is much higher than the lattice temperature. To estimate the spatial scale and absolute electron temperature, we use Eq. (14) from Ref. 10:

$$T_e(0,t)=\left(\frac{3q_0^2T_0}{2\alpha\kappa_0}\right)^{1/3},\quad \delta(t)=\left(\frac{18q_a\kappa_0}{\alpha^2T_0}\right)^{1/3}, \tag{14}$$

where $q_a(t)$ is the absorbed laser intensity, α is the energy exchange rate between electrons and the lattice, κ_0 is the heat conductivity of a metal in equilibrium, and T_0 is the

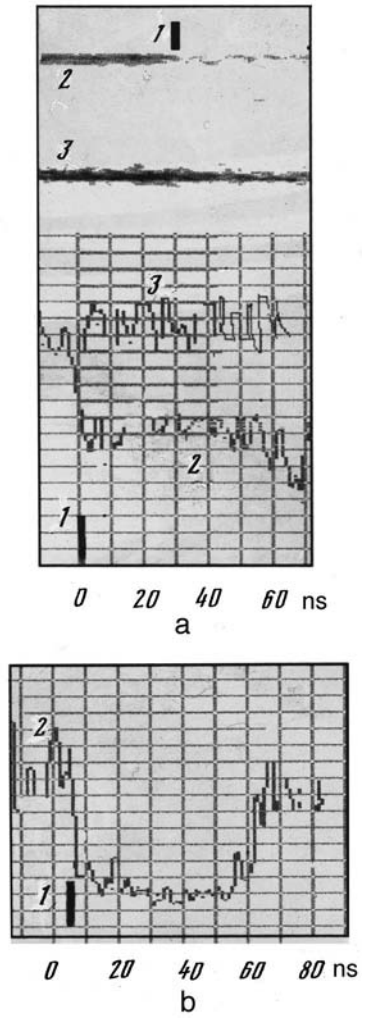


FIG. 5. Typical streak photos of (*2*) *s*- and (*3*) *p*-polarized components of the reflected probe beam at the (*1*) pump pulse fluence $F_0<F<2F_0$: (a) drop in the intensity of the *s*-polarized component in the process of graphite melting; (b) recovery of the *s*-polarized component during recrystallization (the measurements were processed with due account of the actual probe pulse shape). The time resolution is about 3 ps.

initial (room) temperature. Equation (14) applies to the case when the thickness of the heated layer is larger than the skin depth. Assuming that $q_a \approx 10^{11}$ W/cm^2, $\kappa_0 \approx 1$ W/cm K, $\alpha \approx 10^{12}$ W/cm^{3}K, and $T_0 \approx 300$ K, we obtain $T_e \approx 1.6 \times 10^4$ K and $\delta \approx 1.8 \times 10^{-5}$ cm. The lattice of the materials under consideration is heated in this layer in several picoseconds. The maximum lattice temperature at the sample surface is estimated using Eq. (25) from Ref. 10:

$$T_{im} \approx T_0 + \frac{F_a}{c_i} \frac{\beta\mu}{\beta+\mu}, \quad \beta = \sqrt{\frac{\alpha}{\kappa_0}}, \tag{15}$$

where F_a is the laser fluence absorbed, μ is the absorption coefficient, and c_i is the lattice specific heat. Assuming $\mu \approx 3 \times 10^5$ cm^{-1}, we obtain $T_{im} = 2500$ K for zinc and $T_{im} = 6100$ K for carbon at the damage threshold (we assume that the reflectivities of Zn and C averaged over time are 0.5 and 0.3, respectively). Both these temperatures are notably higher than the melting temperatures of the corresponding materials.

Since the target is heated in a very short time, the material density is constant during a time interval of about δ/c_s, during which a rarefaction wave passes through the heated surface layer. Here c_s is the speed of sound in the heated material. During this time (several tens of picoseconds in this specific case) the heated layer is kept under a pressure that can be estimated by the formula $P \approx \Gamma(\rho_0)E_T$, where $\Gamma(\rho_0)$ is the Grüneisen parameter at the initial density ρ_0 and E_T is the laser pulse energy absorbed in unit volume. Our estimates yield the pressure up to 30 kbar for zinc and 150 kbar for carbon near the damage threshold. The latter value is in agreement with the estimate quoted in Ref. 11. This pressure is higher than at the graphite–diamond–liquid triple point on the carbon phase diagram. It is noteworthy that electron–lattice relaxation has little effect on the pressure, since the partial pressures due to electrons and lattice are, by definition, proportional to the corresponding energy densities, whereas the total energy is constant.

In order to produce changes in the reflectivity due to melting, one should destroy the crystal lattice in the skin layer. When the overheating is small, $T_i - T_m \ll T_m$, the melting front propagates from the surface into the crystal bulk. The velocity of this front is lower than the speed of sound,[12] so the time a melting wave needs to traverse the skin layer is at least several tens of picoseconds. Our experiments indicate, however, that the reflected *s*-polarized component drops in a time shorter than 3 ps after the pump pulse. This means that the melting takes place in the bulk of material as a result of homogeneous nucleation. This process was studied[13] taking into account the activation energy caused by lattice deformation around a growing nucleus. Under certain conditions one can expect formation of nonspherical nuclei. This analysis, however, cannot be directly applied to carbon since the liquid phase of this material has a higher density than the solid phase in a pressure range of 70–80 kbar.[11]

The cooling of a metal heated by ultrashort laser pulses is largely controlled by thermal conductivity.[10,14] The contribution of evaporation to the cooling rate is negligible. Hence the cooling time can be estimated by the formula $\tau_c \sim \delta^2/4\chi$, where χ is the heat diffusivity. This estimate yields $\tau_c \approx 300$ ps, which is in satisfactory agreement with our measurements.

6. CONCLUSIONS

The principal results of the reported work are as follows.

We have developed an optical technique for studying the kinetics of phase transformation at surfaces of anisotropic, highly absorbing solids (including polycrystals). The technique allows one to test the presence or absence of long-range order in microscopic regions on crystal surfaces and on bottoms of microcraters, where measurement using small-angle x-ray radiation or electron beams have low efficiency.

Time-resolved measurements of phase transformations on zinc and graphite surfaces due to picosecond laser pulses based on the suggested technique clearly indicate that melting processes in surface layers can be investigated with picosecond resolution.

The work was supported by the Russian Fund for Fun-

damental Research (Grants 96/52-18494, 96/52-18495, and 97-02-16044).

*)E-mail: agranat@tiv.phys.msu.su

[1] J. M. Liu, R. Yen, H. Kurz, and N. Bloembergen, Appl. Phys. Lett. **39**, 755 (1981).
[2] D. van der Linde, K. Sokolowski-Tunten, and J. Bialkowski, Appl. Surf. Sci. **109/110**, 1 (1997).
[3] D. H. Reitze, H. Ahn, and M. C. Downer, Phys. Rev. B **45**, 2677 (1992).
[4] A. M. Malvezzi, N. Bloembergen, and C. Y. Huang, Phys. Rev. Lett. **57**, 146 (1986).
[5] X. Y. Wang and M. C. Downer, Opt. Lett. **17**, 1450 (1992).
[6] T. S. Korsunskaya, Yu. V. Kudryavtsev, and I. V. Lezhnenko, Metallofizika **6**(5), 117 (1984).
[7] S. Williamson, G. Mourou, and J. Li, Phys. Rev. Lett. **52**, 2364 (1984).
[8] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).
[9] S. I. Anisimov, B. L. Kopeliovich, and T. L. Perel'man, Zh. Éksp. Teor. Fiz. **66**, 776 (1974) [Sov. Phys. JETP **39**, 375 (1974)].
[10] S. I. Anisimov and B. Rethfeld, Proc. SPIE **3093**, 192 (1997).
[11] F. P. Bundy, J. Chem. Phys. **38**, 618 (1963).
[12] V. I. Motorin and S. L. Musher, Zh. Tekh. Fiz. **52**, 1200 (1982) [Sov. Phys. Tech. Phys. **27**, 726 (1982)].
[13] V. I. Motorin and S. L. Musher, J. Chem. Phys. **81**, 465 (1984).
[14] B. Rethfeld and S. Anisimov, JETP Lett. **62**, 872 (1995).

Translation provided by the Russian Editorial office.

De Haas–van Alphen effect in unconventional superconductors

M. G. Vavilov and V. P. Mineev*)

L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 3 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2174–2192 (June 1998)

A theory of the de Haas–van Alphen effect in type-II p-wave and D-wave superconductors (the latter corresponds to the B_{1g} one-dimensional representation of group D_{4h}) has been developed. Solutions for the order parameter and density of quasiparticle states near the upper critical field have been calculated. If the curve enclosing the extremal cross section of the Fermi surface in the plane perpendicular to the external magnetic field coincides with the line of nodes of the superconducting order parameter, the effect of the transition to the superconducting state on the amplitude of magnetization oscillations is negligible. If the line of nodes is oriented differently with respect to the applied magnetic field, the de Haas–van Alphen oscillations are suppressed in a manner qualitatively similar to the case of conventional superconductors. © *1998 American Institute of Physics.* [S1063-7761(98)01906-4]

1. INTRODUCTION

The amount of experimental data concerning the de Haas–van Alphen effect in type-II superconductors that have been published recently is quite considerable (see the review by Corcoran *et al.*[1]). If the region of very high magnetic fields where the effect in the normal state takes place[1)] overlaps with the domain of a superconducting mixed state, magnetization oscillations persist in the region below the transition to the mixed state, $H<H_{c2}$, and the frequency of the oscillations as a function of magnetic field remains unchanged, while the amplitude drops with decreasing magnetic field faster than in a normal metal. Nonetheless, the effect is observable in fields down to $H\approx 0.5H_{c2}$. These observations indicate that the Landau quantization persists in the mixed state at magnetic fields considerably below than the upper critical field H_{c2}.

Attempts to interpret this phenomenon were made in a set of theoretical studies.[2–4] A self-consistent theory of the de Haas–van Alphen effect was developed in our previous work.[5] It was found that, at a finite impurity concentration, notwithstanding the high-purity condition $\pi\Gamma<\omega_c$, which should be satisfied in order to detect magnetization oscillations, there is a region of gapless superconductivity in the mixed state near the upper critical field H_{c2}, where the density of states at the Fermi level is finite:

$$N(E=0)\approx N_0\left(1-\frac{\sqrt{\pi^3 n_F}}{\ln n_F}\frac{H_{c2}-H}{H_{c2}}\right). \tag{1}$$

Here N_0 is the density of states in the normal metal, $n_F=\mu/\omega_c$, and μ is the chemical potential. The oscillating component $M^s_{\rm osc}$ of magnetization in the mixed state is suppressed in comparison with its amplitude $M^n_{\rm osc}$ in the normal state:

$$\frac{M^s_{\rm osc}}{M^n_{\rm osc}}\approx 1-\frac{\sqrt{\pi n_F}}{\ln n_F}\frac{H_{c2}-H}{H_{c2}}. \tag{2}$$

Equations (1) and (2) were derived[5] to first order in the order parameter $\Delta^2\sim(H_{c2}-H)/H_{c2}$ under the assumption that $T<\Gamma\ll\omega_c$.

In high-temperature superconductors, which were discovered in recent years, and in heavy-fermion superconductors, superconducting states with anisotropic pairing were hypothesized. In this connection, the need in a theory of the de Haas–van Alphen effect in superconductors with symmetry nodes of the order parameter became obvious. Maki[6] extended his theory[2] to superconducting states with order parameter nodes on the equator and poles of the Fermi sphere. He used an expression for the excitation spectrum derived[7] neglecting the Landau quantization and an order parameter obtained in the momentum representation.

We have developed a quantum theory of the de Haas–van Alphen effect for p-wave states, whose order parameter has been calculated exactly for fields $H\approx H_{c2}$ and arbitrary temperatures.[8] A phase whose symmetry corresponds to the one-dimensional representation B_{1g} of group D_{4h} has also been investigated. For brevity, it will be denoted as the D-phase. The theory of magnetization oscillations in the D-phase is of special interest because recent experimental data indicate that the B_{1g} state can occur in the high-T_c superconductor $YBa_2Cu_3O_{7-x}$.[9] Our results demonstrate that, if the curve enclosing the extremal Fermi surface cross section in the plane perpendicular to the applied magnetic field does not coincide with the line of nodes of the superconductor order parameter, the suppression of the oscillation amplitude is qualitatively similar to the case of conventional superconductivity. On the contrary, if the order parameter nodes are on the curve enclosing the cross section with an extreme area, the transition to the superconducting state has little effect on the oscillation amplitude. Thus, observation of the de Haas–van Alphen effect in the mixed state can be used in identifying unconventional superconducting states.

The paper is organized as follows. The next section presents equations for the Green's function of superconductors

with p and D-wave pairing. The order parameter near the upper critical field for different superconducting states will be derived in Sec. 3, then the corresponding matrix elements of the order parameter will be calculated. Then the self-consistency equation for the order parameter amplitude will be solved. In Sec. 6 the density of states at the Fermi level and the amplitude of de Haas–van Alphen oscillations will be calculated.

2. ELECTRON GREEN'S FUNCTION IN DOPED SUPERCONDUCTORS

The Gor'kov equation for anisotropic superconductors has the form

$$[i\omega-\hat{H}_0(\mathbf{R})-\hat{u}(\mathbf{R})]\hat{G}(\mathbf{R},\mathbf{R}',\omega) - \int d\mathbf{r}\hat{\Delta}(\mathbf{R},\mathbf{r})\hat{G}(\mathbf{R}-\mathbf{r},\mathbf{R}',\omega)=\delta(\mathbf{R}-\mathbf{R}'). \tag{3}$$

Here

$$\hat{H}_0(\mathbf{R})=\begin{pmatrix} H_0(\mathbf{R}) & 0 \\ 0 & -H_0^*(\mathbf{R}) \end{pmatrix},$$

$$\hat{u}(\mathbf{R})=\hat{\tau}_3 u(\mathbf{R})=\begin{pmatrix} u(\mathbf{R}) & 0 \\ 0 & -u(\mathbf{R}) \end{pmatrix},$$

$u(\mathbf{R})$ is the impurity scattering potential, and

$$H_0(\mathbf{R})=\frac{1}{2m}\left(-i\frac{\partial}{\partial\mathbf{r}}+\frac{e}{c}\mathbf{A}(\mathbf{r})\right)^2-\mu \tag{4}$$

is the one-particle Hamiltonian of electrons in the magnetic field. The magnetic field is assumed to be uniform and the same as the external magnetic field, which is justified at $H\sim H_{c2}$ in superconductors with a large Ginzburg–Landau parameter. Here we use eigenfunctions $\phi_l(\mathbf{R})$ of the operator $H_0(\mathbf{R})$, which form a representation of magnetic sublattices.[10] In the Landau gauge $\mathbf{A}(\mathbf{R})=(0, Hx, 0)$ the functions $\phi_l(\mathbf{r})$ have the form

$$\phi_l(\mathbf{r})=\sqrt{\frac{a}{\lambda}}\exp(ik_z z)\sum_m \exp(-iq_x am) \times\exp\left[i\left(q_y+\frac{\pi m}{a}\right)y\right]\varphi_n\left(\frac{x}{\lambda}+\left(q_y+\frac{\pi m}{a}\right)\lambda\right), \tag{5}$$

where

$$\varphi_n(s)=\frac{1}{\sqrt{2^n n!\sqrt{\pi}}}\exp\left(-\frac{s^2}{2}\right)H_n(s), \tag{6}$$

$$H_n(s)=(-1)^n\exp(s^2)\frac{d^n}{ds^n}\exp(-s^2) \tag{7}$$

are Hermite polynomials. The unit cell in the lattice of magnetic translations is a rectangle with sides $a_x=a$ and $a_y=2a$. The quantum number in this case is $l=\{n,k_z,\vec{q}\}$, where $\vec{q}$ is the two-dimensional vector in the first Brillouin zone $-\pi/a<q_x<\pi/a$, $-\pi/2a<q_y<\pi/2a$.

The matrix $\hat{G}(\mathbf{r},\mathbf{r}',\omega)$ contains the normal Green's function $G(\mathbf{r},\mathbf{r}',\omega)$ and anomalous Green's function $F(\mathbf{r},\mathbf{r}',\omega)$, which can be expressed in both the coordinate representation and the representation of $\phi_l(\mathbf{r})$ states:

$$\hat{G}(\mathbf{r},\mathbf{r}',\omega) =\begin{pmatrix} G(\mathbf{r},\mathbf{r}',\omega) & F(\mathbf{r},\mathbf{r}',\omega) \\ F^+(\mathbf{r},\mathbf{r}',\omega) & -G(\mathbf{r}',\mathbf{r},-\omega) \end{pmatrix} =\sum_{ll'}\begin{pmatrix} \phi_l(\mathbf{r})G_{ll'}(\omega)\phi_{l'}^*(\mathbf{r}') & \phi_l(\mathbf{r})F_{ll'}(\omega)\phi_{l'}(\mathbf{r}') \\ \phi_l^*(\mathbf{r})F_{ll'}^+(\omega)\phi_{l'}^*(\mathbf{r}') & -\phi_{l'}(\mathbf{r}')G_{l'l}(-\omega)\phi_l^*(\mathbf{r}) \end{pmatrix}. \tag{8}$$

The summation over the quantum numbers should be performed by the following rule:

$$\sum_l=\sum_{n=0}^{\infty}\int\frac{dk_z}{2\pi}\int_{-\pi/a}^{\pi/a}\frac{dq_x}{2\pi}\int_{-\pi/2a}^{\pi/2a}\frac{dq_y}{2\pi}. \tag{9}$$

The order parameter

$$\hat{\Delta}(\mathbf{R},\mathbf{r})=\begin{pmatrix} 0 & \Delta(\mathbf{R},\mathbf{r}) \\ \Delta^*(\mathbf{R},\mathbf{r}) & 0 \end{pmatrix} \tag{10}$$

is a function of two variables, namely the center-of-mass coordinate $\mathbf{R}$ of a Cooper pair and the relative coordinate $\mathbf{r}$ of the electrons in the pair, and it is determined by the following self-consistency equation:

$$\Delta^*(\mathbf{R},\mathbf{r})=V(\mathbf{r})T\sum_\omega F^+\left(\mathbf{R}-\frac{\mathbf{r}}{2},\mathbf{R}+\frac{\mathbf{r}}{2},\omega\right), \tag{11}$$

where $V(\mathbf{r})$ is the attractive potential between the electrons.

Using the same notation as in Eq. (8) in the Green's function averaged over the positions of the impurities, we obtain the following equation:[11]

$$\hat{G}(\mathbf{R},\mathbf{R}',\omega)=\hat{G}(\mathbf{R},\mathbf{R}',\omega) + \int d\mathbf{R}_1\hat{g}(\mathbf{R},\mathbf{R}_1,\omega)\hat{\Sigma}(\mathbf{R}_1,\omega)\hat{G}(\mathbf{R}_1,\mathbf{R}',\omega) + \int d\mathbf{R}_1\int d\mathbf{r}\,\hat{g}(\mathbf{R},\mathbf{R}_1)\hat{\Delta}(\mathbf{R}_1,\mathbf{r})\hat{G} \times(\mathbf{R}_1-\mathbf{r},\mathbf{R}',\omega). \tag{12}$$

Here $\hat{g}(\mathbf{r},\mathbf{r}',\omega)$ is the Green's function of the undoped normal metal in a magnetic field:

$$\hat{g}(\mathbf{r},\mathbf{r}',\omega)=\begin{pmatrix} g(\mathbf{r},\mathbf{r}',\omega) & 0 \\ 0 & -g(\mathbf{r}',\mathbf{r},-\omega) \end{pmatrix}. \tag{13}$$

The Green's function $g(\mathbf{r},\mathbf{r}',\omega)$ is expressed in terms of the eigenfunctions of the operator H_0 as follows:

$$g(\mathbf{r},\mathbf{r}',\omega)=\sum_l\phi_l(\mathbf{r})g_l(\omega)\phi_l^*(\mathbf{r}'), \tag{14}$$

where

$$g_l(\omega)=(i\omega-\xi_l)^{-1},$$

$$\xi_l=\xi_n(k_z)=\omega_c\left(n+\frac{1}{2}\right)+\frac{k_z^2}{2m^*}-\mu. \tag{15}$$

The expression for the impurity self-energy part in the case of anisotropic pairing has the form

$$\Sigma_{\rm imp}(\mathbf{R},\omega)=\begin{pmatrix}\overline{G}(\mathbf{R},\omega) & 0\\ 0 & -\overline{G}(\mathbf{R},-\omega)\end{pmatrix}, \tag{16}$$

where $\overline{G}(\mathbf{R},\omega)$ is determined by the equation

$$\overline{G}(\mathbf{R},\omega)=n_{\rm imp}u^2G(\mathbf{R},\mathbf{R},\omega), \tag{17}$$

$n_{\rm imp}$ is the impurity concentration, and u is the characteristic amplitude of the scattering potential due to impurity.

Unlike the case of conventional superconductivity, states with anisotropic pairing satisfy

$$\int d\Omega_{\hat r}F\left(\mathbf{R}+\frac{\mathbf{r}}{2},\mathbf{R}-\frac{\mathbf{r}}{2},\omega\right)=0,$$

where $d\Omega_{\hat r}$ denotes an element of solid angle in the space of unit vectors $\hat r=\mathbf{r}/|\mathbf{r}|$. Therefore the nondiagonal elements of the self-energy part vanish.

To third order in Δ, the normal and anomalous parts of the Green's function have the form

$$G_{ll'}(\omega)=G^{(0)}_{ll'}(\omega)+G^{(2)}_{ll'}(\omega), \tag{18}$$

$$F_{ll'}(\omega)=F^{(1)}_{ll'}(\omega)+F^{(3)}_{ll'}(\omega). \tag{19}$$

Equation (12) directly yields

$$G^{(0)}_{ll'}(\omega)=\delta_{ll'}G^{(0)}_l(\omega)=\frac{\delta_{ll'}}{g_l^{-1}(\omega)-\overline{G}^{(0)}_l}, \tag{20}$$

$$G^{(2)}_{ll'}(\omega)=G^{(0)}_l(\omega)\overline{G}^{(2)}_{l'l}(\omega)G^{(0)}_{l'}(\omega)+\sum_{l_1}G^{(0)}_l(\omega)\Delta_{ll_1}(\omega)F^{+(1)}_{l_1l'}(\omega) \tag{21}$$

and for the anomalous Green's function

$$F^{+(1)}_{ll'}(\omega)=-G^{(0)}_l(-\omega)\Delta^*_{ll'}(\omega)G^{(0)}_{l'}(\omega), \tag{22}$$

$$F^{+(3)}_{ll'}(\omega)=-\sum_{l_1}G^{(0)}_l(-\omega)\Delta^*_{ll_1}(\omega)G^{(2)}_{l_1l'}(\omega)+\sum_{l_1}G^{(0)}_l(-\omega)\overline{G}^{(2)}_{ll_1}(\omega)F^{+(1)}_{l_1l'}(\omega). \tag{23}$$

To zeroth order in Δ^2 the solution of the self-consistency equation for the self-energy part is diagonal in l and l', and has the form $\overline{G}^{(0)}_l(\omega)=-i\Gamma_{\rm imp}{\rm sign}\omega$, where $\Gamma_{\rm imp}=\pi n_{\rm imp}u^2N_0$. The correction to second order in Δ is derived from the equation

$$\overline{G}^{(2)}_{ll'}(\omega)=n_{\rm imp}u^2\sum_{pp'}\int d\mathbf{r}\phi_l(\mathbf{r})\phi_p(\mathbf{r})\times\phi^*_{p'}(\mathbf{r})\phi^*_{l'}(\mathbf{r})G^{(2)}_{pp'}(\omega). \tag{24}$$

The system of equations (20)–(24) is, in fact, identical to the system of equations that determines the Green's function in an isotropic superconductor,[5] the only difference being that the matrix element of the order parameter is expressed by

$$\Delta_{ll'}=\int d\mathbf{R}\int d\mathbf{r}\phi^*_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\phi^*_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\Delta(\mathbf{R},\mathbf{r}), \tag{25}$$

which should be derived explicitly for states with anisotropic pairing. In Eq. (25) the system has been shifted in the space of the coordinates $\mathbf{R}$ through $\mathbf{r}/2$.

3. STRUCTURE OF THE ORDER PARAMETER NEAR THE UPPER CRITICAL FIELD

In this section, we derive solutions of the linearized self-consistency equation (11) for phases with anisotropic pairing. Let us consider superconducting states whose order parameter in the mixed state $(\mathbf{R},\hat k)$ has the form

$$\Delta(\mathbf{R},\hat k)=\sum\psi_i(\hat k)\Delta_i(\mathbf{R}). \tag{26}$$

Here $\mathbf{R}$ is the pair center-of-mass coordinate and $\hat k\approx\mathbf{k}/k_F$ is the unit vector pointing in the direction of the relative electron momentum in a pair. Equation (26) yields the Fourier transform (with respect to the relative coordinate $\mathbf{r}$) of the order parameter given by Eq. (11); $\psi_i(\hat k)$ are the basis functions of the irreducible representation of the crystal symmetry group, in which the pairing potential is expanded, and the pair potential is nonvanishing only in a layer of width ϵ_0 near the Fermi surface:

$$V(\hat k,\hat k')=-|g|\sum_i\psi_i(\hat k)\psi^*_i(\hat k'). \tag{27}$$

We consider p-wave states with order parameter $\psi_i(\hat k)=\sqrt{3}\hat k_i$, in which a Cooper pair is in states with spin projections $S_z=\pm1$ with equal probabilities and in the $S_z=0$ state with zero probability, and also a state corresponding to the one-dimensional representation B_{1g} in a crystal with tetragonal symmetry D_{4h} with the function

$$\psi(\hat k)=\sqrt{\frac{15}{4}}\,(\hat k_x^2-\hat k_y^2).$$

In a uniform magnetic field, the linearized equation for the order parameter is written in the form[8]

$$\Delta_i(\mathbf{R})=gT\sum_\omega\sum_j\int d\mathbf{r}\psi^*_i(\hat r)\psi_j(\hat r)\times\tilde g(\mathbf{r},-\omega)\tilde g(\mathbf{r},\omega)\exp[i\mathbf{r}\cdot\mathbf{D}(\mathbf{R})]\Delta_j(\mathbf{R}). \tag{28}$$

Here

$$\mathbf{D}(\mathbf{R})=-i\frac{\partial}{\partial\mathbf{R}}+\frac{2e}{c}\mathbf{A}(\mathbf{R}),$$

e is the absolute value of the electron charge, $\hat r$ is the unit vector aligned with $\mathbf{r}$, and $\tilde g(\mathbf{r},\omega)$ is the electron Green's function in the normal state defined so that it depends only on the difference between electron coordinates:

$$\tilde{g}(\mathbf{r}-\mathbf{r}',\omega)=\exp\left(i\frac{e}{c}\int_{\mathbf{r}'}^{\mathbf{r}}\mathbf{A}(\mathbf{s})d\mathbf{s}\right)g(\mathbf{r},\mathbf{r}',\omega). \tag{29}$$

For simplicity we do not consider the action of the magnetic field on electron spins.

The solution of Eq. (28) has the form of a finite or infinite linear combination of functions $f_N(\mathbf{r})$:

$$\Delta_i(\mathbf{R})=\sum_{N=0} A_N^i f_N(\mathbf{R}). \tag{30}$$

In a square lattice of Abrikosov flux lines (for simplicity we analyze only this case), the functions $f_N(\mathbf{R})$ have the form

$$f_N(\mathbf{R})=\sqrt[4]{2\pi}\sum_{\nu}\exp\left(\frac{2\pi i\nu Y}{a}\right)\varphi_N\left(\sqrt{2}\left(\frac{X}{\lambda}+\frac{\pi\nu\lambda}{a}\right)\right), \tag{31}$$

where the $\varphi_n(S)$ are defined by Eq. (6).

There are three classes of solutions for p-wave phases,[8] which yield maximal values of H_{c2} in superconducting states in the form of the polar phase:

$$\Delta^{\rm pol}(\mathbf{R},\hat{k})=\sqrt{3}\Delta^{\rm pol}\hat{k}_z f_0(\mathbf{R}), \tag{32}$$

A-phase:

$$\Delta^{A}(\mathbf{R},\hat{k})=\sqrt{\frac{3}{2}}\Delta^{A}(\hat{k}_x-i\hat{k}_y)f_0(\mathbf{R}), \tag{33}$$

and the Scharnberg–Klemm phase (SK):

$$\Delta^{\rm SK}(\mathbf{R},\hat{k})=\sqrt{\frac{3}{2}}\Delta^{\rm SK}\left[(\hat{k}_x+i\hat{k}_y)f_0(\mathbf{R})+\frac{1-\beta_0}{\gamma_0}(\hat{k}_x-i\hat{k}_y)f_2(\mathbf{R})\right]. \tag{34}$$

The corresponding equations for H_{c2} are expressed as

$$\alpha_0(H,T)=1, \tag{35}$$

$$\beta_0(H,T)=1, \tag{36}$$

$$[1-\beta_0(H,T)][1-\beta_2(H,T)]=\gamma_0^2(H,T), \tag{37}$$

where

$$\alpha_0(H,T)=6\pi|g|T\sum_{\omega}\int_0^{\infty}dr\ r^2\int_0^{\pi}d\theta\ \sin\theta\cos^2\theta\times\exp\left(-\frac{r^2\sin^2\theta}{2\lambda^2}\right)\tilde{g}(r,-\omega)\tilde{g}(r,\omega),$$

$$\beta_N(H,T)=3\pi|g|T\sum_{\omega}\int_0^{\infty}dr\ r^2\int_0^{\pi}d\theta\ \sin^3\theta\times\exp\left(-\frac{r^2\sin^2\theta}{2\lambda^2}\right)\tilde{g}(r,-\omega)\times\tilde{g}(r,\omega)L_N\left(\frac{r^2\sin^2\theta}{\lambda^2}\right),$$

$$\gamma_0(H,T)=-3\pi|g|T\sum_{\omega}\int_0^{\infty}dr\ r^2\int_0^{\pi}d\theta\ \sin^3\theta\exp\left(-\frac{r^2\sin^2\theta}{2\lambda^2}\right)\tilde{g}(r,-\omega)\tilde{g}(r,\omega)\frac{r^2\sin^2\theta}{\sqrt{2}\lambda^2}.$$

Here the $L_N(s)$ are Laguerre polynomials. The structure of the solution is independent of whether we select the exact or quasiclassical expression,

$$\tilde{g}(\mathbf{r},\omega)=-\frac{m}{2\pi r}\exp\left(ip_F r\ \text{sign}\ \omega-\frac{|\omega|r}{v_F}\right), \tag{38}$$

for the Green's function $\tilde{g}(r,\omega)$. In subsequent calculations of the linear part of the self-consistency equation we will use the latter form.

Finally, the linear combination in Eq. (30) corresponding to the maximum H_{c2} in a tetragonal crystal in a magnetic field aligned with the four-fold axis contains an infinite number of terms. By retaining the first three (the contributions of the rest to H_{c2} and coefficients A_N with numbers $N=0,1,2$ are vanishing corrections), we obtain in the limit $T\to 0$

$$\Delta^D(\mathbf{R},\hat{k})=\sqrt{\frac{15}{4}}\Delta^D(\hat{k}_x^2-\hat{k}_y^2)[f_0(\mathbf{R})+0.15f_4(\mathbf{R})+0.013f_8(\mathbf{R})+\ldots]. \tag{39}$$

The field H_{c2} is determined by the equation

$$\begin{vmatrix}\beta_0-1 & \gamma_0 & 0\\ \gamma_0 & \beta_4-1 & \gamma_4\\ 0 & \gamma_4 & \beta_8-1\end{vmatrix}=0, \tag{40}$$

where

$$\beta_N(H,T)=\frac{15}{4}\pi g\sum_{\omega}\int_0^{\infty}dr\ r^2\int_0^{\pi}d\theta\ \sin^5\theta\times\exp\left(-\frac{r^2\sin^2\theta}{2\lambda^2}\right)\tilde{g}(r,-\omega)\times\tilde{g}(r,\omega)L_N\left(\frac{r^2\sin^2\theta}{\lambda^2}\right),$$

$$\gamma_N(H,T)=\frac{15}{8}\pi g\sum_{\omega}\int_0^{\infty}dr\ r^2\int_0^{\pi}d\theta\ \sin^5\theta\times\exp\left(-\frac{r^2\sin^2\theta}{2\lambda^2}\right)\tilde{g}(r,-\omega)\tilde{g}(r,\omega)\times\sum_{l=0}^{N}\left(-\frac{r^2\sin^2\theta}{\lambda^2}\right)^{l+1}\frac{\sqrt{N!(N+4)!}}{l!(l+4)!(N-l)!}.$$

The expressions for the order parameter in the anisotropic phases given in this section will be used in subsequent sections for calculating the matrix elements $\Delta_{ll'}$.

4. MATRIX ELEMENT OF THE ORDER PARAMETER

This section describes calculation of the matrix element (25) for various phases with anisotropic pairing.

The easiest case is calculation of $\Delta_{ll'}$ for the polar phase. In this phase the order parameter as a function of $\mathbf{r}$ is expressed as

$$\Delta^{\rm pol}(\mathbf{R},\mathbf{r})=\sqrt{3}\Delta^{\rm pol}\int\frac{d\mathbf{k}}{(2\pi)^3}f_0(\mathbf{R})\hat{k}_z e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{41}$$

Integration over $\mathbf{k}$ is performed in a layer of thickness $\delta k=\epsilon_c m/v_F$ near the Fermi surface, where the pairing potential is nonvanishing. Therefore we can set $\hat{k}_z\approx k_z/k_F$ and take $\Delta^{\rm pol}(\mathbf{R},\mathbf{r})$ to be a rapidly decaying function of $\mathbf{r}$ which is nonzero for $r\sim k_F^{-1}$. Rewriting $k_z\exp(i\mathbf{k}\cdot\mathbf{r})=-i\partial_z\exp(i\mathbf{k}\cdot\mathbf{r})$, where $\partial_z=\partial/\partial z$, we obtain (hereafter $l=\{n,k_z,q\}$ and $l'=\{n',k_z',q'\}$)

$$\begin{aligned}\Delta^{\rm pol}_{ll'}&=-i\sqrt{3}\,\frac{\Delta^{\rm pol}}{k_F}\int d\mathbf{R}f_0(\mathbf{R})\int d\mathbf{r}\int\frac{d\mathbf{k}}{(2\pi)^3}\\&\times\left[\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right]^*\partial_z e^{i\mathbf{k}\cdot\mathbf{r}}\\&=i\sqrt{3}\frac{\Delta^{\rm pol}}{k_F}\int d\mathbf{R}f_0(\mathbf{R})\int d\mathbf{r}\int\frac{d\mathbf{k}}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\\&\times\partial_z\left[\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right]^*\\&=\frac{\sqrt{3}}{2}\,\frac{k_z-k_z'}{k_F}\Delta^{\rm pol}\int d\mathbf{R}f_0(\mathbf{R})\phi_l^*(\mathbf{R})\phi_{l'}^*(\mathbf{R}).\end{aligned}$$

Here we have used the fact that integration with respect to the momentum, $\int d\mathbf{k}e^{i\mathbf{k}\cdot\mathbf{r}}$, yields a spherically symmetrical, rapidly decaying function of $\mathbf{r}$, which can be replaced by $\delta(\mathbf{r})$, and the integration over $\mathbf{r}$ becomes trivial.[11]

The remaining integral is identical to the matrix element of the order parameter in an isotropic superconductor. Its calculation is described in Appendix [see Eq. (86) for $N=0$]. Thus the only difference between the matrix element of the order parameter in the cases of a polar phase and an isotropic superconductor is the additional factor $i\sqrt{3}k_z/k_F$ in the former case:

$$\Delta^{\rm pol}_{ll'}=(2\pi)^3\delta(\vec{q}+\vec{q}\,')\,\delta(k_z+k_z')\Delta^{\rm pol}_{nn'}(\vec{q},k_z), \tag{42}$$

where

$$\begin{aligned}\Delta^{\rm pol}_{nn'}(\vec{q})&=i(-1)^{n'}\sqrt{3}\frac{k_z}{k_F}\Delta^{\rm pol}\sqrt{\sqrt{2\pi}\frac{(n+n')!}{2^{n+n'+1}n!n'!}}\\&\times\sum_\nu\exp(2i\nu q_x a)\varphi_{n+n'}\left(\sqrt{2}\left(q_y\lambda+\frac{\pi\lambda\nu}{a}\right)\right).\end{aligned} \tag{43}$$

Now let us analyze the case of the A-phase. The order parameter has the form

$$\Delta^A(\mathbf{R},\mathbf{r})=\sqrt{\frac{3}{2}}\Delta^A\int\frac{d\mathbf{k}}{(2\pi)^3}f_0(\mathbf{R})(\hat{k}_x-i\hat{k}_y)e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{44}$$

We write $(k_x-ik_y)\exp(i\mathbf{k}\cdot\mathbf{r})=-i(\partial_x-i\partial_y)\exp(i\mathbf{k}\cdot\mathbf{r})$ and perform calculations similar to the case of the polar phase:

$$\begin{aligned}\Delta^A_{ll'}&=-i\sqrt{\frac{3}{2}}\,\frac{\Delta^A}{k_F}\int d\mathbf{R}f_0(\mathbf{R})\int d\mathbf{r}\int\frac{d\mathbf{k}}{(2\pi)^3}\\&\times\left[\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right]^*(\partial_x-i\partial_y)e^{i\mathbf{k}\cdot\mathbf{r}}\\&=i\sqrt{\frac{3}{2}}\frac{\Delta^A}{k_F}\int d\mathbf{R}f_0(\mathbf{R})\int d\mathbf{r}\int\frac{d\mathbf{k}}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\\&\times\left[(\partial_x+i\partial_y)\left(\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right)\right]^*\\&=i\sqrt{\frac{3}{2}}\frac{\Delta^A}{2k_F}\int d\mathbf{R}f_0(\mathbf{R})\int d\mathbf{r}\frac{d\mathbf{k}}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\\&\times\left[\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)(\partial_X+i\partial_Y)\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)\right.\\&\left.-\phi_l\left(\mathbf{R}+\frac{\mathbf{r}}{2}\right)(\partial_X+i\partial_Y)\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right]^*\\&=-i\sqrt{\frac{3}{2}}\,\frac{\Delta^A}{2k_F}\int d\mathbf{R}f_0(\mathbf{R})\left[\phi_{l'}(\mathbf{R})\Pi_+\phi_l(\mathbf{R})\right.\\&\left.-\phi_l(\mathbf{R})\Pi_+\phi_{l'}\left(\mathbf{R}-\frac{\mathbf{r}}{2}\right)\right]^*.\end{aligned}$$

Here the derivatives with respect to the components of $\mathbf{r}$ are replaced by derivatives with respect to the $\mathbf{R}$ components, and then the integration over $\mathbf{k}$ and $\mathbf{r}$ is performed. The expressions in brackets are transformed so that the usual differentiation operators can be replaced by raising operators[2)] for transitions between states $\phi_{nk_z\vec{q}}$. We have

$$\begin{aligned}\Delta^A_{ll'}&=-i\sqrt{\frac{3}{2}}\frac{\Delta^A}{\sqrt{2}k_F\lambda}\left(\sqrt{n+1}\right.\\&\times\int d\mathbf{R}f_0(\mathbf{R})\phi^*_{n'k_z'q'}(\mathbf{R})\phi^*_{n+1,k_z,\vec{q}}(\mathbf{R})\\&\left.-\sqrt{n'+1}\int d\mathbf{R}f_0(\mathbf{R})\phi^*_{n'+1,k_z',q'}(\mathbf{R})\phi^*_{nk_zq}(\mathbf{R})\right).\end{aligned} \tag{45}$$

The integration over $\mathbf{R}$ is performed in the Appendix [$N=0$ in Eq. (86)]. Thus, we obtain

$$\begin{aligned}\Delta^A_{nn'}(\vec{q})&=i(-1)^{n'+1}\sqrt{\frac{3}{2}}\frac{\Delta^A}{k_F\lambda}\sqrt{\sqrt{2\pi}\frac{(n+n'+1)!}{2^{n+n'+1}n!n'!}}\\&\times\sum_\nu\exp(2i\nu q_x a)\varphi_{n+n'+1}\left(\sqrt{2}\left(q_y\lambda+\frac{\pi\lambda\nu}{a}\right)\right).\end{aligned} \tag{46}$$

The matrix element of the order parameter for the Scharnberg–Klemm phase is calculated similarly to that of the A-phase. In this case the order parameter is expressed as a linear combination of the functions $(\hat{k}_x+i\hat{k}_y)f_0(\mathbf{R})$ and $(\hat{k}_x-i\hat{k}_y)f_2(\mathbf{R})$:

$$\Delta^{\rm SK}(\mathbf{R},\hat{k})=\sqrt{\frac{3}{2}}\Delta^{\rm SK}[(\hat{k}_x+i\hat{k}_y)f_0(\mathbf{R})+A(\hat{k}_x-i\hat{k}_y)f_2(\mathbf{R})].$$

The parameters Δ and A will be derived from the self-consistency equations. In performing transformations similar to those for the A-phase, one should take into account that electron wave functions are acted on by the lowering operator Π_- in the part of the order parameter matrix element containing $f_0(\mathbf{R})$ (see Footnote 2), whereas in the part containing $f_2(\mathbf{R})$ it is the raising operator Π_+. Taking into account the properties of the raising and lowering operators, we obtain an expression similar to Eq. (45). Integration over $\mathbf{R}$ in the term with $f_2(\mathbf{R})$ is performed using Eq. (86) in the Appendix with $N=2$. Thus, the matrix element of the order parameter for the Scharnberg–Klemm phase is

$$\Delta^{\rm SK}_{nn'}(\vec{q})=i(-1)^{n'-1}\sqrt{\frac{3}{2}}\frac{\sqrt{n+n'}}{k_F\lambda}\Delta^{\rm SK}\left(1+\frac{A}{\sqrt{2}}\right)\times\sqrt{\sqrt{2\pi}\frac{(n+n')!}{2^{n+n'+1}n!n'!}}\times\sum_\nu\exp(2i\nu q_x a)\varphi_{n+n'-1}\left(\sqrt{2}\left(q_y\lambda+\frac{\pi\lambda\nu}{a}\right)\right). \tag{47}$$

Here we assume that $n'\approx n$.

In accordance with the results of the previous section, the order parameter in the D-phase can be expressed as

$$\begin{aligned}\Delta^D(\mathbf{R},\hat{k})&=\sqrt{\frac{15}{4}}\Delta^D(\hat{k}_x^2-\hat{k}_y^2)[f_0(\mathbf{R})+A_4f_4(\mathbf{R})\\&\quad+A_8f_8(\mathbf{R})+\dots]\\&=\frac{\sqrt{15}}{4}\Delta^D[(\hat{k}_x-i\hat{k}_y)^2f_0(\mathbf{R})+(\hat{k}_x+i\hat{k}_y)^2f_0(\mathbf{R})\\&\quad+(\hat{k}_x-i\hat{k}_y)^2A_4f_4(\mathbf{R})+(\hat{k}_x+i\hat{k}_y)^2A_4f_4(\mathbf{R})\\&\quad+(\hat{k}_x-i\hat{k}_y)^2A_8f_8(\mathbf{R})+\dots].\end{aligned} \tag{48}$$

In subsequent calculations we will take into account only the first two terms on the right of Eq. (48). The rest of the terms contribute only small corrections to the physical quantities calculated below.

Next is the expression for the matrix element of the order parameter in the D-wave phase at $n'\approx n$:

$$\Delta^D_{nn'}(\vec{q})=\sqrt{\frac{15}{4}}\Delta^D\frac{n+n'}{2k_F^2\lambda^2}(-1)^{n'}\sqrt{\sqrt{2\pi}\frac{(n+n')!}{2^{n+n'+1}n!n'!}}\times\sum_\nu e^{2i\nu q_x a}\left[\varphi_{n+n'+2}\left(\sqrt{2}\left(q_y\lambda+\frac{\pi\nu\lambda}{a}\right)\right)+\varphi_{n+n'-2}\left(\sqrt{2}\left(q_y\lambda+\frac{\pi\nu\lambda}{a}\right)\right)\right]. \tag{49}$$

It will be shown below that for all phases under consideration the density of electron states is expressed by the formula

TABLE I.

Semiconducting phase	$I_{nn'}(k_z)$	$I(\theta)$
s-wave	1	1
Polar	$3k_z^2/k_F^2$	$3\cos^2\theta$
A	$\frac{3}{2}\frac{n+n'+1}{k_F^2\lambda^2}$	$\frac{3}{2}\sin^2\theta$
Scharnberg–Klemm	$\frac{3}{2}\frac{n+n'}{k_F^2\lambda^2}\left(1+\frac{A}{\sqrt{2}}\right)^2$	$\frac{3}{2}\sin^2\theta\left(1+\frac{A}{\sqrt{2}}\right)^2$
D	$\frac{15}{8}\frac{(n+n')^2}{k_F^4\lambda^4}$	$\frac{15}{8}\sin^4\theta$

$$\int\frac{d\vec{q}}{(2\pi)^2}\Delta_{nn'}(\vec{q})\Delta^*_{nn'}(\vec{q})=\Delta^2I_{nn'}(k_z)\frac{(n+n')!}{2^{n+n'+1}n!n'!}. \tag{50}$$

The factors $I_{nn'}(k_z)$ for each phase are listed in Table I.

The interesting electron states are those near the Fermi surface, so we assume that n can be derived from the approximate equations $\omega_c(n+1/2)+k_z^2/2m^*\approx\mu$ and $n'\approx n$. Then $I_{nn'}(k_z)\approx I(\theta)$, where $\sin^2\theta=n/(k_F\lambda)^2$ and $\cos^2\theta=k_z^2/k_F^2$. Note that the quasiclassical expression for $I(\theta)$ is nothing but the square of the absolute value of the order parameter averaged over the azimuthal angle on the Fermi sphere. In particular, for the A-phase we have $|\hat{k}_x-i\hat{k}_y|^2=\sin^2\theta$. If the contribution due to the function $f_N(\mathbf{R})$ with $N\neq0$ is neglected in the expression for $I(\theta)$ in the D-wave phase, $I(\theta)$ has the form $I(\theta)\sin^4(\theta/2)$, which corresponds to the expression $(\hat{k}_x^2-\hat{k}_y^2)^2$ averaged over the azimuthal angle.

5. SOLUTION OF THE SELF-CONSISTENCY EQUATION FOR THE ORDER PARAMETER

Order parameters for superconducting phases in fields below the upper critical field should be derived from the self-consistency equation (11). In the A and polar phases, where the order parameter is proportional to $f_0(\mathbf{R})$ at $H=H_{c2}$, it is calculated similarly to the case of conventional superconductivity,[5] i.e., a solution proportional to that of linearized equation (11) is sought. Let us calculate the coefficient, i.e., the amplitude of the order parameter Δ for the polar and A phases. The equation for the order parameter amplitude has the form

$$\Delta^2=|g|T\sum_\omega\sum_{nn'}\int\frac{d\vec{q}}{(2\pi)^2}\int\frac{dk_z}{2\pi}\Delta_{nn'}(\vec{q})\times[F^{(1)+}_{nn'}(k_z,\vec{q},\omega)+F^{(3)+}_{nn'}(k_z,\vec{q},\omega)]. \tag{51}$$

The first term on the right of Eq. (51), containing $F^{(1)+}_{nn'}(k_z,\vec{q},\omega)$, is calculated in the quasiclassical approximation[5] for a magnetic field H below the upper critical field

H_{c2} and is identical to that given in Ref. 8. Thus, the self-consistency equation takes the form

$$N_0 \ln\sqrt{\frac{H_{c2}}{H}} = \Delta^2 T \sum_{\omega} \sum_{nn'mm'} \int \frac{dk_z}{2\pi} I_{nm}(k_z) I_{n'm'}(k_z) \times G_n^{(0)}(k_z,-\omega) G_m^{(0)}(k_z,\omega) G_{n'}^{(0)}(k_z,-\omega) \times G_{m'}^{(0)}(k_z,\omega) Y_{mm'}^{nn'}. \quad (52)$$

Here $I_{nm}(k_z)$ is a factor that determines the dependence of the order parameter on the quantum numbers n, n', and k_z (see Table I),

$$Y_{mm'}^{nn'} = 2\pi\lambda^2 \int D_{nm}(\vec{q}) D^*_{n'm}(\vec{q}) D^*_{n'm'}(\vec{q}) D_{nm'}(\vec{q}) \frac{d\vec{q}}{(2\pi)^2}, \quad (53)$$

where $D_{nm}(\vec{q})$ is given by

$$D_{nm}(\vec{q}) = \Delta_{nm}(\vec{q})/\Delta.$$

The upper critical field H_{c2} for p-wave states was given in Ref. 8.

When the conditions $T < \Gamma_{\rm imp} \ll \omega_c$ are satisfied, we can retain in Eq. (52) only the terms with $n = n' = m = m'$. Further calculations are, similar overall to those in the case of conventional superconductivity.[5] We assume that $Y_{nn}^{nn} \approx L$ and perform integration over k_z and the quantum number n, neglecting small oscillating components in the Poisson sum. This procedure reduces to integration over the polar angle θ and quasiparticle energies $\xi_n(k_z)$ measured with respect to the Fermi level.

In the A phase the integral over the polar angle is

$$\int \frac{I^2(\theta)}{\sin\theta} d\theta = \frac{2}{3}. \quad (54)$$

Hence the order parameter amplitude in the A phase is

$$[\Delta^A(H)]^2 \approx \frac{16\pi}{3L_A} n_F \Gamma_{\rm imp}^2 \frac{H_{c2}-H}{H_{c2}}. \quad (55)$$

The integral over θ is divergent at the poles in the case of the polar phase. We have already encountered such a divergence in the calculation of the order parameter amplitude as a function of magnetic field in s-wave states.[5] After introducing a cut-off at $\theta_c \approx 1/k_F\lambda$, we have

$$[\Delta^{\rm pol}(H)]^2 \approx \frac{16\pi}{9L_{\rm pol}} \frac{n_F \Gamma_{\rm imp}^2}{\ln n_F} \frac{H_{c2}-H}{H_{c2}}. \quad (56)$$

The order parameter in the Scharnberg–Klemm phase is a linear combination of the functions $(\hat{k}_x + i\hat{k}_y) f_0(\mathbf{R})$ and $(\hat{k}_x - i\hat{k}_y) f_2(\mathbf{R})$. Therefore the self-consistency equation for $T < \Gamma_{\rm imp} \ll \omega_c$ transforms to the system of equations

$$\Delta^2[1-\beta_0(H)-\gamma_0(H)A] = |g| T \sum_{\omega} \sum_{n} \int \frac{dk_z}{2\pi} \frac{1}{[(\omega+\Gamma_{\rm imp})^2 + \xi_n^2(k_z)]^2} \times \int \frac{d\vec{q}}{(2\pi)^2} \Delta^{(1)}_{nn}(\vec{q}) \Delta^*_{nn}(\vec{q}) \Delta_{nn}(\vec{q}) \Delta^*_{nn}(\vec{q}), \quad (57)$$

$$\Delta^2\{-\gamma_0(H) + [1-\beta_2(H)]A\} = |g| T \sum_{\omega} \sum_{n} \int \frac{dk_z}{2\pi} \frac{1}{[(\omega+\Gamma_{\rm imp})^2 + \xi_n^2(k_z)]^2} \times \int \frac{d\vec{q}}{(2\pi)^2} \Delta^{(2)}_{nn}(\vec{q}) \Delta^*_{nn}(\vec{q}) \Delta_{nn}(\vec{q}) \Delta^*_{nn}(\vec{q}). \quad (58)$$

Here the term linear in Δ^2 is expressed using the quasiclassical formulas of Ref. 8. We have

$$\Delta_{nn}(\vec{q}) \sim \Delta(1 + A/\sqrt{2})$$

[see Eq. (47)], $\Delta^{(1)}_{nn}(\vec{q})$ is the matrix element of $\Delta(\hat{k}_x + i\hat{k}_y) f_0(\mathbf{R})$, and $\Delta^{(2)}_{nn}(\vec{q})$ is the matrix element of $\Delta(\hat{k}_x - i\hat{k}_y) f_2(\mathbf{R})$.

In estimating the integral with respect to the components of the wave vector $\vec{q}$, we derive from Eqs. (57) and (58), as in the case of conventional superconductivity,[5] an algebraic system of equations for Δ and A:

$$1 - \beta_0(H) - \gamma_0(H) A = \frac{3L_{\rm SK}(1+A/\sqrt{2})^3}{32\pi} \frac{|g| N_0 \Delta^2}{n_F \Gamma_{\rm imp}^2}, \quad (59)$$

$$-\gamma_0(H) + (1-\beta_2(H)) A = \frac{3L_{\rm SK}(1+A/\sqrt{2})^3}{32\sqrt{2}\pi} \frac{|g| N_0 \Delta^2}{n_F \Gamma_{\rm imp}^2}. \quad (60)$$

From these two equations, we derive A:

$$A = \frac{\sqrt{2}\gamma_0 + 1 - \beta_0 + |g| N_0 \ln\sqrt{H_{c2}/H}}{\gamma_0 + \sqrt{2}(1 - \beta_2 + |g| N_0 \ln\sqrt{H_{c2}/H})}. \quad (61)$$

Here $\gamma_0 = \gamma_0(H_{c2}) = 1/\sqrt{2}$, $\beta_0 = \beta_0(H_{c2})$, $\beta_2 = \beta_2(H_{c2})$, $1 - \beta_0(H_{c2}) = (\sqrt{3}-1)/2$, and we have used the relationship

$$\beta_0(H) - \beta_0(H_{c2}) = \beta_2(H) - \beta_2(H_{c2}) = |g| N_0 \ln\sqrt{H/H_{c2}}.$$

By substituting the calculated A in Eq. (59) and retaining terms linear in $\ln(H_{c2}/H)$, we obtain

$$\alpha \ln\sqrt{\frac{H_{c2}}{H}} = \frac{3L_{\rm SK}(1+A/\sqrt{2})^3}{32\pi} \frac{\Delta^2}{n_F \Gamma_{\rm imp}^2}, \quad (62)$$

where $\alpha \approx 1.007$.

For the Scharnberg–Klemm phase, we finally obtain

$$[\Delta^{\rm SK}(H)]^2 \approx \frac{16\pi}{3L_{\rm SK}} \left(1 + \frac{A}{\sqrt{2}}\right)^{-3} n_F \Gamma_{\rm imp}^2 \frac{H_{c2}-H}{H_{c2}}. \quad (63)$$

In the remaining calculations, we will take $A = (1 - \beta_0)/\gamma_0$, since expressing A as a function of H would exceed the accuracy of the present calculation.

As in the case of the Scharnberg–Klemm phase, the amplitude of the order parameter and its form for D-wave su-

perconductors are derived from an algebraic system of equations. Suppose that the order parameter has the form

$$\Delta^D(\mathbf{R},\hat{k})=\sqrt{\frac{15}{4}}\Delta^D(\hat{k}_x^2-\hat{k}_y^2)[f_0(\mathbf{R})+A_4 f_4(\mathbf{R})+\dots], \tag{64}$$

where Δ^D is derived from the equation

$$1-\beta_0(H)=\left(\frac{15}{8}\right)^2(\Delta^D)^2|g|T\sum_\omega\sum_n\int\frac{dk_z}{2\pi}\frac{n^4}{k_F^4\lambda^4}\times\left[\frac{(2n)!}{2^{2n+1}n!^2}\right]^2\frac{1}{[(\omega+\Gamma_{\rm imp})^2+\xi_n^2(k_z)]^2}L_D(n) \tag{65}$$

resulting from Eq. (11) after multiplication by $\Delta^D(\hat{k}_x^2-\hat{k}_y^2)f_0(\mathbf{R})$ and integration over $\mathbf{R}$ and $\hat{k}$. The function $L_D(n)$ is determined by an expression similar to Eq. (53) and is a function of the integer parameter n. Hence in the remaining calculations we will neglect the dependence on n and consider L_D to be a numerical constant equal to this function averaged over n. By retaining terms up to the first order in $H_{c2}-H$ in Eq. (65), we obtain the order parameter amplitude for the D-wave phase:

$$(\Delta^D)^2=\frac{7\pi\cdot 2^9}{45L_D}n_F\Gamma_{\rm imp}^2\frac{H_{c2}-H}{H_{c2}}. \tag{66}$$

6. DENSITY OF STATES AND AMPLITUDE OF MAGNETIZATION OSCILLATIONS

Now we will find the density of states and amplitudes of magnetization oscillations in p-wave and D-wave superconductors. Note that the procedure for calculating these characteristics is quite similar to that in the case of s-wave superconductors. The only difference is in the integration over the polar angle θ.

The density of states $N(E)$ is expressed through the electron Green's function:

$$N(E)=-\frac{1}{\pi}\,\mathrm{Im}\sum_l G_{ll}(E),\qquad G_{ll'}(E)=G_l^{(0)}\delta_{ll'}+G_{ll'}^{(2)}. \tag{67}$$

The sum over the Landau level numbers is calculated using the Poisson formula:

$$\sum_n\dots=\sum_r\int dn e^{2\pi i r n}\dots \tag{68}$$

The nonoscillating component of the density of states in unconventional superconductors is determined by the term with $r=0$ and is similar to the expression (1) for the corresponding parameter for isotropic s-wave superconductors:

$$N^{\rm pol}(E=0)=N_0\left(1-\frac{3\sqrt{\pi}}{16}\frac{(\Delta^{\rm pol})^2}{\sqrt{n_F}\Gamma_{\rm imp}^2}\right) \tag{69}$$

in the polar phase,

$$N^A(E=0)=N_0\left(1-\frac{3\sqrt{\pi}}{32}\frac{(\Delta^A)^2}{\sqrt{n_F}\Gamma_{\rm imp}^2}\right) \tag{70}$$

in the A-phase,

$$N^{\rm SK}(E=0)=N_0\left(1-\frac{\sqrt{3\pi}}{32}\left(1+\frac{1-\beta_0}{\sqrt{2}\gamma_0}\right)^2\frac{(\Delta^{\rm SK})^2}{\sqrt{n_F}\Gamma_{\rm imp}^2}\right) \tag{71}$$

in the Scharnberg–Klemm phase, and

$$N^D(E=0)=N_0\left(1-\frac{45\sqrt{\pi}}{2^9}\frac{(\Delta^D)^2}{\sqrt{n_F}\Gamma_{\rm imp}^2}\right) \tag{72}$$

in the D-wave phase.

We will use Δ^2 for all these phases given by Eqs. (55), (56), (63), and (66), and determine the regions of applicability of expansions to first order in Δ^2, which are derived from the condition $N(E=0)>0$. We have for the polar phase

$$\frac{H_{c2}-H}{H_{c2}}<\frac{3L_{\rm pol}\ln n_F}{\pi^{3/2}\sqrt{n_F}}, \tag{73}$$

for the A-phase

$$\frac{H_{c2}-H}{H_{c2}}<\frac{2L_A}{3\pi^{3/2}\sqrt{n_F}}, \tag{74}$$

for the Scharnberg–Klemm phase

$$\frac{H_{c2}-H}{H_{c2}}<\frac{2L_{\rm SK}}{\pi^{3/2}\sqrt{n_F}}\left(1+\frac{1-\beta_0}{\sqrt{2}\gamma_0}\right) \tag{75}$$

and for the D-wave phase

$$\frac{H_{c2}-H}{H_{c2}}<\frac{4L_D}{7\pi^{3/2}\sqrt{n_F}}. \tag{76}$$

Now let us consider oscillating components in the density of states. Instead of integrating over n, we integrate over $\xi_n(k_z)=\xi$. Integrations over ξ and k_z are performed independently, and

$$\int\frac{dk_z}{2\pi}\exp\left(-2\pi i\frac{k_z^2}{2m\omega_c}r\right)=\frac{1}{2\pi\lambda\sqrt{r}}\exp\left(-i\frac{\pi}{4}\right).$$

The principal contribution to this integral comes from the energy band near the equatorial band of the Fermi sphere of width $\lambda^{-1}\ll k_F$. Since the order parameter in this region is almost constant, its dependence on k_z can be neglected in calculating the integral over k_z, and its value on the equatorial band of the Fermi sphere ($k_z=0$) can be used. For the Poisson sum component with number r ($r\neq 0$) we have

$$N^{(r)}(E)=\frac{\sqrt{m^3\omega_c}}{2\pi^2}\frac{(-1)^r}{\sqrt{r}}A_r(\Delta)\times\mathrm{Im}\exp\left[i\frac{2\pi r}{\omega_c}(E+\mu)-i\frac{\pi}{4}\right]\exp\left(-\frac{2\pi r\Gamma_{\rm imp}}{\omega_c}\right), \tag{77}$$

which determines oscillating components of the density of states with different $A_r(\Delta)$ in different superconducting phases.

In the polar phase the matrix element of the order parameter is zero in the equatorial plane, so in the approximation $\Gamma_{\rm imp} \ll \omega_c$ the value of $A_r(\Delta)$ is close to unity throughout the region where the expansion to the first order in Δ^2 applies [i.e., in fields H determined by condition (73)]. Thus, the oscillating component of the density of states in the polar phase is the same as in the normal metal.

For other phases under consideration, the value of $A_r(\Delta)$ coincides qualitatively with $A_r(\Delta)$ for conventional superconductors:

$$A_r^A(\Delta^A) = 1 - \frac{3}{16\sqrt{\pi}} \frac{(\Delta^A)^2}{\sqrt{n_F}\Gamma_{\rm imp}^2} \tag{78}$$

for the A-phase,

$$A_r^{\rm SK}(\Delta^{\rm SK}) = 1 - \frac{3}{16\sqrt{\pi}} \left(1 + \frac{1-\beta_0}{\sqrt{2}\gamma_0}\right)^2 \frac{(\Delta^{\rm SK})^2}{\sqrt{n_F}\Gamma_{\rm imp}^2}, \tag{79}$$

$$A_r^D(\Delta^D) = 1 - \frac{15}{2^6\sqrt{\pi}} \frac{(\Delta^D)^2}{\sqrt{n_F}\Gamma_{\rm imp}^2} \tag{80}$$

for the Scharnberg–Klemm and D-wave phases.

Using Eqs. (55), (63), and (66), we express $A_r(\Delta)$ as functions of magnetic field:

$$A_r^A(H) = 1 - \frac{\sqrt{\pi n_F}}{L_A} \frac{H_{c2}-H}{H_{c2}}, \tag{81}$$

$$A_r^{\rm SK}(H) = 1 - \frac{\sqrt{\pi n_F}}{L_{\rm SK}} \left(1 + \frac{1-\beta_0}{\sqrt{2}\gamma_0}\right)^{-1} \frac{H_{c2}-H}{H_{c2}}, \tag{82}$$

$$A_r^D(H) = 1 - \frac{14\sqrt{\pi n_F}}{3L_D} \frac{H_{c2}-H}{H_{c2}}. \tag{83}$$

The oscillating component of the magnetization is given by the equation

$$M^{(r)} = -\frac{\partial \Omega^{(r)}}{\partial H} = \frac{1}{2\pi^3}\left(\frac{e}{c}\right)^{3/2} \sqrt{H}\mu \frac{(-1)^r}{r^{1/2}} A_r(H) \times \sin\left(\frac{2\pi r\mu}{\omega_c} + \frac{\pi}{4}\right) \frac{2\pi^2 T/\omega_c}{\sinh(2\pi^2 Tr/\omega_c)} \times \exp\left(-\frac{2\pi r\Gamma_{\rm imp}}{\omega_c}\right). \tag{84}$$

The factor $A_r(H)$ is determined by Eqs. (81)–(83) for the A, SK, and D-wave phases, and equals unity for the polar phase.

7. CONCLUSIONS

Three types of superconducting states in p-wave superconductors have been analyzed: one state with an order parameter having a line of nodes on the Fermi sphere equator and two axial states whose order parameter have nodes at the Fermi sphere poles. The oscillation amplitude in the polar phase is the same as in a normal metal because electron states near the equator, which are the main contributors to the oscillating component of magnetization, are not affected by Cooper pairing. This result, naturally, also applies to all superconducting states in which the order parameter line node coincides with the line defining the extremal cross section of the Fermi surface. Thus, measurements of the de Haas–van Alphen effect can be used in identification of unconventional superconducting states.

In the axial phases the de Haas–van Alphen effect is suppressed more strongly than in the s-wave phase. The faster growth in the order parameter amplitude as a function of $1-H/H_{c2}$ in the axial phases is due to the absence of a divergence in the integral with respect to the polar angle calculated in solving the self-consistency equation. The D-wave phase has, in general, the same properties as the axial p-wave phases.

Finally, we emphasize once again that all calculations of this work were performed in the limit $T < \Gamma \ll \omega_c$ in the approximation linear in $(H_{c2}-H)/H_{c2}$.

The work was supported by the Ministry of Science and Technology of Russian Federation (Statistical Physics program) and the Russian Fund of Fundamental Research (Grant No. 96-15-96632 within the program in support of leading scientific schools).

APPENDIX

In this Appendix we perform the integration over the component of vector $\mathbf{R}=(X,Y,Z)$ in the equation

$$D_{ll'}^N = \int f_N(\mathbf{R})\phi_l^*(\mathbf{R})\phi_{l'}^*(\mathbf{R})d\mathbf{R}. \tag{85}$$

It is clear that the integral can be separated into integrals over each coordinate. We have

$$\int dZ\ \exp[-i(k_z+k_z')Z] = 2\pi\delta(k_z+k_z'),$$

$$\int dY \exp\left(\frac{2\pi i l Y}{a}\right) \exp\left[-i\left(q_y + \frac{\pi m}{a}\right)Y\right] \times \exp\left[-i\left(q_y' + \frac{\pi m'}{a}\right)Y\right] = 2\pi\delta_{2l,m+m'}\delta(q_y+q_y').$$

We have used the fact that $\vec{q}$ is the two-dimensional vector in the first Brillouin zone $-\pi/2a < q_y,\ q_y' < \pi/2a$, where a is the Abrikosov lattice constant. In integrating over the X coordinate we perform the change of variables $\xi = (X - \pi l\lambda^2/a)/\lambda$. By introducing the variable $\eta = [q_y + \pi(m - m')\lambda/2a]\lambda$ and noting that $q_y' = -q_y$, we express the integral over X as

$$I_{nn'}^N = \int d\xi \varphi_N(\sqrt{2}\xi)\varphi_n(\xi+\eta)\varphi_{n'}(\xi-\eta).$$

In calculating this integral, we use the generating function for Hermite polynomials and integrate over ξ. We obtain

$$\sum_{n,n',N}^{\infty} I^N_{nn'}(\eta)\frac{u^n}{n!}\frac{v^{n'}}{n'!}\frac{w^N}{N!} = \sqrt{\frac{\pi}{2}}\exp\left[-\frac{(u-v)^2}{2} + 2\sqrt{2}\eta\frac{u-v}{\sqrt{2}}\right] \times \exp(-\eta^2)\exp[\sqrt{2}w(u+v)].$$

By expressing the first exponential function on the right-hand side of this equation in terms of the generating function for Hermite polynomials as functions of $\sqrt{2}\eta$, we obtain the equation

$$\sum_{n,n',N}^{\infty} I^N_{nn'}(\eta)\frac{u^n}{n!}\frac{v^{n'}}{n'!}\frac{w^N}{N!} = \sum_{l=0}^{\infty} e^{-\eta^2}H_l(\sqrt{2}\eta)\frac{(u-v)^l}{\sqrt{2}^l l!} \times \exp[\sqrt{2}w(u+v)].$$

Expanding $\exp[\sqrt{2}w(u+v)]$ in a power series and equating the coefficients in front of terms with equal powers of w, u, and v, we find

$$I^N_{nn'}(\eta) = \sqrt{\pi}e^{-\eta^2}H_{n+n'-N}(\sqrt{2}\eta)\frac{2^N n!n'!}{\sqrt{2}^{n+n'+1}(n+n'-N)!} \times \sum_{j=0}^{N}(-1)^{n'+j-N}C_N^j C_{n+n'-N}^{n-j},$$

where the C_n^k are binomial coefficients.

We note that $\exp[ia(mq_x+m'q'_x)] = \exp[i(m+m')\times(q_x+q'_x)/2]\exp[i(m-m')(q_x-q'_x)/2]$ and replace summation over m and m' by summation over $m+m'$ and $m-m'$. The summation over $m+m'$ is trivial due to the Kronecker delta $\delta_{2l,m+m'}$, and since we have $\eta = \eta(m-m',q_y)$ the summation over l reduces to

$$\sum_l \exp[ila(q_x+q'_x)] = \frac{2\pi}{a}\delta(q_x+q'_x)$$

provided that $-\pi/a < q_x,\ q'_x < \pi/a$. Finally, we have

$$D_{ll'} = (2\pi)^3\delta(k_z+k'_z)\delta(\vec{q}+\vec{q}\,') \times \sqrt{\sqrt{2\pi}\frac{(n+n'-N)!}{2^{n+n'+1}n!n'!N!}}\frac{n!n'!}{(n+n'-N)!} \times \sum_{j=0}^{N}(-1)^{n'+j-N}C_N^j C_{n+n'-N}^{n-j} \times \sum_{\nu} e^{2i\nu q_x a}\varphi_{n+n'-N}(\sqrt{2}(q_y+\pi\nu/a)\lambda). \quad (86)$$

*)E-mail: mineev@landau.ac.ru

1)We recall that the de Haas–van Alphen effect in the normal state occurs when the temperature and sample purity satisfy $\omega_c > 2\pi^2 T$ and $\omega_c > 2\pi^2 T_D$. Here $\omega_c = eH/m^*c$ is the cyclotron frequency. The parameter $T_D = 1/2\pi\tau = \Gamma/2\pi$ is traditionally called the Dingle temperature, where τ is the electron mean free time and Γ is the corresponding level width. Planck's constant $\hbar$ is hereafter assumed to be unity.

2)The raising and lowering operators are defined as $\Pi_\pm = (\Pi_y \mp i\Pi_x)$, where $\Pi = (-i\partial_x, -i\partial_y + e/cA_y, -i\partial_z)$, and have the property $\Pi_+\phi_{nk\vec{q}}(\mathbf{R}) = \lambda^{-1}\sqrt{2(n+1)}\phi_{n+1,k,\vec{q}}(\mathbf{R})$, $\Pi_-\phi_{nk\vec{q}}(\mathbf{R}) = \lambda^{-1}\sqrt{2n}\phi_{n-1,k,\vec{q}}(\mathbf{R})$.

[1] R. Corcoran, N. Harrison, C. J. Haworth *et al.*, Physica B **206–207**, 534 (1995).
[2] K. Maki, Phys. Rev. B **44**, 2861 (1991).
[3] M. J. Stephen, Phys. Rev. B **45**, 5481 (1992).
[4] S. Dukan and Z. Tesanovic, Phys. Rev. Lett. **74**, 2311 (1995).
[5] M. G. Vavilov and V. P. Mineev, Zh. Éksp. Teor. Fiz. **112**, 1873 (1997) [JETP **85**, 1024 (1997)].
[6] K. Maki, Int. J. Mod. Phys. B **7**, 82 (1993).
[7] U. Brandt, W. Pesch, and L. Tewordt, Z. Phys. **201**, 209 (1967).
[8] K. Scharnberg and R. A. Klemm, Phys. Rev. B **22**, 5233 (1980).
[9] D. J. van Harlingen, Rev. Mod. Phys. **67**, 515 (1995).
[10] Yu. A. Bychkov and E. I. Rashba, Zh. Éksp. Teor. Fiz. **85**, 1826 (1983) [Sov. Phys. JETP **58**, 1062 (1983)].
[11] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics*, Dover Publ., Inc., New York (1993).

Translation provided by the Russian Editorial office.

Kinetics of first-order phase transitions in the asymptotic stage

S. A. Kukushkin*) and A. V. Osipov

Institute of Problems of Mechanical Engineering, Russian Academy of Sciences, 199178 St. Petersburg, Russia

(Submitted 24 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 2193–2208 (June 1998)

We construct an asymptotic theory that describes the kinetics of first-order phase transitions. The theory is a considerable refinement of the well-known Lifshits–Slezov theory. The main difference between the two is that the Lifshits–Slezov theory uses for the first integral of the kinetic equation an approximate solution of the characteristic equation, which is valid in the entire range of sizes except for the blocking point, i.e., it uses a nonuniformly applicable approximation. At the same time, the behavior of the characteristic solution near the blocking point determines the asymptotic behavior of the size distribution function of the nuclei for the new phase. Our theory uses a uniformly applicable solution of the characteristic equation, a solution valid at long times over the entire range of sizes. This solution is used to find the asymptotic behavior of all basic properties of first-order phase transitions: the size distribution function, the average nucleus size, and the nucleus density. © *1998 American Institute of Physics.* [S1063-7761(98)02006-X]

1. INTRODUCTION

The study of processes taking place in first-order phase transitions has prompted unflagging interest (see, e.g., the recent review articles in Refs. 1–4, a monograph,[5] and the references cited therein). There are several reasons for this. First, the physical phenomena that accompany the growth of the new phase are of a universal nature, since the evolution of a system undergoing a first-order phase transition obeys unified laws, which usually depend only on the mechanisms of new-phase growth rather than on the specific properties of the system, whether it is the early Universe, supersaturated vapor, an electron–hole liquid, or a biological system. Second, a first-order phase transition is a process consisting of several loosely linked stages, and almost every stage exhibits interesting nonlinear phenomena,[6–15] (in particular, nonlinear waves are generated[16,17]). Third, in view of the important role that first-order phase transitions play in engineering and various technologies, today we have a vast body of experimental data extracted from observations of the processes of formation, growth, and evolution of new-phase nuclei in various systems.

Theoretical studies of first-order phase transitions have been actively carried out for more than 16 years. Important results have been achieved during that period, which are basic to thermodynamics and the general theory of first-order phase transitions.[18–26] Unresolved problems abound, however. One is linked to the description of the kinetics of first-order phase transitions. The main difficulty here is that a typical first-order phase transition consists of at least three stages (without accounting for the symmetry changes in a first-order phase transition, the development of instabilities, and the interaction of new-phase nuclei). These are the stage in which nuclei of the new phase are formed, the stage of separate nucleus growth, and the Ostwald ripening stage, in which large nuclei grow at the expense of decay of small nuclei.[27] The processes in each stage are essentially nonlinear,[28,29] and, besides, each stage has its own features, which makes a unified description of the stages impossible (roughly speaking, each stage has its own small parameter). The first two stages were described theoretically in Ref. 13 within the scope of the theory of perturbations in $1/i_c$, where i_c is the number of particles in a critical nucleus under maximum supersaturation. In particular, the overall number of the growing nuclei and their size distribution were calculated. But it proved impossible in principle to incorporate the Ostwald ripening stage into this theory. Actually, this is not required, since in the Ostwald ripening phase the system ''forgets'' its initial state and, irrespective of this state, finds itself in a certain asymptotic regime.[27]

Many researchers have studied the Ostwald ripening stage,[5,30–43] but undoubtedly the contribution of I. M. Lifshits and V. V. Slezov to the development of this theory was the greatest. They were the first to calculate[31] the asymptotic behavior of the basic characteristics of the first-order phase transition, the supersaturation, the critical size, and the concentration of the nuclei of the new phase, and proved that for long times the nucleus-size distribution function has a self-similar form, independent of the initial conditions. They also obtained an estimate for this function in what became known as the zeroth approximation (this term was used in several later papers). The essence of the Lifshits–Slezov (LS) approximation is as follows. The time-dependent term in the characteristic equation, whose solution yields the first integral of the kinetic equation, is ignored (Eq. (15) in Ref. 31), since with the passage of time this term tends to zero. In this connection Lifshits and Slezov noted that such a simplifying assumption is possible only far from the blocking point, since at the blocking point the other terms tend to zero at the same rate. From the viewpoint of rigorous perturbation

theory,[44] the LS approximation corresponds to the first term of the nonuniformly admissible series in powers of t, so that the solution found cannot be the asymptotic one (the blocking point is the source of nonuniformity). The fact that the size distribution function in the Lifshits–Slezov theory is not asymptotic has been mentioned in several papers. Kahlweit's paper[34] is the best-known. However, Kahlweit was convinced that the source of nonuniformity is not the blocking point, and for this he was rightly criticized by Slezov.[45] Marqusee and Ross[36] proposed seeking the solution of the kinetic equation immediately as a power series in t:

$$g(r,t)=t^{-y_0}[g_0(z)+t^{-y_1}g_1(z)+t^{-y_2}g_2(z)+\cdots], \quad (1)$$

where $z=r/t^{\beta}$, g is the distribution of the nuclei of the new phase over the radius r, t is the time variable, and β and $y_0, y_1, \ldots$ are positive constants. Applying an ingenious method, they were able to decouple the equations for g_0 and g_1 and find their solution. It was established that g_0 coincides with the result of the LS theory, from which Marquese and Ross concluded that this solution was the asymptotic one. Actually, this is not the asymptotic solution but the solution corresponding to the zeroth approximation, since the series (1) is divergent, at least at the blocking point. Marquese and Ross[36] did not study the convergence of (1), and the blocking point makes the series (1) nonuniformly admissible even in the next order, i.e., $g_1(z)/g_0(z)\to\infty$ as z tends to the blocking point. As is well known, this is the most important aspect of any asymptotic theory or perturbation theory.[44] The same error was also made in some other works on the asymptotic theory of first-order phase transitions.

Thus, the problem of finding an asymptotic solution for the nucleus-size distribution function in first-order phase transitions has yet to be solved. The present paper treats this problem. The plan is as follows. We start by examining the evolution of the nuclei of the new phase in the layered growth of a thin film on a substrate from supersaturated vapor subjected to sudden supersaturation. We have chosen this example because it is a simple and yet realistic first-order phase transition with the simplest new phase growth mechanism possible. The essence of the proposed theory is illustrated by this example of layered film growth most vividly. Then we briefly analyze the general case of a first-order phase transition.

2. ASYMPTOTIC THEORY OF FIRST-ORDER PHASE TRANSITIONS IN THE CASE OF LAYERED FILM GROWTH

Suppose that initially n_1 atoms are adsorbed on a unit area of a solid substrate, with n_1 greater than n_{1e}, the equilibrium concentration of adsorbed atoms (adatoms). Then a phase transition begins in the system, and in the case of layered growth[1] disklike nuclei of the new phase grow on the surface. By $g(R,t)$ we denote the nucleus distribution function over the disk radius R at time t. Then, if the atoms do not leave the substrate, which is the case if the temperature T_s of the supersaturated vapor of adatoms and the substrate is low, the conservation of matter on the substrate is expressed by the following law:

$$\frac{w}{h}n_1+\pi\int_0^{\infty}R^2g(R,t)dR=Z, \quad (2)$$

where w is the volume occupied by a single atom in the new phase, h is the disk height, Z the extent to which the adsorbate fills the substrate (in the given model Z is assumed to be constant and, to simplify matters, much larger than wn_{1e}/h in the Ostwald ripening stage). Let $\xi=n_1/n_{1e}-1$ be the degree of supersaturation in the system and σ the interphase energy per unit length of the periphery of a disk. Then the radius R_c of a critical nucleus in the Ostwald ripening stage is equal to $\sigma w/k_BT_sh\xi$, where k_B is Boltzmann's constant. We introduce R_*, the characteristic radius in the system, equal to $\sigma w^2n_{1e}/k_BT_sh^2Z$. In this case Eq. (2) becomes

$$\frac{R_*}{R_c}+\frac{\pi}{Z}\int_0^{\infty}R^2g(R,t)dR=1. \quad (3)$$

In turn, the nucleus-size distribution function obeys the continuity equation in the size space:[31]

$$\frac{\partial g}{\partial t}+\frac{\partial(v_Rg)}{\partial R}=0, \quad (4)$$

$$g(R,0)=N_0g_0(R), \quad \int_0^{\infty}g_0(R)dR=1. \quad (5)$$

Here we have allowed for the fact that in the Ostwald ripening stage the nucleation process has already been completed and that there are N_0 nuclei, whose size distribution is given by the normalized function $g_0(R)$. Both N_0 and $g_0(R)$ are calculated by perturbation-theory expansions in $\varepsilon=1/\pi R_c^2hw$ (Ref. 13). Our goal is to describe the evolution of the existing nuclei at a later stage. When the so-called boundary-kinetics mechanism[1] is realized, the rate of separate nucleus growth is

$$v_R=\frac{R_*^2}{t_*}\left(\frac{1}{R_c}-\frac{1}{R}\right), \quad (6)$$

where $t_*=l_0R_*/\alpha ZD$ is the characteristic growth time, l_0 is the lattice parameter of the substrate, α is the probability that an adatom attaches itself to a nucleus ($\alpha\ll 1$ in the given growth mechanism), and D is the coefficient of diffusion of adatoms over the substrate. At this point we introduce the dimensionless variables

$$T=\frac{t}{t_*}, \quad r=\frac{R}{R_*}, \quad r_c=\frac{R_c}{R_*}, \quad F=\frac{gR_*}{N_0} \quad (7)$$

and a dimensionless constant $\delta=\pi N_0R_*^2\alpha^{-1}\ll 1$. The kinetic equation and the law of adatom conservation then become

$$\frac{\partial F}{\partial T}+\frac{\partial}{\partial r}\left[\left(\frac{1}{r_c(T)}-\frac{1}{r}\right)F\right]=0, \quad (8)$$

$$r_c(T)=\left[1-\delta\int_0^{\infty}r^2F(r,T)dr\right]^{-1}. \quad (9)$$

Plugging (9) into (8) yields a master equation for the kinetics of first-order phase transitions in the Ostwald ripening stage (for the given type of phase transition). This equation de-

scribes the following physical processes. Initially, due to the smallness of δ, the critical radius r_c remains essentially unchanged ($r_c \approx 1$) and the size of most nuclei is greater than r_c, so that they grow. The distribution function F moves in the size space as an integral whole at a rate $1/r_c \approx 1$, i.e., $F(r,T) \approx F_0(r-T)$. This stage is described by nucleation theory.[13] The integral on the right-hand side of Eq. (9) increases with T, and so does r_c, which gradually catches up with the main part of the size spectrum. The nuclei smaller than the critical size begin to be evaporated, adding to the supply of adatoms. After time T_r of approximately $1/\sqrt{\delta}$ has elapsed, the critical size catches up completely with the main part of the size spectrum, i.e.,

$$\delta \int_0^\infty r^2 F dr \approx 1.$$

This means that the supply of adatoms that were initially on the substrate has been exhausted. Now supercritical-size nuclei grow at the expense of subcritical-size nuclei. The increase in the critical size is limited by the average size of the nuclei, i.e., the system evolves in a self-consistent manner. This is the Ostwald ripening stage.

To describe this stage analytically, it is convenient to change variables, so that the blocking point on the size axis can be isolated explicitly. There are several ways of doing this, one of which was suggested by Lifshits and Slezov.[31] Below we use another change of variables, which we believe has more physical meaning and is simpler:

$$f = 2 r_c F, \quad u = \frac{r}{2 r_c}, \quad \tau = \frac{1}{4} \int_0^T r_c^{-2}(T') dT'. \tag{10}$$

Then Eqs. (8) and (9) take the form

$$\frac{\partial f}{\partial \tau} - \frac{\partial}{\partial u}\left\{\left[\frac{(u-1)^2}{u} + \gamma(\tau) u\right] f\right\} = 0, \tag{11}$$

$$r_c(\tau) = \sqrt{\frac{1 - 1/r_c}{4\delta \int_0^\infty u^2 f(u,\tau) du}}, \tag{12}$$

$$\gamma(\tau) = \frac{d \ln r_c}{d\tau} - 1. \tag{13}$$

From Eq. (11) we see that here $u = 1$ is the blocking point. Note that for $r_c \gg 1$ the expression for γ simplifies,

$$\gamma(\tau) = -\frac{1}{2} \frac{d \ln \int_0^\infty u^2 f(u,\tau) du}{d\tau} - 1, \tag{14}$$

and becomes independent of the single parameter δ of the first-order phase transition. This explains why all systems undergoing a first-order phase transition with the same new-phase growth mechanism evolve in the same way. The general approach to solving the system (11)–(13) can be as follows. First the kinetic equation (11) with a given initial condition is solved. Then the solution is substituted in (12). If we then substitute the resulting expression in (13), we arrive at an equation for the function $\gamma(\tau)$. By finding at least an arbitrary solution of this equation we can describe the kinetics of a first-order phase transition in the stage of separate growth of new-phase nuclei and in the Ostwald ripening stage.

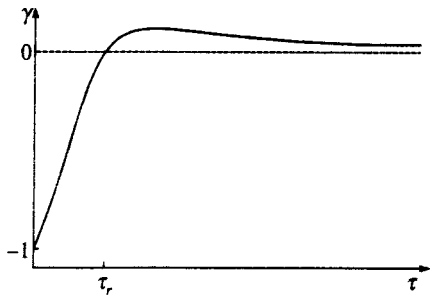


FIG. 1. A schematic of the γ vs. τ dependence.

The general solution of Eq. (11) has the form

$$f(u,\tau) = f_0(C(u,\gamma(\tau))) \frac{\partial C}{\partial u}, \tag{15}$$

where $C(u,\gamma)$ is an integral of the kinetic equation, which is found by solving the characteristic equation

$$\frac{du}{d\tau} = -\frac{(u-1)^2}{u} - \gamma(\tau) u, \tag{16}$$

$$u(\tau_0) = C. \tag{17}$$

It is Eq. (16) that is responsible for the evolution of the system in the final stage of the first-order phase transition. Applying the Lifshits–Slezov method[31] to the given case, we can find the asymptotic behavior of the function γ as $\tau \to \infty$:

$$\gamma(\tau) \to \frac{1}{4\tau^2}. \tag{18}$$

If this condition is not met, the law of conservation of matter (12) is violated. Figure 1 depicts a schematic of the function $\gamma(\tau)$ for the given case. At time $\tau = 0$ the value of γ is -1, then γ increases, and at a certain time τ_r reaches the value $\gamma = 0$. After γ reaches its maximum value $\gamma_{\max} > 0$, it decreases and finally reaches its asymptotic value $\gamma = 1/4\tau^2$. Figure 2 depicts the solution of the characteristic equation (16) for different initial conditions for the simple model function

$$\gamma(\tau) = \frac{\tau - 1}{1 + 4\tau^3}.$$

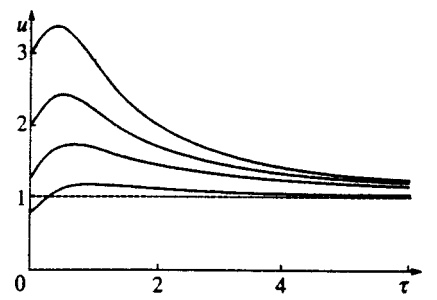


FIG. 2. Solution of the characteristic equation (16) for different initial conditions.

The initial increase in u in Fig. 2 corresponds to the stage of separate nucleus growth by attaching new adatoms that appear on the substrate at the moment when supersaturation emerges (at this stage $r_c \approx 1$). A further increase in the critical radius r_c slows down the rate at which u increases, and later reduces u. This reduction gradually reaches the self-consistent asymptotic regime, corresponding to the Ostwald ripening regime. In this stage u slowly approaches the blocking point $u=1$. In the final analysis, the rate of this approach determines the nucleus-size distribution function in terms of the integral $C(u,\gamma)$. To study this dependence analytically, it is convenient to make the following change of variables:

$$v = \frac{1}{u-1} - \ln|u-1|.$$

Then

$$\frac{dv}{d\tau} = 1 + \gamma(\tau)[1+\psi(v)]^2, \tag{19}$$

where ψ is the solution of the transcendental equation

$$\psi + \ln|\psi| = v. \tag{20}$$

For long times ($\tau > 50$) we can assume $(1+\psi)^2 = v^2$, i.e., in this approximation

$$\frac{dv}{d\tau} = 1 + \gamma(\tau)v^2. \tag{21}$$

To reduce this Riccati equation to a linear one, we do another change of variables, $v = y/y'$. The equation becomes

$$y'' + \gamma(\tau)y = 0. \tag{22}$$

This second-order linear equation is the central equation in the theory of Ostwald ripening, since it determines the dependence of y on γ and hence of C on γ at long times. This means it controls the size distribution function for the new-phase nuclei.

Now let us turn to methods of solving Eq. (22). The simplest one was proposed by Lifshits and Slezov. They suggested ignoring the second term in (22) entirely, since at long times γ tends to zero.[31] In this case $y_{\rm LS}(\tau) = A\tau + B$, where A and B are integration constants, and $v_{\rm LS} = \tau + B/A$. Hence

$$v - \tau = -\frac{1}{1-u} - \tau - \ln|1-u| \tag{23}$$

is the desired integral in the zeroth Lifshits–Slezov approximation. It is convenient to write this integral in the form

$$C_{\rm LS} = \frac{1}{1-u}\exp\left(-\frac{1}{1-u} - \tau\right). \tag{24}$$

By inserting (24) in the general solution (15) and the result into the conservation law (14) we arrive at an expression for f_0:

$$f_{0\,LS}(C) = \begin{cases} \text{const}\cdot C^2, & C > 0, \\ 0, & C \leqslant 0. \end{cases}$$

Only in this case will the ''initial condition'' for the linear equation, $\gamma|_{\tau\to\infty} \to 0$ (the law of conservation of matter), be met. Thus, the size distribution function is given by the following expression:

$$f(u,\tau) = \begin{cases} e^{-2\tau}\varphi(u), & 0 \leqslant u < 1, \\ 0, & u \geqslant 1. \end{cases}$$

As (24) implies, the function φ in the Lifshits–Slezov approximation is[1]

$$\varphi_{\rm LS}(u) = \frac{2u}{(1-u)^4}\exp\left(-\frac{2u}{1-u}\right) \tag{25}$$

(the value of the constant is usually selected in such a way that φ is normalized to unity). The same result can be obtained by using the Marquese–Ross approach.[36] However, the result is in no way asymptotic since the starting equation $y_{\rm LS} = A\tau + B$ is invalid for long times. Indeed, for the given rate at which γ tends to zero we have $y \to \text{const}\cdot\sqrt{\tau}\,\ln\tau$, i.e., the point $\tau = \infty$ is the source of nonuniformity.[44]

For most first-order phase transitions $\gamma(\tau)$ is a slowly varying function. The main thing is that the characteristic time of variation of the dimensions of an individual nucleus is usually much shorter than the characteristic time of variation of the critical nucleus (the so-called time-scale hierarchy[46]). For this reason it is convenient to represent the function γ in (22) in the form $\gamma(\varepsilon\tau)$ and seek the solution of Eq. (22) in the form of a uniformly applicable series in ε (Ref. 44). This equation has been analyzed by a number of researchers,[47–49] who found that the functional dependence of y on γ is determined entirely by the points on the τ axis at which γ vanishes (the so-called regression points) and by the behavior of γ at these points. For the given type of first-order phase transition, γ has two singular points: $\tau = \tau_r$ and $\tau = \infty$ (see Fig. 1). By applying the Langer transformation[47] we can find the dependence of y on γ in the zeroth order in ε over the entire range of τ: for $0 < \tau < \tau_r$, $\tau_r < \tau < \infty$, in the vicinity of τ_r, and as $\tau \to \infty$. The four functions could, in principle, give a complete description of the kinetics of the first-order phase transition together with Eqs. (12) and (13) since, in contrast to the Lifshits–Slezov solution, they are uniformly applicable.[44] However, this approach is impractical, if only because the value of τ_r is not known beforehand. The right-hand side of Eq. (13), after (12) and (15) have been substituted in it, becomes an extremely complicated function of γ with τ_r unknown. For this reason we are forced to limit ourselves to studying the asymptotic behavior of first-order phase transitions. The fact that there is a regression point leads to a situation in which the system ''forgets'' its initial state and reaches a certain asymptotic regime (models with a strictly negative γ are known to generate nonphysical divergences[30]).

Thus, combining the Langer method with (18) makes it possible to use Eq. (22) to find the long-time asymptotic behavior of y:

$$y \to \text{const}\cdot\frac{\int\sqrt{\gamma(\tau)}\,d\tau}{\sqrt[4]{\gamma(\tau)}}, \tag{26}$$

$$v \to -\frac{4\gamma}{\gamma'} \to 2\tau. \tag{27}$$

The fact that v does not tend to τ as $\tau \to \infty$ (as it does in the Lifshits–Slezov theory) but to 2τ also follows directly from Eq. (21) combined with (18). We also see that the difference of the given behavior from that predicted by the Lifshits–Slezov theory is that as $\tau \to \infty$ (i.e., in the vicinity of the blocking point), the term that Lifshits and Slezov discards in not small and provides a contribution equal to that of the term taken into account. A numerical solution of Eq. (16) (see Fig. 2) also implies that u tends to $1-1/2\tau$ rather than to $1-1/\tau$. Thus, the asymptotic behavior of the integral of the kinetic equation is given by the following expression:

$$C = \frac{1}{1-u}\exp\left(-\frac{1}{1-u} - 2\tau\right), \tag{28}$$

so that for C positive we have $f_0(C) = \text{const} \cdot C$, and the desired asymptotic size distribution function is

$$\varphi(u) = \frac{u}{(1-u)^3}\exp\left(-\frac{u}{1-u}\right). \tag{29}$$

What sets this function apart from the Lifshits–Slezov function is the difference in the average nucleus size:

$$\int_0^\infty u\varphi_{\mathrm{LS}}(u)du = \frac{1}{2},$$

$$\overline{u} = \int_0^\infty u\varphi(u)du \approx 0.5963, \quad (\overline{u^2} \approx 0.3854).$$

In other words, while in the Lifshits–Slezov theory the average nucleus size coincides with the critical size, in our theory the average size is shifted toward larger values by approximately 20%. To make the comparison of the results with the experimental data more convenient, it is advisable to introduce a new variable, $x = u/\overline{u} = R/\overline{R}$ (in the Lifshits–Slezov theory $x = 2u$), and recalculate the distribution functions:

$$\varphi_{\mathrm{LS}}(x) = \frac{8x}{(2-x)^4}\exp\left(-\frac{2x}{2-x}\right), \tag{30}$$

$$\varphi(x) = \frac{x_0 x}{(x_0-x)^3}\exp\left(-\frac{x}{x_0-x}\right), \tag{31}$$

where $x_0 = 1/\overline{u} \approx 1.677$. These functions are depicted in Fig. 3. We see that our asymptotic distribution function is higher and narrower than the Lifshits–Slezov function. As predicted by Lifshits and Slezov, the region where the two functions differ most strongly lies near the blocking point $x = x_0$ (approximately in the interval from 1.4 to 2). In all fairness, it must be said that the absolute error of the Lifshits–Slezov function is moderate and often smaller than the experimental error. Nevertheless, experimenters have repeatedly pointed out that the experimental size distribution function actually passes above the Lifshits–Slezov function.[5] In fact, this served as the impetus for improving the theories of Ostwald ripening.[34,37–40]

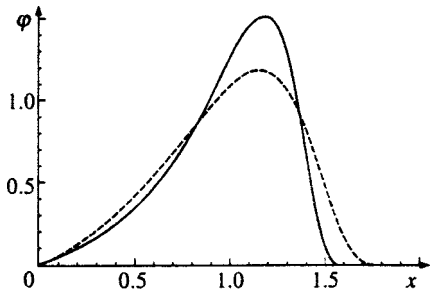


FIG. 3. The asymptotic nucleus distribution function over $x = R/\overline{R}$ (the solid curve corresponds to the results obtained in the present theory, and the dashed curve corresponds to the results obtained by the Lifshits–Slezov theory).

To be precise, the asymptotic formula (29) is valid only for finite or rapidly decreasing initial distribution functions. As shown in Ref. 45, the ''tails'' of the very slowly decreasing initial distribution functions may change the asymptotic behavior (18) of γ, with the same being true for the size distribution functions. However, according to the general nucleation theory,[13] finite nucleation time leads to the finiteness of the initial distribution function.

The above results make it possibly to easily establish the asymptotic behavior of other important characteristics of a first-order phase transition: the average size and the concentration of the new-phase nuclei. Indeed, from (10) and (13) it follows that

$$r_c \to e^{\tau}, \quad T \to e^{2\tau}, \quad \tau \to \frac{1}{2}\ln t \tag{32}$$

as $\tau \to \infty$. This readily leads to the desired asymptotic behavior:

$$\overline{R}(t) = \text{const} \cdot \sqrt{t}, \tag{33}$$

$$N(t) = \frac{\text{const}}{t}. \tag{34}$$

In this asymptotic theory it is impossible to find the constants in (33) and (34), since this requires information about the regression point τ_r. Hence it is easier to find them from experiment by calculating the points of intersection of the straight lines $\ln \overline{R}$ and $\ln N$ with the ordinate $\ln t = 0$.

Now let us describe the asymptotic stage of Ostwald ripening for a general first-order phase transition.

3. ASYMPTOTIC THEORY OF FIRST-ORDER PHASE TRANSITIONS FOR AN ARBITRARY NEW-PHASE GROWTH MECHANISM

The continuity equation in the size space and the law of conservation of matter for arbitrary (but real) first-order phase transitions can be written as follows:

$$\frac{\partial g}{\partial t} + \frac{\partial (v_R g)}{\partial R} = 0, \tag{35}$$

$$g(R,0) = N_0 g_0(R), \quad \int_0^\infty g_0(R)dR = 1, \tag{36}$$

$$v_R=\frac{R_*^2}{t_*}\left(\frac{R_*}{R}\right)^{p-1}\left(\frac{1}{R_c}-\frac{1}{R}\right), \tag{37}$$

$$\frac{R_*}{R_C}+\beta_d\int_0^\infty R^d g(R,t)dR=\left(\frac{t}{t_*}\right)^n. \tag{38}$$

Here R_*, t_*, and β_d are the constants characterizing a first-order phase transition, p is the new-phase nucleus growth index (the cases $p=1, 2, 3$ correspond to all known growth mechanisms[1]), d is the dimensionality of the nuclei ($d=1, 2, 3$; see Ref. 5), $n\geqslant 0$ is the growth index for the total mass of matter (it is assumed that at least at long times this mass increases according to a power law[5,43]). We introduce the dimensionless variables

$$T=\frac{t}{t_*},\quad r=\frac{R}{R_*},\quad r_c=\frac{R_c}{R_*},\quad F=\frac{gR_*}{N_0} \tag{39}$$

and the dimensionless constant $\delta=\beta_d N_0 R_*^d$. Then Eqs. (35)–(38) become

$$\frac{\partial F}{\partial T}+\frac{\partial}{\partial r}\left[\frac{1}{r^{p-1}}\left(\frac{1}{r_c(T)}-\frac{1}{r}\right)F\right]=0, \tag{40}$$

$$r_c(T)=\left[T^n-\delta\int_0^\infty r^d F(r,T)dr\right]^{-1}. \tag{41}$$

Next, following the general scheme described above, we introduce the variables

$$u=\frac{p}{p+1}\frac{r}{r_c},\quad f=\frac{p+1}{p}r_c F,$$

$$\tau=\frac{p+1}{2}\int_0^T\left[\frac{p}{(p+1)r_c(T')}\right]^{p+1}dT' \tag{42}$$

and write (40) and (41) in the following form:

$$\frac{\partial f}{\partial\tau}-\frac{\partial}{\partial u}\left\{\left[\frac{2}{p(p+1)u^p}(p-(p+1)u+u^{p+1})+\gamma(\tau)u\right]f\right\}=0, \tag{43}$$

$$r_c(\tau)=\left[\frac{\left(\frac{2}{p+1}\right)^n\left(\frac{p+1}{p}\right)^{n(p+1)+d}(\int_0^\tau r_c^{p+1}(\tau')d\tau')^n-\left(\frac{p}{p+1}\right)^d\frac{1}{r_c}}{\delta\int_0^\infty u^d f(u,\tau)du}\right]^{1/d}, \tag{44}$$

$$\gamma(\tau)=\frac{d\ln r_c}{d\tau}-\frac{2}{p(p+1)}. \tag{45}$$

Equations (44) and (45) imply that for large values of r_c the function γ ceases to depend on δ and the dependence on τ gets simpler:

$$\gamma(\tau)=\frac{n}{d}\frac{d\ln\int_0^\tau r_c^{p+1}(\tau')d\tau'}{d\tau}-\frac{1}{d}\frac{d\ln\int_0^\infty u^d f(u,\tau)du}{d\tau}-\frac{2}{p(p+1)}. \tag{46}$$

The general solution of Eq. (43) has the form

$$f(u,\tau)=f_0(C(u,\gamma(\tau)))\frac{\partial C}{\partial u}, \tag{47}$$

where $C(u,\gamma)$ is an integral of the kinetic equation, which is found by solving the characteristic equation

$$\frac{du}{d\tau}=-\frac{2}{p(p+1)u^p}[u^{p+1}-(p+1)u+p]-\gamma u. \tag{48}$$

In the neighborhood of the blocking point $u=1$ Eq. (48) becomes independent of p:

$$\frac{du}{d\tau}=-(u-1)^2-\gamma. \tag{49}$$

The Lifshits–Slezov method[31] makes it possible to use Eq. (49) to find the function γ as $\tau\to\infty$:

$$\gamma\to\frac{1}{4\tau^2}. \tag{50}$$

If this condition is not met, the law of conservation of matter is violated. This leads to the following asymptotic behavior of r_c and T:

$$r_c\to\exp\frac{2\tau}{p(p+1)}, \tag{51}$$

$$T\to\exp\frac{2\tau}{p}. \tag{52}$$

Now from (52) we find that

$$\tau\to\frac{p}{2}\ln t, \tag{53}$$

and hence

$$\overline{R}(t)\to\text{const}\cdot t^{1/(p+1)}, \tag{54}$$

$$N(t)\to T^n r_c^{-d}\to\text{const}\cdot t^{-(d/(p+1)-n)}. \tag{55}$$

Of course, at $d=3$ formulas (54) and (55) coincide with the corresponding results of the Lifshits–Slezov theory.[5] Note that the Ostwald ripening stage is realized only in the case of so-called weak sources,

$$n<\frac{d}{p+1}, \tag{56}$$

when the critical size catches up with the main part of the nucleus size spectrum. Otherwise, new particles enter the system so fast that supersaturation increases instead of de-

creasing, the nucleation stage is not terminated, and Ostwald ripening does not begin. The same condition follows from the analysis of the nucleation stage.[10]

Now let us establish the asymptotic behavior of the distribution function. We perform the transformation $u \to v_p(u)$, which maps the blocking point to infinity:

$$v_p = -\frac{p(p+1)}{2} \int_0^u \frac{x^p dx}{x^{p+1}-(p+1)x+p}. \tag{57}$$

Then the characteristic equation (48) becomes

$$\frac{dv_p}{d\tau} = 1 + \gamma(\tau)\frac{p(p+1)}{2}\frac{u^{p+1}(v_p)}{u^{p+1}(v_p)-(p+1)u(v_p)+p}. \tag{58}$$

At long times the characteristic equation becomes the Riccati equation

$$\frac{dv_p}{d\tau} = 1 + \frac{v_p^2}{4\tau}, \tag{59}$$

whose general solution is

$$v_p = 2\tau\left(1 + \frac{1}{\ln \tau + A_0}\right), \tag{60}$$

where A_0 is the integration constant. At long times v tends to 2τ (the Lifshits–Slezov theory yields $v_p \to \tau$), so that

$$C(u,\tau) = \exp\frac{v_p(u)-2\tau}{p} \tag{61}$$

is the asymptotic value of the integral of the kinetic equation. Plugging (61) into the general solution (47), from the initial condition (the law of mass conservation) we find that

$$f_0(C) = \begin{cases} \text{const}\cdot C^{d/(p+1)-n}, & C>0, \\ 0, & C\leqslant 0. \end{cases}$$

Hence the asymptotic behavior of the new-phase nucleus distribution function has the following form:

$$f(u,\tau) = \begin{cases} \exp\left[-\frac{2\tau}{p}\left(\frac{d}{p+1}-n\right)\right]\varphi_p(u), & u<1, \\ 0, & u\geqslant 1, \end{cases}$$

where

$$\varphi_p(u) = \frac{u^p}{u^{p+1}-(p+1)u+p}\exp\left[\frac{d-n(p+1)}{2} \times \int_u^{u_p}\frac{x^p dx}{x^{p+1}-(p+1)x+p}\right]. \tag{62}$$

The constant v_p in (62) is usually found from the normalization condition

$$\int_0^1 \varphi_p(u)du = 1.$$

As expected, at $p=1$, $d=2$, $n=0$, and $u_p=0$ the expression (62) becomes (29). An analysis of the φ_p vs. u dependence for $p=1,2,3$, which correspond to real new-phase growth mechanisms, is given in the Appendix. Figure 4 depicts the size distribution functions for $p=2$ and $d=3$ and different values of n. We see that the closer n is to its limit $n_c = d/(p+1)$, the sharper the peak (which moves toward the blocking point) in the $\varphi(u)$ curve. Hence by changing n we can control the structure of the new phase.[5] At $n \geqslant n_c$ the Ostwald ripening does not even start, so that the distribution function in this case evolves in accordance with the predictions of nucleation theory.[13]

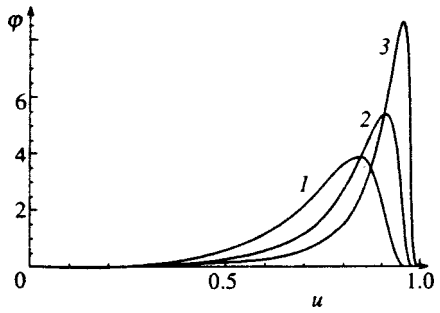


FIG. 4. The asymptotic function of nuclei distribution for $d=3$ and $p=2$: curve *1*, $n=0$; curve *2*, $n=0.55$; and curve *3*, $n=0.8$.

Thus, the use of a uniformly admissible approximation for the solution of the characteristic equation makes it possible to find the asymptotic behavior of the basic characteristics of first-order phase transitions: the new-phase nucleus concentration, the average size of the nuclei, and the size distribution function.

The authors would like to express their deep gratitude to V. V. Slezov for his critical analysis and useful results. The work was made possible by grants from the Russian Fund for Fundamental Research (Grants Nos. 96-03-32396 and 98-03-32791) and the Integration Fund (Grant No. 589).

APPENDIX

To find the distribution function φ_p, we must calculate the integral in (62) (or (57)):

$$I_p = \int \frac{x^p dx}{x^{p+1}-(p+1)x+p}. \tag{A1}$$

Note that for positive integral values of p we have

$$x^{p+1}-(p+1)x+p = (x-1)^2\sum_{n=1}^{p} nx^{p-n}, \tag{A2}$$

so that

$$I_1 = \int \frac{x dx}{(x-1)^2} = \frac{1}{1-x} + \ln|1-x|, \tag{A3}$$

$$I_2 = \int \frac{x^2 dx}{(x-1)^2(x+2)} = \frac{1}{3(1-x)} + \frac{4}{9}\ln|x+2| + \frac{5}{9}\ln|1-x|, \tag{A4}$$

$$I_3 = \int \frac{x^3 dx}{(x-1)^2(x^2+2x+3)} = \frac{1}{6(1-x)} + \frac{7}{18} \times \ln|1-x| + \frac{11}{36}\ln(x^2+2x+3) + \frac{\sqrt{2}}{36}\arctan\frac{x+1}{\sqrt{2}}, \tag{A5}$$

etc. This sequence can be continued, but for $p \geqslant 4$ the growth mechanism becomes too exotic, and, besides, the real roots of the equation $\Sigma n x^{p-n} = 0$ do not obey simple relationships. Plugging (A3)–(A5) into (62), we get

$$\varphi_1(u) = \frac{(d/2-n)u}{(1-u)^{2+d/2-n}} \exp - \left(\frac{d}{2}-n\right)\frac{u}{1-u}, \tag{A6}$$

$$\varphi_2(u) = \frac{4^{(d-3n)/9}(d-3n)u^2}{2(1-u)^{2+5(d-3n)/18}(2+u)^{1+2(d-3n)/9}} \times \exp\left(-\frac{d/3-n}{2}\frac{u}{1-u}\right), \tag{A7}$$

$$\varphi_3(u) = \frac{C_3 u^3 \exp\left[-\frac{\sqrt{2}}{18}\left(\frac{d}{4}-n\right)\arctan\frac{u+1}{\sqrt{2}}\right]}{(1-u)^{2-(7/9)(d/4-n)}(u^2+2u+3)^{1+(11/18)(d/4-n)}} \times \exp\left(-\frac{d/4-n}{3}\frac{u}{1-u}\right), \tag{A8}$$

with C_3 the normalization constant.

*)E-mail: ksa@math.ipme.ru

[1] S. A. Kukushkin and A. V. Osipov, Prog. Surf. Sci. **51**, 1 (1996).
[2] S. A. Kukushkin and A. V. Osipov, Khim. Fiz. **15**, 5 (1996).
[3] A. I. Olemskoĭ and I. V. Koplyk, Usp. Phys. Nauk **165**, 1105 (1995) [Phys. Usp. **38**, 1061 (1995)].
[4] V. G. Boĭko, Kh.-I. Mogel', V. M. Sysoev, and A. V. Chalyĭ, Usp. Phys. Nauk **161**, No. 2, 77 (1991) [Sov. Phys. Usp. **34**, 141 (1991)].
[5] S. A. Kukushkin and V. V. Slezov, *Dispersive Systems on Solid Surfaces: Mechanisms of Thin Film Formation* [in Russian], Nauka, St. Petersburg (1996).
[6] J. J. Kozak, Adv. Chem. Phys. **40**, 229 (1979).
[7] K. Binder and D. Stauffer, Adv. Phys. **25**, 343 (1976).
[8] Yu. M. Mikhaĭlova and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **59**, 1368 (1970) [Sov. Phys. JETP **32**, 747 (1971)].
[9] V. P. Skripov and A. V. Skripov, Usp. Phys. Nauk **128**, 193 (1979) [Sov. Phys. Usp. **22**, 389 (1979)].
[10] A. V. Osipov, Thin Solid Films **227**, 111 (1993).
[11] A. V. Osipov, Thin Solid Films **231**, 173 (1995).
[12] S. A. Kukushkin and A. V. Osipov, Phys. Rev. E **53**, 4964 (1996).
[13] S. A. Kukushkin and A. V. Osipov, J. Chem. Phys. **107**, 3247 (1997).
[14] S. A. Kukushkin and A. V. Osipov, J. Phys. Chem. Solids **58**, 1115 (1997).
[15] A. S. Zel'tser, T. K. Soboleva, and A. É. Filippov, Zh. Éksp. Teor. Fiz. **108**, 356 (1995) [JETP **81**, 193 (1995)].
[16] S. A. Kukushkin and A. V. Osipov, Thin Solid Films **227**, 119 (1993).
[17] S. A. Kukushkin and A. V. Osipov, Surf. Sci. **329**, 135 (1995).
[18] L. D. Landau, Zh. Éksp. Teor. Fiz. **7**, 627 (1937).
[19] Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **12**, 525 (1942).
[20] C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404 (1952).
[21] J. W. Cahn, Acta Metall. **3**, 255 (1961).
[22] N. G. van Kampen, Phys. Rev. **135**, A362 (1964).
[23] J. S. Langer, Ann. Phys. (N.Y.) **41**, 108 (1967).
[24] M. E. Fisher, Physics (Long Island City, NY) **3**, 255 (1967).
[25] A. Z. Patashinskiĭ and B. I. Shumilo, Zh. Éksp. Teor. Fiz. **77**, 1417 (1979) [Sov. Phys. JETP **50**, 712 (1979)].
[26] A. A. Chernov, in *Modern Crystallography*, Vol. 3, Nauka, Moscow (1980), p. 7 [English transl.: Springer, Berlin (1984)].
[27] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* [in Russian], Nauka, Moscow (1979) [English transl.: Pergamon Press, Oxford (1981)].
[28] A. V. Osipov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1213 (1994) [Phys. Solid State **36**, 664 (1994)].
[29] A. V. Osipov, J. Phys. D **28**, 1670 (1995).
[30] O. M. Todes, Zh. Fiz. Khim. **20**, 630 (1946).
[31] I. M. Lifshits and V. V. Slezov, Zh. Éksp. Teor. Fiz. **35**, 479 (1958) [Sov. Phys. JETP **8**, 331 (1959)].
[32] I. M. Lifshits and V. V. Slezov, Fiz. Tverd. Tela (Leningrad) **1**, 1401 (1959).
[33] C. Z. Wagner, Elektrochem. **65**, 581 (1961).
[34] M. Kahlweit, Appl. Radiat. Isot. **5**, 1 (1975).
[35] K. Binder, Phys. Rev. B **15**, 4425 (1977).
[36] J. A. Marquese and J. J. Ross, Chem. Phys. **79**, 373 (1983).
[37] M. Tokuyama and K. Kawasaki, Physica A **123**, 386 (1984).
[38] G. Venzl, Phys. Rev. A **31**, 3431 (1985).
[39] P. W. Voorhees, J. Stat. Phys. **38**, 231 (1985).
[40] M. Marder, Phys. Rev. A **36**, 858 (1987).
[41] A. Nakahara, T. Kawakatsu, and K. Kawasaki, J. Chem. Phys. **95**, 4407 (1991).
[42] S. A. Kukushkin, Acta Metall. Mater. **43**, 1484 (1994).
[43] V. V. Slezov, Phys. Rev. **17**, 1 (1995).
[44] A. H. Nayfeh, *Perturbation Methods*, Wiley, New York (1973).
[45] V. V. Slezov, J. Phys. Chem. Solids **39**, 367 (1978).
[46] F. M. Kuni and A. P. Grinin, Kolloidn. Zh. **46**, 460 (1984).
[47] R. E. Langer, Trans. Am. Math. Soc. **67**, 461 (1949).
[48] I. Imai, Phys. Rev. **80**, 1112 (1950).
[49] H. J. Moriguchi, J. Phys. Soc. Jpn. **14**, 1771 (1959).

Translated by Eugene Yankovsky

Magnetic ordering and phase transitions in planar antiferromagnetic systems with a Kagomé lattice

R. S. Gekht*) and I. N. Bondarenko

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 28 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 2209–2220 (June 1998)

We study the process of magnetic ordering in planar antiferromagnetic systems with a Kagomé lattice. It is found that if the interaction between next-nearest-neighbor spins is taken into account, the heat capacity of such systems has a singularity at a finite temperature T. On the basis of a scaling analysis of finite-size systems we study the behavior of thermodynamic quantities in the neighborhood of a phase transition. We find that the phase transition at the critical point is due to discrete- and continuous-symmetry breaking, in which the long-range chiral order and the power-law translational spin order emerge simultaneously. Finally, we calculate the temperatures of the transition to different (with three and nine spins per unit cell) ordered states. © *1998 American Institute of Physics.* [S1063-7761(98)02106-4]

1. INTRODUCTION

Lately there has been an upsurge of interest in phase transitions in the low-temperature range in compounds with a Kagomé lattice. Because of the special geometry of the lattice (triangles in a layer alternate with hexagons), the spin systems are highly frustrated. As the temperature drops, the ordering proceeds much more slowly in comparison to ordinary frustrated systems. It is known[1,2] that this situation occurs because in systems with a coordination number much smaller than, say, in triangular antiferromagnets, at large values of S in addition to states with nontrivial global degeneracy there can be local degenerate states. As a result, when there is interaction between the nearest-neighbor spins, there is no single finite temperature at which a phase transition to a magnetically ordered state can occur. The additional interaction between next-nearest-neighbor spins partially lifts the degeneracy and may lead to a phase transition at finite temperatures.[3] Nevertheless, since the frustration effects are still present, the process of ordering and stabilization of structure is slower than in nonfrustrated systems.

Ising systems with a Kagomé lattice have been studied fairly recently. As in Ising systems with a triangular lattice, in the classical ground state the entropy per spin is finite (interaction of nearest neighbors), but the decrease in the spin–spin correlation functions at $T=0$ follows an exponential law rather than a power law (superfrustrated systems[4,5]). Heisenberg systems with a Kagomé lattice were under intensive study at the beginning of the 1990s. The excitations of such systems have a null spectrum in the entire magnetic Brillouin zone.[6] Quantum[7] and thermal[2,3] fluctuations lift the degeneracy and select states with a planar spin configuration. XY systems have not been studied so thoroughly. It is known that as $T\to 0$, the spins in such systems are less ordered than in Heisenberg systems. Here the correlation function of an XY system is similar to the correlation function of the three-state Potts model[8] $(T\to 0)$, while the correlation length of Heisenberg systems diverges in the zero-temperature limit.[2,8]

In the $MFe_3(OH)_6(SO_4)_2$ family of compounds (M=H_3O, Na, K, Rb, Ag, NH_4, Tl, Pb, and Hg) called jarosites (the name has its origin in mineralogy) and in their chromium analog $KFe_3(OH)_6(CrO_4)_2$, the magnetic iron ions Fe^{3+} form a Kagomé lattice in the c plane.[9–11] The crystalline structure of such compounds is hexagonal (the space group is $R\bar{3}m$). According to the experimental data, the interactions between nearest-neighbor spins inside a layer and between layers is antiferromagnetic.[12] Neutron-diffraction, Mössbauer, and other measurements involving jarosites show that in jarosites at low temperatures magnetic ordering can be accompanied by formation of triangular structures in the x plane.[11–13]

In our work we studied the phase transitions in compounds of the jarosite type. Since in such compounds the neighboring layers with Fe^{3+} are separated by nonmagnetic ions of S, O, K, and OH, layer-to-layer exchange is much smaller than the intraplanar exchange J_1. It has also been established that in some substances, e.g., with M=K, the spins in a layer are perpendicular to the c axis as a result of magnetic ordering.[12] Below we allow for the interaction between nearest-neighbor and next-nearest-neighbor spins on a Kagomé lattice separated by distances Δ_1 and Δ_2, respectively,

$$H=J_1\sum_{i\Delta_1}\mathbf{S}_i\cdot\mathbf{S}_{i+\Delta_1}+J_2\sum_{i\Delta_2}\mathbf{S}_i\cdot\mathbf{S}_{i+\Delta_2}, \tag{1}$$

and limit our study to systems with XY-like spins: $\mathbf{S}_i=S(\cos\theta_i,\sin\theta_i)$.

As for Ising systems with a Kagomé lattice, it is known[14] that phase transitions are possible only when next-nearest-neighbors interact antiferromagnetically $(J_2<0)$, but compounds with Ising spins have yet to be found. In contrast to such compounds, XY systems have continuous symmetry in

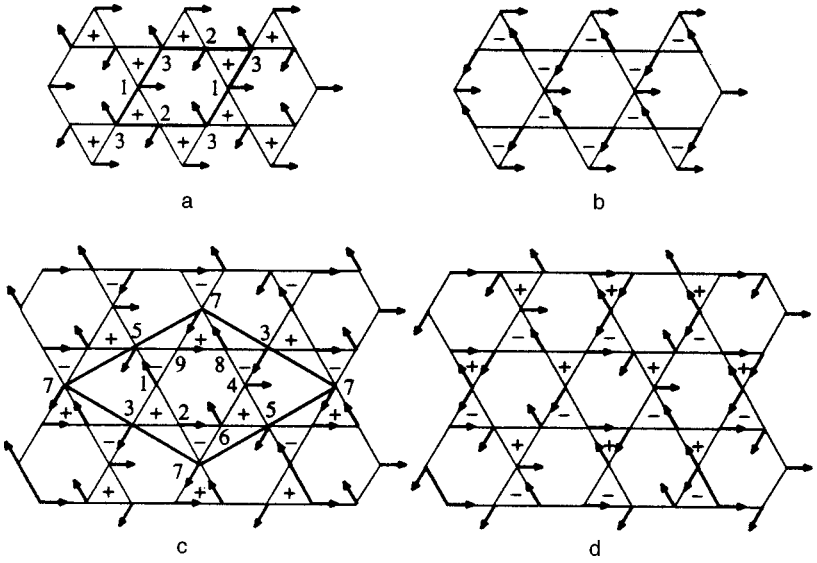


FIG. 1. Degenerate ground states for $j>0$ (a and b) and $j<0$ (c and d); the ''plus'' and ''minus'' indicate the sign of the parameter k on the elementary triangles. The heavy lines depict the unit magnetic cells with three (a) and nine (c) spins.

the plane. Further, in contrast to Heisenberg systems, they also have discrete symmetry, since at $T=0$ the chiral parameter specified for each elementary triangle,[15]

$$\mathbf{k}=\frac{2}{3\sqrt{3}}((\mathbf{S}_1\mathbf{S}_2)+(\mathbf{S}_2\mathbf{S}_3)+(\mathbf{S}_3\mathbf{S}_1)) \tag{2}$$

(the spins are numbered clockwise), takes a value of $+1$ or -1. The situation resembles triangular antiferromagnets with planar spins,[16,17] but here, first, the chiral parameter does not change sign for $J_2>0$ and, second, the unit cell on a Kagomé lattice has nine spins instead of three for $J_2<0$. We find that although for next-nearest-neighbor antiferromagnetic interaction ($J_2>0$) the ordering process is slower than that for ferromagnetic interaction ($J_2<0$), in both cases there exists a finite critical temperature at which translational spin and chiral orders emerge simultaneously.

2. THE LOW-TEMPERATURE RANGE

The ground state on a Kagomé lattice strongly depends on the sign of the exchange interaction J_2 between next-nearest-neighbors. For antiferromagnetic exchange, $J_2>0$, this state has a structure with three spins per unit cell (Fig. 1a), while for $J_2<0$ the structure consists of nine spins (Fig. 1c). In both cases the spin configurations are continuously degenerate with respect to rotations in the plane and are two-fold symmetric. For $J_2>0$ the discrete degeneracy is characterized by a k of fixed sign (Figs. 1a and 1b), while for $J_2<0$ the value of k changes sign in neighboring elementary triangles (Figs. 1c and 1d). A transition between two equivalent states amounts to surmounting an energy barrier proportional to $|J_2|$. We expect that in the low-temperature range the related excitations are suppressed and the system can be described in the harmonic approximation. Let us examine the properties of the phases at low temperatures for states with three and nine spins per unit magnetic cell.

In the state with three spins per unit cell, $J_2>0$, the Hamiltonian in the quadratic approximation in $\psi_k=(\psi_{k1},\psi_{k2},\psi_{k3})$ (here $\psi_{k\alpha}$ are the Fourier transforms of the deviation of the sublattice α from the equilibrium structure) can be written as

$$H=-(J_1+J_2)S^2N+\frac{1}{2}S^2\sum_k \psi_k M_k \psi_{-k}, \tag{3}$$

where the elements of the 3×3 matrix M_k are

$$\begin{aligned}
M_{11}&=M_{22}=M_{33}=2(J_1+J_2),\\
M_{12}&=M_{21}=-J_1\cos\left(\frac{k_x}{2}+\frac{\sqrt{3}}{2}k_y\right)\\
&\quad -J_2\cos\left(\frac{3}{2}k_x-\frac{\sqrt{3}}{2}k_y\right),\\
M_{23}&=M_{32}=-J_1\cos\left(\frac{k_x}{2}-\frac{\sqrt{3}}{2}k_y\right)\\
&\quad -J_2\cos\left(\frac{3}{2}k_x+\frac{\sqrt{3}}{2}k_y\right),\\
M_{31}&=M_{13}=-J_1\cos k_x-J_2\cos\sqrt{3}k_y.
\end{aligned} \tag{4}$$

When k is small, for the smallest eigenvalue of the matrix M_k we obtain

$$\lambda_1=\frac{1}{2}(J_1+3J_2)k^2 \tag{5}$$

($\lambda_2=\lambda_3\simeq3(J_1+J_2)$). In the low-temperature range we have the following expressions for the energy $E=\langle H\rangle$, the spin–spin correlation function, and the chiral parameter $k(T)$:

$$E=-(J_1+J_2)S^2N\left[1-\frac{T}{2(J_1+J_2)S^2}\right], \tag{6}$$

$$\langle\mathbf{S}_o\cdot\mathbf{S}_r\rangle=\exp\left[-\frac{\langle(\psi_o-\psi_r)^2\rangle}{2}\right]\sim r^{-\eta(T)}, \tag{7}$$

where o and r belong to the same sublattice,

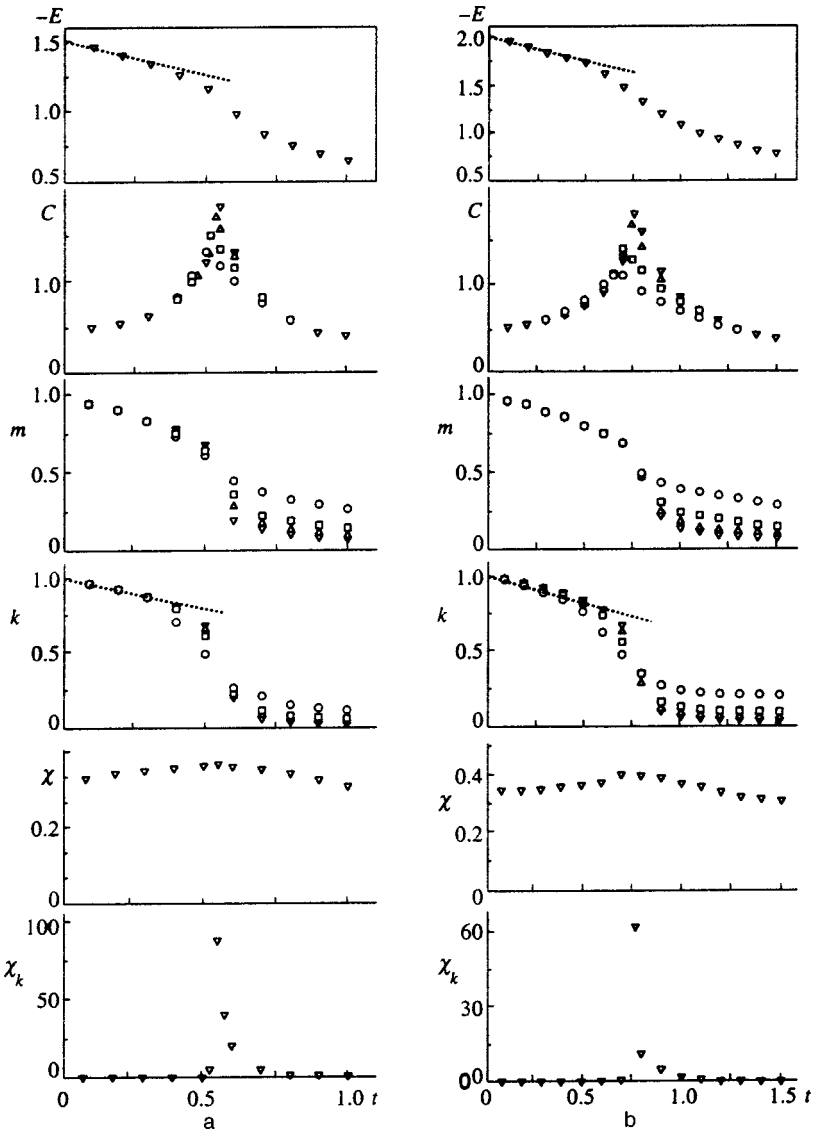


FIG. 2. Energy, heat capacity, magnetization, chiral parameter, and susceptibilities χ and χ_k versus the normalized temperature $t=T/J_1S^2$ at $j=0.5$ (a) and $j=-0.5$ (b). The symbols ○, □, △, and ▽ correspond to $L=12$, 24, 36, and 48.

$$\eta(T)=\frac{T}{\pi(J_1+3J_2)S^2}, \tag{8}$$

$$k(T)=\frac{1}{N}\left\langle \sum_R k(R)\right\rangle = 1-\frac{T}{2(J_1+3J_2)S^2} \tag{9}$$

(R stands for the coordinates of the points of the dual lattice).

In the states with nine spins per unit cell, $J_2<0$, the smallest eigenvalue of M_k, the spin–spin correlation function, and the chiral parameter $k(T)$ in the low-temperature range are given by the same expressions (5)–(9) but with $-2J_2$ substituted for J_2.

The process of ordering of planar spins on a Kagomé lattice was studied for arbitrary T by the Monte Carlo method. In comparison to a triangular lattice, the number of spins on a Kagomé lattice is smaller by 1/4, or $N=3L^2/4$, where L in our calculations varied from 12 to 48. The heat capacity and the magnetic susceptibility were found by numerical calculations from the fluctuations of the energy and magnetization, respectively. We also calculated the mean square of the sublattice magnetization:

$$m^2=\frac{1}{N_\alpha}\left\langle \sum_{N_\alpha} M_\alpha^2\right\rangle \tag{10}$$

($N_\alpha=3$ for $J_2>0$ and $N_\alpha=9$ for $J_2<0$; M_α is the sublattice magnetization), the parameter $k(T)$, and the corresponding susceptibility χ_k.

The temperature dependence of the thermodynamic quantities for $j=\pm 0.5$ ($j=J_2/J_1$) is depicted in Fig. 2. At low temperatures the behavior of the energy can be described by the harmonic approximation (6) for $j=0.5$ and by the same expression with $-2J_2$ substituted for J_2 in (6) for $j=-0.5$. Deviations from the linear dependence emerge for $T/J_1S^2>0.3$ in Fig. 2a and for $T/J_1S^2>0.5$ in Fig. 2b. The parameter $k(T)$ behaves in the linear region in a similar way, in accordance with the expected relationships of type (9).

The exponent $\eta(T)$ for the spin–spin correlation function can be determined from the dimensional relationship

$$m^2\sim L^{-\eta(T)}. \tag{11}$$

We calculated the parameter $\eta(T)$ from the slope of the asymptotic straight lines for $-\ln m^2$ as a function of $\ln L$.

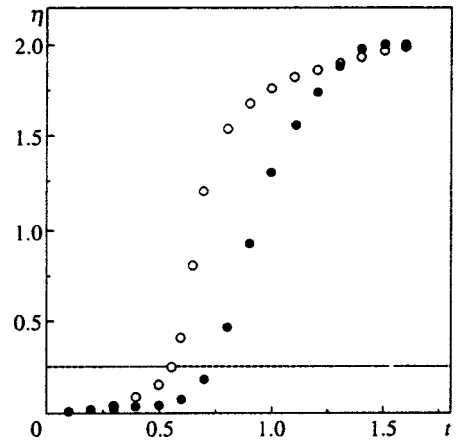


FIG. 3. Temperature dependence of η. The symbols ○ and ● correspond to diagrams with $j=0.5$ and $j=-0.5$, respectively.

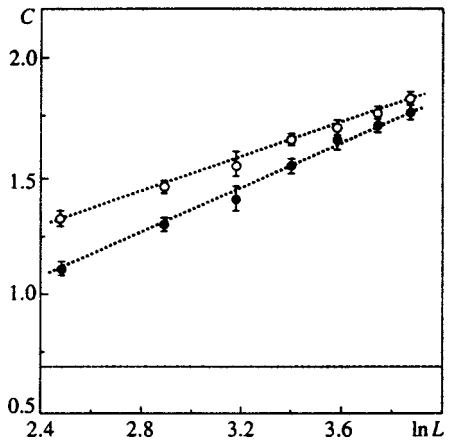


FIG. 4. The maximum in heat capacity as a function of ln L. The symbols ○ and ● correspond to the same values of j as in Fig. 3.

The results for different values of T are depicted in Fig. 3. As the temperature increases, deviations from the linear dependence emerge at the same values of T as for the internal energy.

3. PHASE TRANSITION

The appreciable difference between the antiferromagnetic systems with $J_2=0$ and $J_2\neq 0$ manifests itself in the behavior of the heat capacity and the susceptibilities (Fig. 2). For instance, when we have $J_2\neq 0$, the heat capacity and the chiral susceptibility have a peak that increases with lattice size and becomes sharper, while the homogeneous susceptibility χ has a broad maximum in a specific temperature range. The dimensional dependence of the height of the heat-capacity peak is depicted in Fig. 4: obviously, the logarithmic divergence is due to a phase transition in the parameter k.

We expect that in the limit $N\to\infty$ the behavior of k is described by the following formula:

$$k^2N=[k(N\to\infty)]^2N+O(N). \tag{12}$$

The dimensional dependence of k^2N on N at $j=\pm 0.5$ is depicted in Figs. 5a and 5b. The values of $k(T)$ for an infinite system were calculated from the slope of the asymptotic straight lines (dotted lines). On the basis of these data, we constructed (Figs. 5c and 5d) the dependence of $-\ln k$ on

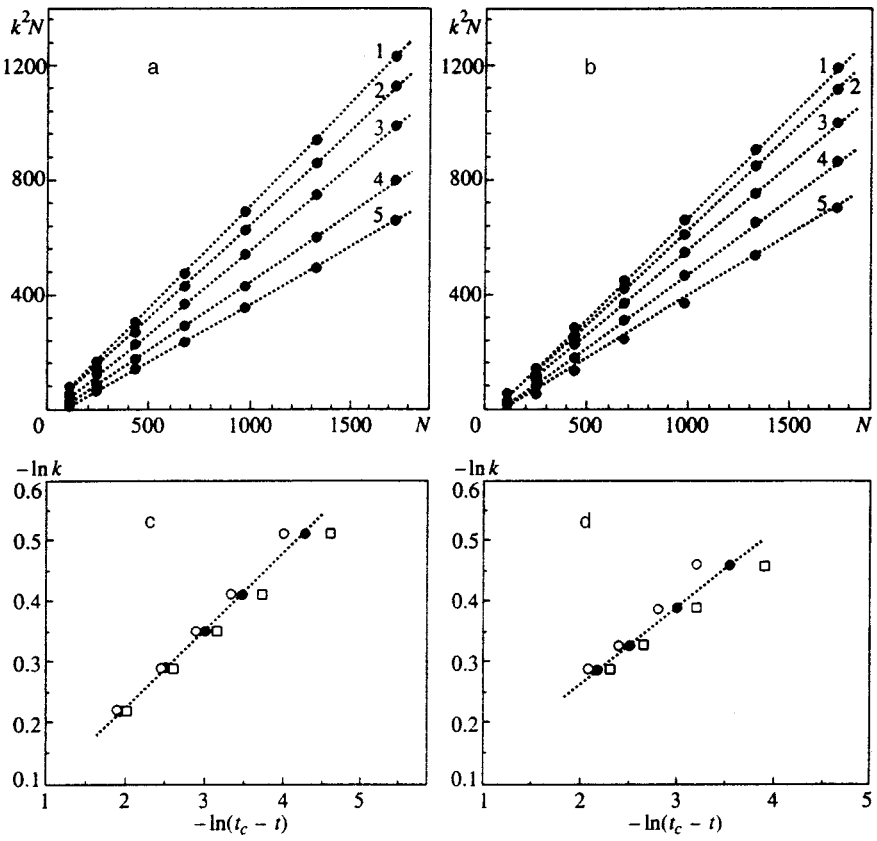


FIG. 5. (a,b)—Dimensional dependence of k^2N at different temperatures. The slopes of the asymptotic straight lines (dotted lines) yield the value of k^2 for an infinite system. The straight lines *1–5* correspond to $t=0.36$, 0.41, 0.46, 0.51, and 0.53 at $j=0.5$ (a) and $t=0.52$, 0.57, 0.62, 0.67, and 0.72 at $j=-0.5$ (b). (c,d)—The parameter k extrapolated to an infinite system as a function of the normalized temperature t (log–log scale) at $j=0.5$ (c) and $j=-0.5$ (d). The symbols ○, ●, and □ correspond to $t_c=0.55$, 0.54, and 0.53 at $j=0.5$ (c) and $t_c=0.74$, 0.73, and 0.72 at $j=-0.5$ (d). The dotted lines have a slope $\beta=0.12\pm 0.01$.

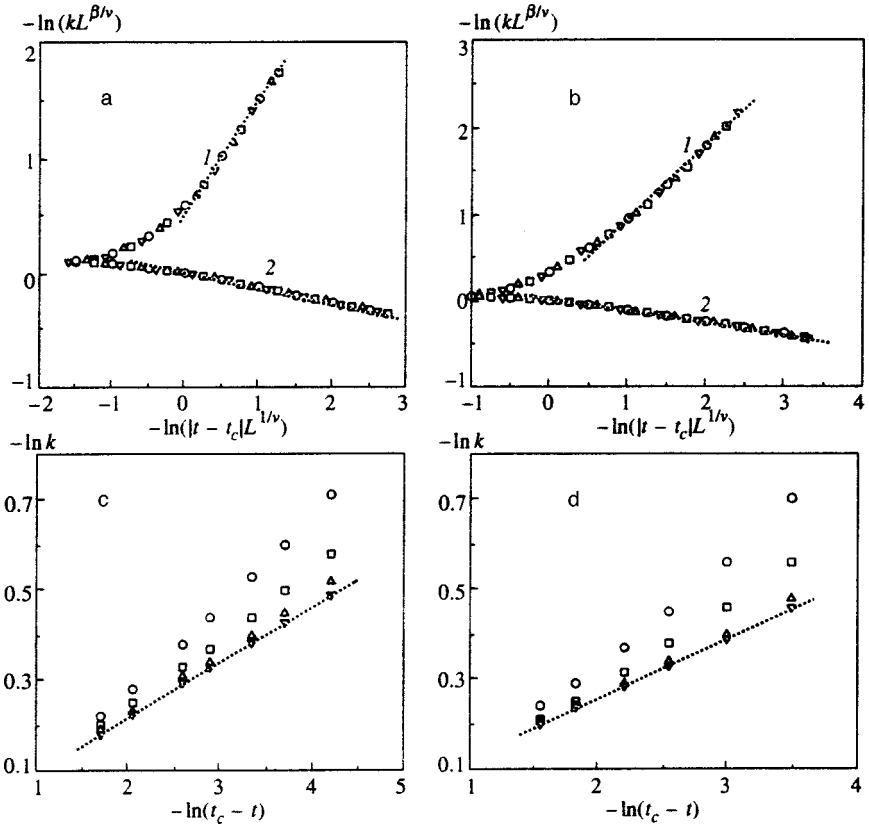


FIG. 6. (a,b)—The scaling functions for the parameter k above and below T_c (curves *1* and *2*, respectively) at $j=0.5$ (a) and $j=-0.5$ (b). The symbols ○, □, △, and ▽ correspond to $L=12$, 24, 36, and 48. The dotted straight lines have a slope $\nu-\beta=7/8$ for $T>T_c$ and $-\beta=-1/8$ for $T<T_c$. (c,d)—The temperature dependence of the parameter k for finite-size systems (log–log scale); $t_c=0.535$ at $j=0.5$ (c) and $t_c=0.726$ at $j=-0.5$ (d). The symbols ○, □, △, and ▽ correspond to $L=12$, 24, 36, and 48. The dotted straight lines have a slope $\beta=1/8$.

$-\ln(t_c-t)$ for different trial values of $t_c(=T_c/J_1S^2)$. The critical temperature t_c is found under the assumption that the chiral parameter is a power function: $k(t)\sim(t-t_c)^{\beta}$. Figures 5c and 5d show that for any sign of j a straight line with a slope $\beta=0.12\pm0.01$ emerges at $t_c=0.54\pm0.01$ for $j=0.5$ and at $t_c=0.73\pm0.01$ for $j=-0.5$.

We also did a finite-size-scaling analysis under the assumption that

$$kL^{\beta/\nu}=F_k(|t-t_c|L^{1/\nu}), \tag{13}$$

where F_k is the scaling function.[18] Below t_c the relationship (13) reduces to $k\sim(t_c-t)^{\beta}$ in the limit $L\to\infty$, so that for F_k we have

$$F_k\sim x^{\beta} \tag{14}$$

as $x\to\infty$. On the other hand, above t_c the parameter k is proportional to $1/\sqrt{N}\sim 1/L$, so that in this case

$$F_k(x)\sim x^{\beta-\nu} \tag{15}$$

as $x\to\infty$. The best values of t_c, β, and ν, obtained from the conditions that the data for different lattice sizes lie on a single curve (Figs. 6a and 6b) and the limiting relations (14) and (15) are valid, are as follows: $t_c=0.535$ at $j=0.5$ and $t_c=0.726$ at $j=-0.5$, and $\beta=1/8$ and $\nu=1$ irrespective of the sign of j. We see that the calculated values of the transition temperatures and the critical exponents in Figs. 6a and 6b are in good agreement with the similar calculated values in Figs. 5c and 5d.

For these values $t_c=0.535$ ($j=0.5$) and $t_c=0.726$ ($j=-0.5$) we have also found the $-\ln k$ vs. $-\ln(t_c-t)$ dependence for different values of L (Figs. 6c and 6d). Near the transition temperature the data of the numerical calculations deviate from a straight line (the dotted line) because of the finiteness of L. In the region where the data for different lattice sizes lie on a common straight line, the lines correspond to the slope $\beta=1/8$ (as in the previous calculations).

Above T_c, scaling analysis of the chiral susceptibility χ_k was done on the basis of the following relationship:

$$t\chi_kL^{-\gamma/\nu}=F_{\chi}(|t-t_c|L^{1/\nu}). \tag{16}$$

Obviously, as $x\to\infty$, the scaling function $F_{\chi}(x)$ assumes the following form:

$$F_{\chi}(x)\sim x^{-\gamma}\ (t>t_c), \tag{17}$$

since in the thermodynamic limit $L\to\infty$ we must have $t\chi_k\sim|t-t_c|^{-\gamma}$. The values of γ and ν were chosen from the conditions that the numerical data for lattices with different Ls lie on the same curve and that the limit (17) holds. The best coincidence at $t_c=0.535$ for the case $j=0.5$ and at $t_c=0.726$ for the case $j=-0.5$ was obtained with $\nu=1$ and $\gamma=7/4$ (Figs. 7a and 7b).

Thus, the foregoing results show that, irrespective of the sign of j (and hence of the number of spins per unit cell), the critical behavior in a phase transition is described by the critical exponents of two-dimensional Ising systems. This fact is not accidental and is due to the symmetry of the systems with respect to sign reversal of k.

In determining the temperature of the Berezinskii–Kosterlitz–Thouless transition it is convenient to study the correlation function

$$g(r)=\langle\cos 3(\psi_0-\psi_r)\rangle\sim r^{-9\eta_{xy}(T)}, \tag{18}$$

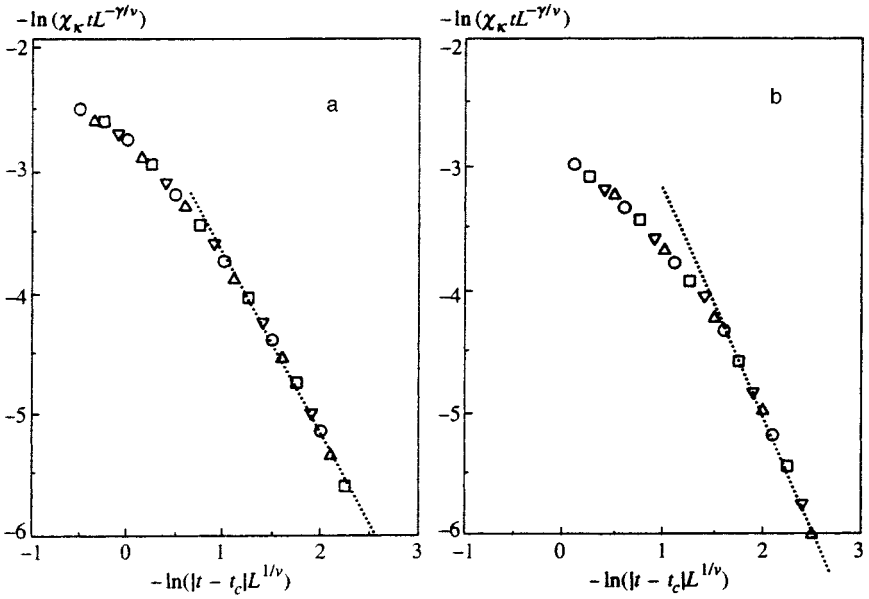


FIG. 7. The scaling functions for the chiral susceptibility above T_c at $j=0.5$ (a) and $j=-0.5$ (b). The symbols ○, □, △, and ▽ correspond to $L=12$, 24, 36, and 48. The dotted straight lines have a slope $-\gamma=-7/4$.

which makes it possible to isolate the contribution of continuous fluctuations at T below the temperature of the Ising transition and to correctly determine the phase transition if it occurs at temperature higher than that for the transition in discrete variables. Figure 8 depicts the power-function behavior of $g(r)$ for $j=\pm 0.5$ at different temperatures. Using the Berezinskii–Kosterlitz–Thouless criterion $\eta_{xy}(T_{\mathrm{BKT}})=1/4$, we found that the phase transition with continuous-symmetry breaking occurs at $t_{\mathrm{BKT}}=0.542\pm 0.003$ at $j=0.5$ and $t_{\mathrm{BKT}}=0.733\pm 0.003$ at $j=-0.5$, where $t_{\mathrm{BKT}}=T_{\mathrm{BKT}}/JS^2$. Within the accuracy of the calculations, t_{BKT} coincides with t_c, so that a phase transition in the system is realized at a single temperature, irrespective of the sign of $j(=\pm 0.5)$. Note that, to the accuracy of calculations, the behavior of η in (8) yields the same value of t_{BKT}. In this case for $\eta=1/4$ we have $t_{\mathrm{BKT}}=0.537\pm 0.002$ at $j=0.5$ and $t_{\mathrm{BKT}}=0.729\pm 0.003$ at $j=-0.5$. Similar calculations for other values of j that are not too close to zero show that both transitions occur simultaneously. The t_c-j phase diagram is depicted in Fig. 9. The neighborhood of the point $j=0$ where the two phase transitions may be expected to occur is probably very small and requires more exact calculations and extensive computer time.

In $KFe_3(OH)_6(SO_4)$, the magnetic susceptibility has a broad maximum at $T_c=60$ K (Ref. 10); the exchange interactions J_1 and J_2 are antiferromagnetic, with J_2 known to be smaller than J_1 by a factor of ten. At $j=0.1$ we have $t_c=0.22$. Thus, the exchange interaction between the nearest-neighbor Fe^{3+} ions with spins $S=5/2$ can be expected to be 44 K.

4. CONCLUSION

We have studied the magnetic properties of planar antiferromagnetic systems with a Kagomé lattice. We have found that with allowance for exchange interactions between next-nearest-neighbor spins there is a phase transition in the system at finite temperatures. In the low-temperature phase there is long-range order in the parameter k, and the correlation function decreases according to a power law. Scaling analysis of finite-size systems shows that k vanishes at the same temperature at which the chiral susceptibility diverges, and their behavior is described fairly well by the critical exponents of two-dimensional Ising systems. We have also found that the temperature of a Berezinskii–Kosterlitz–Thouless transition and the temperature of an Ising transition

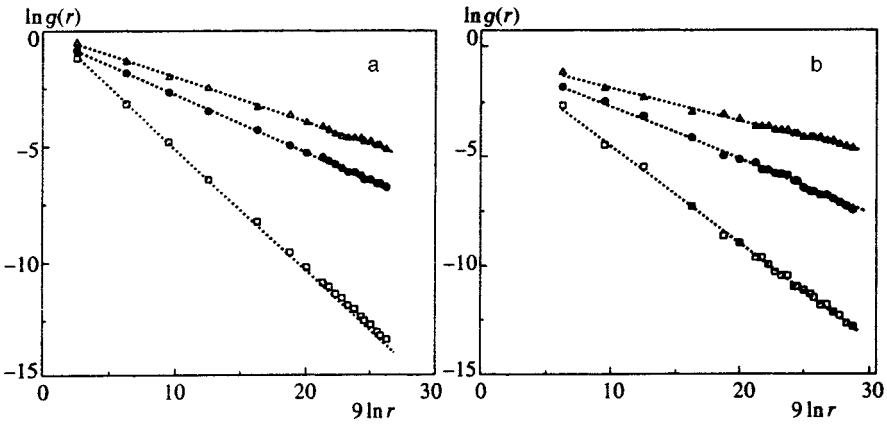


FIG. 8. Spatial dependence of the correlation function $g(r)$ for $L=48$. The symbols △, ●, and □ correspond to $t=0.519$, 0.542, and 0.565 and the slope of the dotted lines $\eta_{xy}=0.18$, 0.25, and 0.5 (a), and $t=0.664$, 0.733, and 0.804 and $\eta_{xy}=0.12$, 0.25, and 0.45 (b).

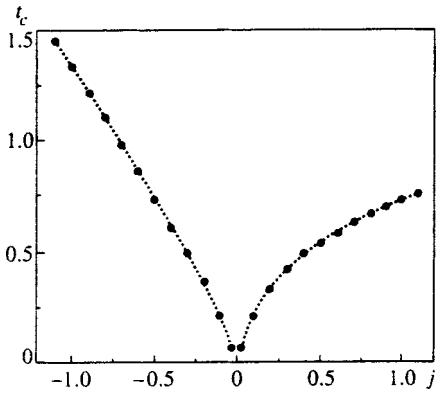


FIG. 9. The phase diagram in the $t_c - j$ plane for planar antiferromagnetic systems with a Kagomé lattice.

coincide to within the accuracy of the calculations. We expect that our result can be used for a detailed experimental study of jarosite-type compounds. Note that in real systems with weak interplanar interaction there is a narrow but finite temperature range where the critical behavior is three-dimensional. However, the vast body of experimental data suggests that, say, for the layered XY ferromagnet Rb_2CrCl_4 (Ref. 19), the Ising antiferromagnet K_2CoF_4 (Ref. 20), the triangular antiferromagnet VCl_2 (Ref. 21), and other magnetic materials (Ref. 22) the behavior outside this range is two-dimensional, although there is three-dimension long-range order in the system.

In conclusion we note that in Ising-like Heisenberg antiferromagnets, where due to distortions in the 120° structure there is a finite magnetic moment on each elementary triangle of the Kagomé lattice, a drop in temperature can lead to a phase transition with discrete- and continuous-symmetry breaking.[23] Therefore, we expect that the behavior of such systems is in many respects similar to the behavior of the planar (XY) antiferromagnetic systems considered in this paper.

This work was supported by the Krasnoyarsk Regional Scientific Fund (Project 6F0061).

*)E-mail: theor@iph.krasnoyarsk.su

[1] P. Chandra, P. Coleman, and I. Ritchey, J. Phys. (Paris) **33**, 591 (1993).
[2] J. T. Chalker, P. C. W. Holdsworth, and E. F. Shender, Phys. Rev. Lett. **68**, 855 (1992).
[3] A. B. Harris, C. Kallin, and A. J. Berlinsky, Phys. Rev. B **45**, 2899 (1992).
[4] A. Sütö, Z. Phys. B **44**, 121 (1981).
[5] R. S. Gekht and V. I. Ponomarev, Phase Transit. **20**, 27 (1990).
[6] C. Zeng and V. Elser, Phys. Rev. B **42**, 8436 (1990).
[7] A. Chubukov, Phys. Rev. Lett. **69**, 832 (1992).
[8] D. A. Huse and A. D. Rutenberg, Phys. Rev. B **45**, 7536 (1992).
[9] R. Wang, W. F. Bradley, and H. Steinfink, Acta Crystallogr. **18**, 249 (1965).
[10] A. Bonnin and A. Lecerf, C. R. Acad. Sci. **262**, 1782 (1966).
[11] M. G. Townsend, G. Longworth, and E. Roudaut, Phys. Rev. B **33**, 4919 (1986).
[12] M. Takano, T. Shinjo, and T. Takada, J. Phys. Soc. Jpn. **30**, 1049 (1971).
[13] A. Keren, K. Kojima, L. P. Le, G. M. Luke, W. D. Wu, Y. J. Uemura, M. Takano, H. Dabkowska, and M. J. P. Gingras, Phys. Rev. B **53**, 6451 (1996).
[14] T. Takagi and M. Mekata, J. Phys. Soc. Jpn. **62**, 3943 (1993).
[15] J. Villain, J. Phys. (France) **38**, 385 (1977).
[16] S. Miyashita and H. Shiba, J. Phys. Soc. Jpn. **53**, 1145 (1984).
[17] D. H. Lee, J. D. Joannopoulos, J. W. Negele, and D. P. Landau, Phys. Rev. B **33**, 450 (1986).
[18] *Finite Size Scaling and Numerical Simulation of Statistical Systems*, V. Privman (Ed.), World Scientific, Singapore (1990).
[19] S. T. Bramwell, P. C. W. Holdsworth, and M. T. Hutchings, J. Phys. Soc. Jpn. **64**, 3066 (1995).
[20] H. Ikeda and K. Hirakawa, Solid State Commun. **14**, 529 (1974).
[21] H. Kadowaki, K. Ubukoshi, K. Hirakawa *et al.*, J. Phys. Soc. Jpn. **56**, 4027 (1987).
[22] E. J. Samuelsen, Phys. Rev. Lett. **31**, 936 (1973).
[23] A. Kuroda and S. Miyashita, J. Phys. Soc. Jpn. **64**, 4509 (1995).

Translated by Eugene Yankovsky

Temperature dependence of electric resistance and magnetoresistance of pressed nanocomposites of multilayer nanotubes with the structure of nested cones

V. I. Tsebro,*) and O. E. Omel'yanovskiĭ,

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924 Moscow, Russia

E. F. Kukovitskiĭ and N. A. Sainov

Kazan Physicotechnical Institute, 420029 Kazan, Russia

N. A. Kiselev and D. N. Zakharov

A. V. Shubnikov Institute of Crystallography, Russian Academy of Sciences, 177333 Moscow, Russia
(Submitted 16 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 2221–2228 (June 1998)

Bulk samples of carbon multilayer nanotubes with the structure of nested cones (fishbone structure) suitable for transport measurements, were prepared by compressing under high pressure ($\sim$25 kbar) a nanotube precursor synthesized through thermal decomposition of polyethylene catalyzed by nickel. The structure of the initial nanotube material was studied using high-resolution transmission electron microscopy. In the low-temperature range (4.2–100 K) the electric resistance of the samples changes according to the law $\ln R \propto (T_0/T)^{1/3}$, where $T_0 \sim 7$ K. The measured magnetoresistance is quadratic in the magnetic field and linear in the reciprocal temperature. The measurements have been interpreted in terms of two-dimensional variable-range hopping conductivity. It is suggested that the space between the inside and outside walls of nanotubes acts as a two-dimensional conducting medium. Estimates suggest a high value of the density of electron states at the Fermi level of about 5×10^{21} eV^{-1}cm^{-3}. © *1998 American Institute of Physics.* [S1063-7761(98)02206-9]

Investigations of electric transport properties of carbon nanotubes has attracted great attention recently. According to theoretical concepts,[1] an isolated nanotube can be either a metal, or semimetal, or insulator, depending on such structural parameters as its diameter, chirality, and the number of concentric layers in it. Despite enormous difficulties in measurements of electric parameters of isolated nanotubes or nanotube bundles, several attempts undertaken recently have been successful.[2–4] The latest published measurements[4] clearly indicate the presence of both metallic and insulating nanotubes in a single set of samples prepared in the same conditions. The authors emphasized that each multilayer nanotube manifested its specific conducting properties, thus indicating a strong correlation between structural and electric parameters.

In this connection, it is interesting to study, in addition to the transport properties of isolated carbon nanotubes, the conducting properties of bulk nanotube materials, in which contacts between nanotubes and/or their sections are randomly distributed. In our previous publication[5] we reported on the conductivity temperature dependence and structure (see also Ref. 6) of carbon nanotube films fabricated by evaporating graphite in an electron beam. The data of those experiments were interpreted in terms of a three-dimensional model of hopping conductivity with a Coulomb gap about the Fermi level (the resistivity was described by the law $\ln \rho \propto (T_0/T)^{1/2}$). The density of states at the Fermi level for films that contained, as shown by structural investigations, mostly one-layer carbon nanotubes (isolated or assembled in bundles) and had a relatively high conductivity was estimated to be $g(\mu) \sim 10^{21}$ eV^{-1}cm^{-3}. On the other hand, films containing multilayer carbon nanotubes were characterized by fairly large values of resistivity, which changed with temperature to Mott's law, $\ln \rho \propto (T_0/T)^{1/4}$. In this case, estimates of the density of states, $g(\mu) \sim 10^{18}$ eV^{-1}cm^{-3}, corresponded to $g(\mu)$ for amorphous carbon. Amorphous carbon in significant quantities was detected on the outside surfaces of multilayer nanotubes in such films by electron microscopy,[5,6] and it seems that the conductivity of such films can be attributed to the presence of carbon.

It is well known that, in addition to one-layer and multilayer nanotubes with walls made of coaxial carbon layers, there are nanotubes whose walls consist of nested truncated cones (these are the so-called fishbone-type structures).[7] Such nanocones are usually detected at the ends of carbon nanotubes, but can also exist in the form of independent objects among products of arc discharges in a helium atmosphere,[8] commonly used in synthesizing carbon nanotubes.

In our recent work[9–11] we demonstrated that thermal decomposition of polyethylene with nickel used as a catalyst is a fairly efficient technique for fabrication of large quantities of fishbone nanotubes. This technique allows one to manufacture in a relatively short time considerable quantities (several grams) of fairly homogeneous nanotube material. According to the data of thermal analysis in oxiding atmo-

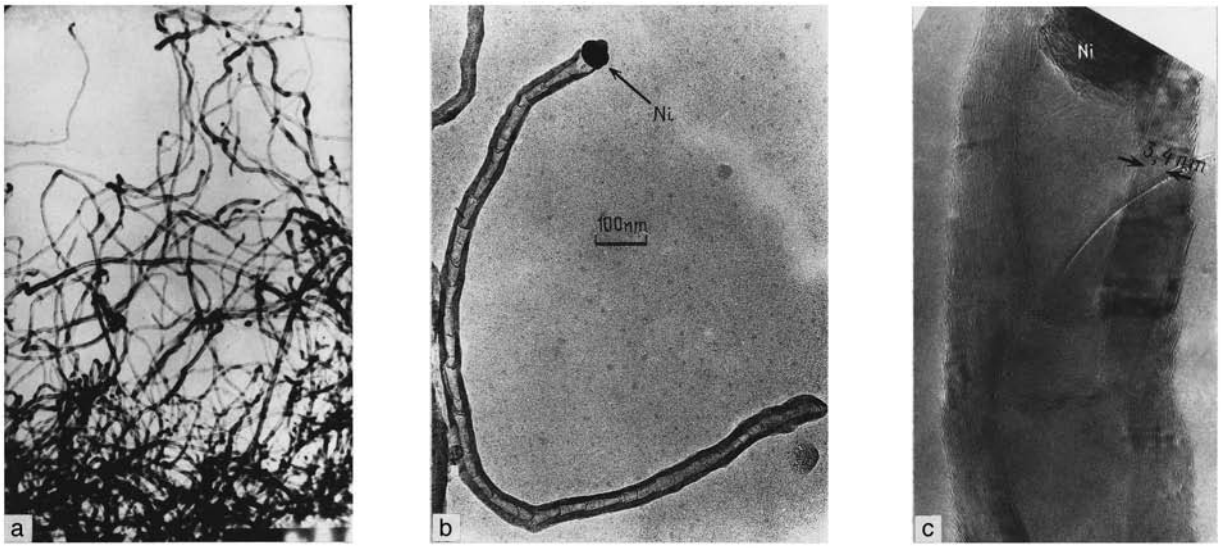


FIG. 1. Electron micrographs of nanotubes in the composite material at (a) low, (b) intermediate, and (c) high resolution.

sphere, the nickel content in this material is less then 15% by mass. Nickel is present in the material in the form of nanoparticles, which can be eliminated completely by thermal processing of the nanocomposite in vacuum at temperatures of up to 2800°C.[10]

In this paper we present our measurements of electric resistance versus temperature and magnetoresistance of bulk nanocomposite samples fabricated by pressing the initial powder of carbon fishbone nanotubes. The structure of the carbon phase in the initial powder was imaged by a Philips EM 430ST transmission electron microscope of high resolution at an accelerating voltage of 200 kV. These measurements demonstrated that the major part of the initial carbon material was multilayer carbon nanotubes with lengths of several micrometers, outside diameter of 40–50 nm, and internal channel diameter of 9–20 nm. The tubes consisted of almost rectilinear sections with lengths of 100–300 nm turned with respect to one another. Figure 1 shows as an example electron micrographs of the composite nanotube material at (a) low, (b) medium, and (c) high resolution.

The analysis of micrographs indicated that the nanotube walls were composed in most cases of 40–65 tapered graphite layers. The taper angle varied along the tubes in the range of 16–35°. The inside diameter was also variable. The dimensions and shapes of wider sections of the inside channel corresponded to those of catalytic nickel nanoparticles, which were detected in most cases at the ends of the nanotubes. We observed either so-called bamboo structures (with taper angles of 20 to 25°) or, more frequently, fishbone structures with larger taper angles.

Bulk samples that could be used in transport measurements were fabricated by cold pressing of nanotube powder under high ($\sim$25 kbar) pressure. Samples were shaped as bars with dimensions of $\sim 1\times 2\times 3$ mm. Contacts for measuring current and voltage across samples were made from a conducting epoxy paste. Note that the samples were fairly strong and their resistivity at room temperature was relatively low: $\rho(300\ \mathrm{K})\sim 1\ \Omega$cm. The resistance was measured as a function of temperature down to the liquid-helium temperature in magnetic fields of up to 75 kOe.

In all samples under investigation, the resistance changed with temperature most rapidly (about one order of magnitude) in the temperature range between liquid helium and $\sim$100 K, and the resistance followed the law

$$R(T)=R_0\ \exp[(T_0/T)^{1/3}], \tag{1}$$

which is typical of variable-range hopping conductivity in two dimensions. Figure 2 shows as an example two curves of ln R vs. $T^{-1/3}$ plotted for samples Nos. 14 and 15.

It is known[12] that in this case T_0 in Eq. (1) is given by

$$T_0=\frac{13.8}{k_B g^*(\mu)a^2}, \tag{2}$$

where $g^*(\mu)$ is the two-dimensional density of states at the Fermi level and a is the localization length.

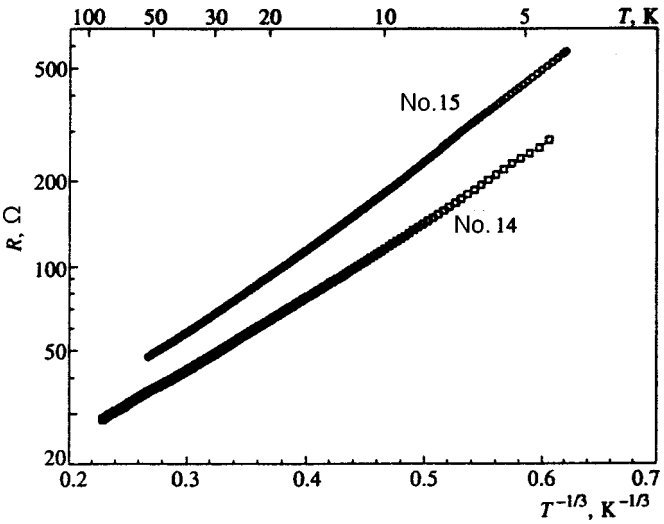


FIG. 2. Logarithmic resistance of samples Nos. 14 and 15 as a function of $T^{-1/3}$.

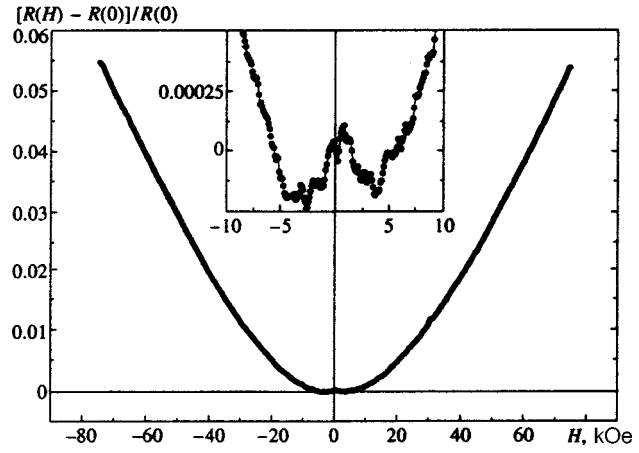


FIG. 3. Magnetoresistance of sample No. 15 versus magnetic field at $T=4.2$ K. The inset shows the section of negative magnetoresistance at low magnetic fields on an extended scale.

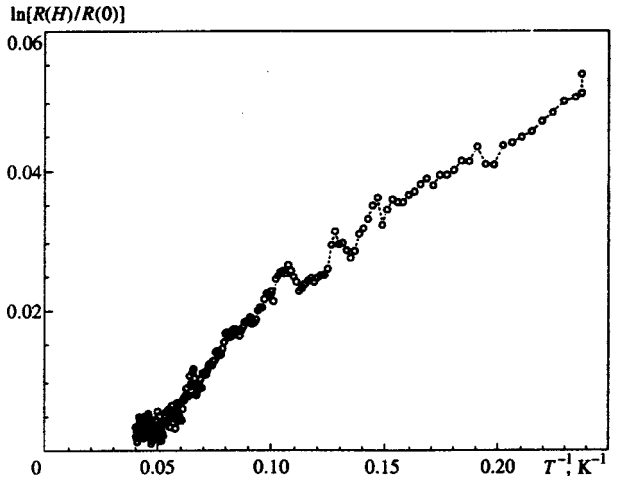


FIG. 4. Magnetoresistance of sample No. 15 versus temperature in a magnetic field of 75 kOe plotted in coordinates $\ln[R(H)/R(0)]$ and T^{-1}.

To the best of our knowledge, this is the first observation of the dependence $\ln R \propto (T_0/T)^{1/3}$ in a system with a relatively low resistivity. Another interesting feature of our measurements is low T_0 (for example, we found $T_0=7.3$ K in sample No. 15, and in all tested samples T_0 was within the interval of 6.5–7.5 K), which directly indicates, in accordance with Eq. (2), that the density of states at the Fermi level is high.

In this connection, it is of interest to measure the magnetoresistance, especially as a function of temperature, since these measurements would allow us to estimate directly the localization length a and then derive the two-dimensional density of states $g^*(\mu)$ using Eq. (2).

It is known[12] that in systems with variable-range hopping conductivity, the magnetoresistance is positive and (in moderate magnetic fields) is given by the expression

$$\ln \frac{\rho(H)}{\rho(0)} = t\left(\frac{a}{\lambda}\right)^4 \left(\frac{T_0}{T}\right)^{3/p} \equiv A(T)H^2, \tag{3}$$

where λ is the magnetic length, t is a dimensionless factor of about 0.0025, and $p=D+1$ (where D is the system dimensionality). Since $p=3$ holds in the case under consideration, it follows from Eq. (3) that the magnetoresistance at a fixed magnetic field should be inversely proportional to the temperature.

An example of magnetoresistance measurements versus magnetic field at $T=4.2$ K for sample No. 15 is given in Fig. 3. One can see that the magnetoresistance is adequately described by a quadratic function of H in the range of moderate magnetic fields, $H<30$ kOe, and in higher magnetic fields it tends to a linear function.

The behavior of magnetoresistance in low magnetic fields is especially interesting. As a rule, the magnetoresistance is negative on the section of the curve around zero and becomes positive in fields higher than 7 kOe. As a result, we have a small, broad region of negative resistivity at about 3–4 kOe. Moreover, several additional narrow local minima (see the inset to Fig. 3) are observed superposed on this broad peak. Note that the peaks in the inset to Fig. 3 are not caused by noise, although their amplitudes are very small. Experiments with repeated accumulation and averaging of the signal dedicated to testing the reproducibility of such measurements were performed (the results obtained by this procedure are the ones plotted in the inset to Fig. 3), and these experiments proved that the curves were reproducible, even after warming the samples to the room temperature. It seems that the negative magnetoresistance of the samples and local minima are due to the discrete structure of the conducting network formed by nanotubes. The broadest minimum in the magnetoresistance at 3–4 kOe is tentatively related to the average cell dimension in the network, and local minima are ascribed to some additional characteristic dimensions in the random network. When the applied magnetic field reaches a value such that the magnetic flux through a network cell equals the magnetic flux quantum hc/e, the amplitude of the tunneling between nanotubes increases, which causes a drop in the total resistance of the system. A simple estimate yields a cell dimension of the conducting network of about 120 nm at $H_{\min}\sim 3.5$ kOe, which seems plausible, given the structure of the nanotube material shown by the electronic microscope.

The magnetoresistance of sample No. 15 as a function of temperature under a magnetic field of 75 kOe is plotted in Fig. 4 in terms of $\ln[R(H)/R(0)]$ and T^{-1}. It is clear that the magnetoresistance at low temperatures is reasonably well described by a linear function of T^{-1}, in accordance with Eq. (3). The localization length derived from these measurements is $a=17$ nm. Thus, the two-dimensional density of states at the Fermi level estimated using these data and Eq. (2) is $g^*(\mu)\sim 7.5\times 10^{15}$ eV^{-1}cm^{-2}.

Assuming that the space between the inside and outside walls of nanotubes acts as a two-dimensional medium, we can estimate the three-dimensional density of states $g(\mu)$ at the Fermi level. Using the relation $g^*(\mu)=g(\mu)d$, where d is the average nanotube wall thickness (in this specific case it is about 15 nm), we have $g(\mu)\sim 5\times 10^{21}$ eV^{-1}cm^{-3}.

It seems also interesting to estimate the two-dimensional and three-dimensional densities, n_S and n_V, of current carriers. This can be done using the equation

$$n_S = 2g^*(\mu)\epsilon_0(T), \tag{4}$$

where $\epsilon_0(T)$ is the energy band near the Fermi level containing current carriers contributing to the hopping conductivity.[12] In the two-dimensional case, this band width is given by the equation

$$\epsilon_0(T) = (k_B T)^{2/3}/[g^*(\mu)a^2]^{-1/3}. \tag{5}$$

At $T = 25$ K we find from Eqs. (4) and (5) $n_S \sim 9 \times 10^{12}$ cm^{-2}, hence $n_V \sim 6 \times 10^{19}$ cm^{-2}.

Thus, we have interpreted the low-temperature transport measurements of pressed samples of randomly distributed carbon nanotubes with a nested-cones structure in terms of the two-dimensional variable-range hopping conductivity. We have assumed that the space between the inside and outside walls on nanotubes acts as a two-dimensional medium. In our previous publication[5] the low-temperature properties of carbon nanotubes were interpreted in terms of the three-dimensional model of hopping conductivity with a Coulomb gap in the density of states near the Fermi level. In both cases, the resistance is described at low temperatures by the law $\ln \rho \propto (T_0/T)^{1/n}$ with small T_0, which implies that these carbon nanotube materials, with their various morphologies, are characterized by very high densities of electron states at the Fermi level of $\sim 10^{21}$ eV^{-1}cm^{-3}, which is a value typical of metals. This result is important for understanding the fundamental electronic properties of carbon nanotubes and related materials and may also prove quite useful from the viewpoint of practical applications.

The work was supported by the Russian Scientific Technological Program *Topical Issues in Physics of Condensed Media*, branch *Fullerenes and Atomic Clusters* (project No. 96147) and International Center for Science and Technology (project No. 079).

*)E-mail: tsebro@sci.lebedev.ru

[1] M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes*, Academic Press (1996), p. 814.
[2] L. Langer, V. Bayot, E. Grivei *et al.*, Phys. Rev. Lett. **76**, 479 (1996).
[3] H. Dai, E. W. Wong, and C. M. Lieber, Science **272**, 523 (1996).
[4] T. W. Ebbesen, H. J. Lezec, H. Hiura *et al.*, Nature (London) **382**, 54 (1996).
[5] O. E. Omel'yanovskii, V. I. Tsebro, O. I. Lebedev *et al.*, JETP Lett. **62**, 503 (1995).
[6] N. A. Kiselev, O. I. Lebedev, A. N. Kiselev *et al.*, Inst. Phys. Conf. Ser. **146**, 65 (1995).
[7] Y. Saito and T. Yoshikawa, J. Cryst. Growth **134**, 154 (1993).
[8] M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes*, Academic Press (1996), p. 777.
[9] E. F. Kukovitskii, L. A. Chernozatonskii, S. G. L'vov, and N. N. Melnik, Chem. Phys. Lett. **266**, 323 (1997).
[10] N. A. Kiselev, D. N. Zakharov, J. Sloan *et al.*, Mol. Mat. **10**(1–4) (1998).
[11] E. F. Kukovitskii, S. G. L'vov, and N. A. Sainov, Mol. Mat. **10**(1–4) (1998).
[12] B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors* Springer, New York (1984) [Russ. orig. Nauka, Moscow (1979)].

Translation provided by the Russian Editorial office.

Local geometry of the Fermi surface and the skin effect in layered conductors

N. A. Zimbovskaya*)

Ural State Mining and Geology Academy, 620144 Ekaterinburg, Russia
(Submitted 6 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2229–2243 (June 1998)

This paper studies theoretically how local anomalies in the Gaussian curvature of the Fermi surface of a layered conductor affect its surface impedance at high frequency under conditions where the surface of the conductor is parallel to conducting planes. It is shown that the fact that the curvature vanishes or has singularities in the effective sections of the Fermi surface can substantially change the value and frequency dependence of the impedance. A theory of cyclotron resonance in a magnetic field normal to the conductor's surface is constructed and the features of this phenomenon reflecting the presence of local anomalies in the curvature of the Fermi surface are studied. © *1998 American Institute of Physics.* [S1063-7761(98)02306-3]

1. INTRODUCTION

Most superconducting materials with large critical parameters created in the last two decades are layered structures with metallic-type conductivity. A characteristic feature of these materials is the strong anisotropy of the conductivity in the nonsuperconducting state: the conductivity in the layer plane is much higher than that in the direction normal to the layers. It is common to assume that anisotropy in electrical conductivity is a manifestation of the quasi-two-dimensional nature of the energy spectrum of the charge carriers in layered conductors. The Fermi surface of such conductors is a system of weakly rippled cylinders (isolated or connected by links) whose axes are perpendicular to the layers. The experimental data (see, e.g., Refs. 1–6) support this assumption. However, the study of Fermi surfaces of layered conductors is far from completion. Many aspects of the profiles of periodically pinched cylinders have yet to be investigated. The local geometric characteristics of the Fermi surface strongly affect the high-frequency properties of layered conductors, just as they do in ordinary metals.[7–14] The aim of the present work is to study the effect of the local geometry of the Fermi surface on the skin effect in layered conductors. The exposure of the features of the skin effect related to the specific geometric characteristics of the Fermi surface should create additional possibilities for reconstructing the Fermi surfaces of such materials from the experimental data.

2. MODEL

The Fermi surface of a conductor with a quasi-two-dimensional energy spectrum can be described by the following equation:

$$E_F=\sum_{n=0}^{\infty} E_n(p_x,p_y)\cos\frac{anp_z}{\hbar}, \tag{1}$$

where $\mathbf{p}$ is the electron quasimomentum, $E_n(p_x,p_y)$ are coefficients with dimensions of energy, p_z is the projection of the quasimomentum on the direction normal to the layers, and a is the distance between the layers. If we ignore the anisotropy of the energy spectrum in the layer plane, instead of (1) we can write the simpler equation

$$E_F=\frac{p_\perp^2}{2m_\perp}+\sum_{n=1}^{\infty} E_n(p_\perp)\cos\frac{anp_z}{\hbar}, \tag{2}$$

where $p_\perp$ is the projection of the quasimomentum on the layer plane, and $m_\perp$ is the effective mass corresponding to the motion of the quasiparticles in that plane. Equation (2) describes an axisymmetric open Fermi surface with the axis directed along a normal to the layers.

The usual approach in theoretical papers devoted to the study of the electron properties of layered conductors is to keep only the first few terms in the sum over n in (2). As a rule, only the first term is taken into account, which corresponds to results obtained in the tight-binding approximation. This paper uses a different approach to describing the electron energy spectrum of the charge carriers in layered conductors, in which the Fermi surface is given by the equation

$$E_F=\frac{p_\perp^2}{2m_\perp}-\eta v_0 p_0 E\left(\frac{p_z}{p_0}\right), \tag{3}$$

where $v_0=(2E_F/m_\perp)^{1/2}$, $p_0=\pi\hbar/a$, $E(p_z/p_0)$ is an even function periodic in its argument p_z/p_0 with a period equal to 2, and η is a dimensionless parameter characterizing the extent to which the Fermi surface is rippled. The quantity $-\eta v_0 p_0 E(p_z/p_0)$ is the sum of the trigonometric series in (2). By selecting the type of this function we can obtain a Fermi surface in the form of pinched cylinders with different profiles. This approach provides broad possibilities in analyzing the effect of the shape of the Fermi surface on the observed characteristics of layered conductors.

Let us assume that the function $E(p_z/p_0)$ in the interval $-p_0\leqslant p_z\leqslant p_0$ is described by the expression

$$E\left(\frac{p_z}{p_0}\right)=\frac{1}{rl}\left[1-\left|\frac{p_z}{p_0}\right|^l\right]^r, \tag{4}$$

where the parameters r and l take values greater than unity. The model specified by (3) and (4) makes it possible to describe a broad class of Fermi surfaces in the form of pinched cylinders with different profiles.

The Gaussian curvature of the Fermi surface (3) and (4) is

$$K(p_z)=m_\perp^2\left(v_z^2+\frac{p_\perp^2}{m_\perp}\frac{\partial v_z}{\partial p_z}\right)(p_\perp^2+m_\perp^2 v_z^2)^{-2}, \tag{5}$$

where v_z is the projection of the quasiparticle velocity on the symmetry axis of the Fermi surface. At $l=r=2$ the curvature of the Fermi surface in its sections by the planes $p_z=0$ and $p_z=\pm p_0$ takes the following values:

$$K(0)=\frac{\delta S}{S_{\max}}\frac{1}{p_0^2}, \tag{6}$$

$$K(\pm p_0)=-\frac{2\delta S}{S_{\min}}\frac{1}{p_0^2}, \tag{7}$$

where $S_{\max}$ and $S_{\min}$ are the maximum and minimum sectional areas of the Fermi surface: $S_{\max}=S(0)$, $S_{\min}=S(\pm p_0)$, and $\delta S=S_{\max}-S_{\min}=(\pi/2)m_\perp \eta v_0 p_0$. Thus, if the Fermi surface remains a pinched cylinder ($\eta\neq 0$), its curvature at all points of the sections with extremal diameters is finite. Similar results are obtained if the tight-binding approximation is used to describe the electron energy spectrum.

For $r\neq 2$ and $l=2$ the curvature of the Fermi surface near $p_z=0$ remains finite and $K(0)$ is still described by (6). However, the asymptotic behavior of the curvature of the Fermi surface near $p_z=\pm p_0$ is different:

$$K(p_z)=-2(r-1)\frac{\delta S}{S_{\min}}\frac{1}{p_0^2}\left[1-\left(\frac{p_z}{p_0}\right)^2\right]^{r-2}. \tag{8}$$

Thus, for $1<r<2$ the curvature of the Fermi surface has singularities in these sections. For $r>2$ the curvature $K(p_z)$ vanishes at $p_z=\pm p_0$. The corresponding sections of the Fermi surface are lines of parabolic points. The larger the value of the parameter r, the more the Fermi surface near these sections resembles a cylinder.

The anomalies in the curvature of the Fermi surface near $p_0=0$ can be described by the model (3) and (4) with $r=2$ and $l\neq 2$. Here the curvature of the Fermi surface near $p_z=0$ is described by the asymptotic expression

$$K(p_z)=(l-1)\frac{\delta S}{S_{\max}}\frac{1}{p_0^2}\left|\frac{p_z}{p_0}\right|^{l-2}. \tag{9}$$

For $1<l<2$ the curvature of the Fermi surface has a singularity at $p_z=0$; if $l>2$, the Fermi surface near $p_z=0$ resembles a cylinder, and the larger the value of l the closer the resemblance. Finally, if $r\neq 2$ and $l\neq 2$, we have a surface in the form of a pinched cylinder with curvature singularities in all the sections with extremal diameters. The profiles of the Fermi surfaces described by (3) and (4) are depicted schematically in Fig. 1.

Thus, the proposed model makes it possible to analyze the effect of local curvature anomalies of the Fermi surface on the observed characteristics of layered conductors. This makes it preferable to the tight-binding approximation, which is commonly used to conduct specific calculations (see, e.g., Refs. 15–19).

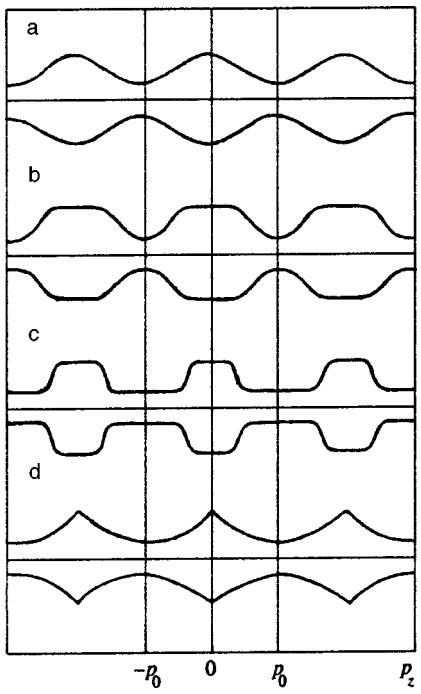


FIG. 1. Profiles of corrugated cylinders described by (3) and (4) for different values of the parameters r and l: (a) $l=r=2$, (b) $r=2$ and $l>2$, (c) $r>2$ and $l>2$, and (d) $r=2$ and $1<l<2$.

3. CALCULATIONS AND RESULTS

We assume that the conductor fills the half-space $z>0$ and its surface is parallel to the layer planes. We also assume that a plane electromagnetic wave is incident normally on the surface. Since in layered organic metals the ratio v_z/v_0 is small, we can limit ourselves to mirror reflection of electrons from the boundary. In this case the surface impedance tensor is diagonal:

$$Z_{\alpha\alpha}=8i\omega\int_0^\infty \frac{dq}{4\pi i\omega\sigma_{\alpha\alpha}(\omega,q)-c^2q^2}, \tag{10}$$

where ω and $\mathbf{q}$ are the frequency and wave vector of the wave ($\mathbf{q}=(0,0,q)$), and $\sigma_{\alpha\alpha}(\omega,q)$ are the diagonal components of the electrical conductivity tensor, $\alpha=x,y$.

In this geometry the components σ_{xx} and σ_{yy} of the conductivity tensor for an axisymmetric Fermi surface and, correspondingly, the components of the surface impedance tensor are equal to each other:

$$\sigma_{xx}=\sigma_{yy}=\sigma=\frac{2ie^2}{(2\pi\hbar)^3 m_\perp}\int_{-p_0}^{p_0}\frac{S(p_z)dp_z}{\tilde{\omega}-qv_z}, \tag{11}$$

where we have written $\tilde{\omega}=\omega+i/\tau$ and τ the quasiparticle effective relaxation time. The maximum value of the longitudinal component of the velocity, v_z, is of order ηv_0. For small values of q, where the parameter u ($u=\omega/\eta q v_0$) assumes values much larger than unity, we can expand σ in a series in inverse powers of u:

$$\sigma=\sigma_0(1+Q_2u^{-2}+Q_4u^{-4}+\cdots)=\sigma_0[1+\Phi(1/u)]. \tag{12}$$

The leading approximation for σ_0 equals $iNe^2/\tilde{m}_\perp\tilde{\omega}$, where N is the charge carrier number density, and is independent of the specific features of the local geometry of the Fermi surface. The dimensionless coefficients Q_{2n} are specified by

$$Q_{2n}=\left(\left(\frac{\omega}{\tilde{\omega}}\right)^{2n}\int_0^{p_0}S(p_z)\overline{v_z^{2n}}dp_z\right)\Big/\left(\int_0^{p_0}S(p_z)dp_z\right), \tag{13}$$

where $\overline{v_z}=v_z/\eta v_0$. Their values depend on the parameters r and l, which determine the function $v_z(p_z)$. This, however, has no effect on the expansion (12).

For q large ($u\ll 1$), the conductivity can be expanded in a power series in u. If the Fermi surface given by (3) and (4) has no curvature anomalies in the extremal sections ($r=l=2$), the expansion is

$$\sigma=\frac{\pi}{i}\frac{\tilde{\omega}}{\omega}\,\sigma_0(u+\Lambda_1u^2+\Lambda_2u^3+\cdots)=\frac{\pi}{i}\frac{\tilde{\omega}}{\omega}\sigma_0(1+f(u)). \tag{14}$$

The first expansion coefficients in (14) are

$$\Lambda_1=-\frac{ib}{\pi}\frac{\tilde{\omega}}{\omega},\quad \Lambda_2=g\left(\frac{\tilde{\omega}}{\omega}\right)^2, \tag{15}$$

where b and g are dimensionless constants of order unity.

In calculating the surface impedance (10) it is convenient to integrate with respect to u and divide the integration range into regions of small and large values of u. In each region we can then employ the corresponding asymptotic behavior of the conductivity.

Thus ($Z_{xx}=Z_{yy}=Z$),

$$Z=Z_1+Z_2, \tag{16}$$

where

$$Z_1=\frac{8\eta v_0}{c^2}\int_0^1\frac{du}{\pi(\tilde{\omega}/\omega)\xi^2u^3(1+f(u))+i}, \tag{17}$$

$$Z_2=-\frac{8i\eta v_0}{c^2}\int_1^\infty\frac{du}{\xi^2u^2(1+\Phi(1/u))+1}. \tag{18}$$

Integration with respect to u in (17) and (18) can easily be done in the limits of large and small absolute values of the anomaly parameter ξ:

$$\xi=\eta\frac{\omega_p}{\sqrt{\omega\tilde{\omega}}}\frac{v_0}{c},$$

where $\omega_p=\sqrt{4\pi Ne^2/m_\perp}$ is the plasma frequency. For $|\xi|\ll 1$ the main contribution to the surface impedance is provided by the region of large values of u and is

$$Z\approx\frac{4\pi}{ic}\frac{\sqrt{\omega\tilde{\omega}}}{\omega_p}\left(1+\frac{\xi^2}{2\pi}Q_2\right). \tag{19}$$

In the opposite limit $|\xi|\gg 1$, the principal part of the impedance is determined by the region of small values of u and is

$$Z\approx\frac{8\pi}{3\sqrt{3}c}\left(\frac{\omega^2}{\omega_p^2}\frac{\eta v_0}{\pi c}\right)^{1/3}\left\{1-i\sqrt{3}-\frac{2\Lambda_1}{3\xi^{2/3}}\left(\frac{\omega}{\pi\tilde{\omega}}\right)^{1/3}(\sqrt{3}-i)\right\}. \tag{20}$$

The leading approximation for the conductivity in the region of small values of q (large u) is independent of q. Thus, for $|\xi|\ll 1$, the link between the electric field and current is local, which is characteristic of the skin effect. The skin depth δ is given by the following expression:

$$\frac{1}{\delta}=\frac{\omega}{\eta v_0}\xi',$$

where $\xi=\xi'+i\xi''$. At low frequencies ($\omega\tau\ll 1$),

$$\xi'=\xi''=\eta\frac{\omega_p}{\omega}\frac{v_0}{c}\sqrt{\frac{\omega\tau}{2}},$$

i.e., $|\xi|=\sqrt{2}l/\delta\omega\tau$. Here $l=\eta v_0\tau$ is the mean free path of the charge carriers along the normal to the layer plane. The inequality $|\xi|\ll 1$ is valid under normal skin effect conditions ($l\ll\delta$). At high frequencies ($\omega\tau\gg 1$) we have $|\xi|=\xi'=l/\delta\omega\tau$. Due to the presence of the large factor $\omega\tau$ in the denominator of the expression for $|\xi|$, the inequality $|\xi|\ll 1$ is valid if $l<\delta$, a condition that can easily be met in layered conductors.

The leading term in the asymptotic expression for the impedance with $|\xi|\gg 1$ corresponds to an anomalous skin effect with a skin depth

$$\delta=\frac{2\eta v_0}{\sqrt{3}\omega}\left(\frac{\omega}{\pi\tilde{\omega}\xi^2}\right)^{1/3}=\frac{2}{\sqrt{3}}\left(\frac{c^2\eta v_0}{\pi\omega\omega_p^2}\right)^{1/3}.$$

When $\omega\tau\ll 1$ holds, $|\xi|$ takes a value of order $(l/\delta)^{3/2}/\omega\tau$, while in the opposite limit $\omega\tau\gg 1$ this parameter is of order $(l/\delta\omega\tau)^{3/2}$. Since the mean free path in the direction perpendicular to the conducting layers is small, for layered organic metals it is essentially impossible to meet the condition $l\gg\delta$. This means that, in contrast to ordinary metals, it is impossible to observe an anomalous skin effect in such substances in the high-frequency range ($\omega\tau\gg 1$).

In the limit $\omega\tau\ll 1$, the skin effect is anomalous ($|\xi|\gg 1$) for $l>\delta$. This condition can easily be met at moderate frequencies, since the skin depth increases with decreasing frequency. The intermediate frequency range $\omega\tau\sim 1$ is optimal for realizing an anomalous skin effect in layered conductors. The maximum value of $|x|$ is reached at $\omega\tau=1/\sqrt{2}$ and is of order $\eta\omega_p\tau v_0/c$. The ratio $\omega_p\tau v_0/c$ in a pure ($\tau\sim 10^{-8}$ s) layered conductor is of order 10^3–10^4. Thus, at moderate values of η ($\eta\sim 10^{-2}$) the maximum value of $|\xi|$ may reach 10^2. This means that in layered conductors both the normal skin effect ($|\xi|\ll 1$) and the anomalous skin effect ($|\xi|\gg 1$) can be present, although the latter is observed in a frequency range narrower than that in ordinary metals. On the other hand, at very small values of η, when the Fermi surface is for all practical purposes a cylinder, the condition

$|\xi| \ll 1$ is met over the entire frequency range. Accordingly, we can use the asymptotic formula (19) for the surface impedance corresponding to a normal skin effect at all frequencies.

The leading terms in the asymptotic expressions for the surface impedance in both limits, the normal skin effect ($|\xi| \ll 1$) and the anomalous skin effect ($|\xi| \gg 1$), are independent of the specific characteristics of the Fermi surface and coincide with the results obtained by Gokhfel'd and Peschanskiĭ,[16] who used a model of the Fermi surface based on the tight-binding approximation for electrons.

Let us now examine a conductor whose Fermi surface has anomalies of the Gaussian curvature in the effective sections. For definiteness, we assume that in the model (3) and (4) we have $l=2$ and $r \neq 2$. This corresponds to curvature anomalies at $p_z = \pm p_0$. In this case the asymptotic expansion of the conductivity in the region where the parameter u is small contains an additional term σ_a:

$$\sigma_a = \frac{\pi}{2i}\sigma_0 \mu_\beta \left(u \frac{\tilde{\omega}}{\omega} \right)^{\beta+1}, \tag{21}$$

where

$$\mu_\beta = (\beta+1)\left(1 - i \tan \frac{\pi\beta}{2} \right), \quad -\beta = \frac{r-2}{r-1}.$$

If the sections of the Fermi surface at $p_z = \pm p_0$ are lines of parabolic points ($r>2$), the parameter β takes negative values ($-1<\beta<0$). The shape of the Fermi surface in the vicinity of these sections is close to cylindrical, and as $\beta \to -1$, the Fermi surface resembles a cylinder more closely. For $1<r<2$ the parameter β assumes positive values. In the given case, the curvature of the Fermi surface in the vicinity of the sections corresponding to $p_z = \pm p_0$ becomes anomalously large. In the low-frequency limit, Eq. (21) can be written as const/$(ql)^\gamma$, where $\gamma = 1/(r-1)$. The same asymptotic behavior of the contribution to the conductivity of the quasicylindrical section of the Fermi surface of a $3D$-metal was obtained by Kaganov and Contreras[10] (see also Ref. 20).

A comparison of (21) and (14) suggests that if the Fermi surface of a layered conductor near the sections with an extremal (in our case minimum) diameter closely resembles a cylinder, σ_a exceeds all other terms in the expansion of the conductivity in powers of u. Accordingly, the leading approximation for the impedance in the event of an anomalous skin effect is determined by the contribution of the quasicylindrical section of the Fermi surface:

$$Z = \frac{8\pi}{c} W(\beta) \frac{\sqrt{\omega\tilde{\omega}}}{i\omega_p} \left(\xi \frac{\omega}{\tilde{\omega}} \right)^{(\beta+1)/(\beta+3)}, \tag{22}$$

where

$$W(\beta) = \left[\frac{2\cos(\pi\beta/2)}{\pi(\beta+1)} \right]^{1/(\beta+3)} \left(1 + i \cot \frac{\pi}{\beta+3} \right) \frac{1}{\beta+3}$$

$$\equiv w(\beta)\left(1 + i \cot \frac{\pi}{\beta+3} \right). \tag{23}$$

In the limit $\beta \to -1$, the complex-valued function $W(\beta)$ tends to 1/2 and (21) becomes the leading term in (18). Here the dependence on the parameter η, which characterizes the extent to which the Fermi surface is rippled, disappears. Thus, the presence of wide cylindrical belts on a highly pinched Fermi surface leads to the same result for the surface impedance as in the case of a weakly pinched Fermi surface.

Equation (23) implies that, to within a complex-valued constant ζ ($|\zeta| \sim 1$),

$$Z = \zeta |Z_0| (1 - i\omega\tau)^{-\beta/(\beta+3)} (\delta/l)^{-\beta/(\beta+3)}. \tag{24}$$

Here Z_0 is the leading approximation of the impedance for an anomalous skin effect in the case where the Fermi surface of the conductor has no curvature anomalies in the effective sections (the first term in (20)), and δ is the skin depth for the anomalous skin effect. We see that the impedance depends on the mean free path of the charge carriers. Kaganov and Contreras[10] found that such dependence exists only if there are quasicylindrical sections on the Fermi surface. If there are no such sections, the leading approximation Z_0 of the impedance for an anomalous skin effect is independent of l.

In the limit $\beta \to -1$, the exponent of δ/l in (24) assumes values close to 1/2. Thus, if the Fermi surface in the vicinity of the effective sections at $p_z = \pm p_0$ resembles a cylinder, the impedance for an anomalous skin effect is proportional to $1/\sqrt{l}$, in the same way as it is for a normal skin effect. For $\omega\tau \ll 1$ and $\beta = -2/3$ ($r=4$), Eq. (24) coincides with the result of Ref. 10.

At sufficiently high frequencies ($\omega\tau > 1$), the real part of the surface impedance (24) is

$$R = \frac{8\pi}{c} w(\beta) \left(\eta \frac{\omega_p}{\omega} \frac{v_0}{c} \right)^{(\beta+1)/(\beta+3)} \times \frac{\omega}{\omega_p} \left[\cot \frac{\pi}{\beta+3} - \frac{\beta}{\beta+3} \frac{1}{\omega\tau} \right]. \tag{25}$$

For negative values of β not too close to -1 we have $R \sim \omega^{2/(\beta+3)}$. The real part of the impedance increases with frequency faster than in the case of a conductor for which the curvature of the Fermi surface in the effective section is finite. In the limit $\beta \to -1$ the frequency dependence of R disappears, as it does for a purely cylindrical surface.

Let us assume that the parameter β, which characterizes the shape of the Fermi surface, takes positive values. This means that the curvature of the Fermi surface at $p_z = \pm p_0$ becomes infinite. The term σ_a ceases to be the leading term in the asymptotic expression for the conductivity for small u. However, for $0<\beta<1$, when the singularity in the curvature is not pronounced, σ_a is the first correction to the leading approximation. Here the asymptotic behavior for an anomalous skin effect is given by

$$Z = \frac{8\pi}{3\sqrt{3}} \left(\frac{\omega^2}{\omega_p^2} \frac{\eta v_0}{\pi c} \right)^{1/3} \left\{ 1 - i\sqrt{3} - \frac{U(\beta)}{2\sqrt{3}} \frac{1}{(\pi\xi^2)^{\beta/3}} \left(\frac{\tilde{\omega}}{\omega} \right)^{2\beta/3} \right\}, \tag{26}$$

where

$$U(\beta) = \frac{(1+\beta^2)}{\cos(\pi\beta/2)} \left[\cot \frac{\pi(1+\beta)}{3} - i \right]. \tag{27}$$

The first correction to the leading approximation for the impedance is now larger than in the absence of curvature anomalies in the effective belts on the Fermi surface. The frequency dependence of this correction also changes. The additional term is proportional to $\omega^{2(1+\beta)/3}$ rather than to $\omega^{4/3}$ (the latter case corresponds to a Fermi surface without curvature anomalies).

Note that the surface impedance for an anomalous skin effect is also described by expressions of the form (26) in the case where narrow neighborhoods of some (but not all) extremal sections of the Fermi surface resemble narrow cylindrical bands. Here the anomalous additional term in the conductance is

$$\sigma_a=\frac{\pi}{2i}\rho\mu_\beta\sigma_0\left(u\frac{\tilde{\omega}}{\omega}\right)^{\beta+1} \tag{28}$$

and contains a small positive factor ρ describing the relative number of effective charge carriers related to the cylindrical section of the Fermi surface. The parameter β takes negative values in the interval $-1<\beta<0$. For small values of ρ and moderate values of β not too close to -1, σ_a is smaller than the leading term in the expansion (14) for σ and must be taken into account as the first-order correction. Here for the component of the surface impedance tensor we arrive at a result that differs from (26) in that the term describing the first correction contains an additional factor ρ and $\beta<0$ in this term.

The above analysis can be repeated for a Fermi surface that has anomalies in the Gaussian curvature at $p_z=0$. Such a surface is described by Eqs. (3) and (4) with $r=2$ and $l\neq 2$. As a result we arrive at expressions that coincide with (20)–(26) with a parameter β, which characterizes the shape of the Fermi surface near the effective section $p_z=0$, expressed in terms of l $(-\beta=(l-2)/(l-1))$.

Finally, let us consider the case where the effective sections of the Fermi surface have curvature anomalies. If $r>2$ and $l>2$ hold, the neighborhood of each section of the Fermi surface with an extremal diameter resembles a cylinder. Here the asymptotic behavior of the surface impedance for an anomalous skin effect retains its form (22). The same expression (22) describes the asymptotic behavior of the surface impedance of a layered conductor for an anomalous skin effect in the case where the curvature of the Fermi surface of the conductor is anomalously large in all sections with minimum and maximum diameters ($1<r<2$ and $1<l<2$). In both cases the value of β is expressed in terms of the larger of the two parameters, r and l.

If one of the parameters, r or l, or both are much larger than unity, the asymptotic expression for the surface impedance ceases to depend on the rippling parameter η and assumes the form (19). Thus, the characteristic features of the skin effect in layered conductors are determined not by the smallness of the second term in Eq. (3) by itself but by the form of the function $E(p_z/p_0)$, which specifies the profile of this surface. The fact that usually for layered conductors the projection of the charge-carrier velocity onto the normal to the layer plane is small in comparison to the velocity components describing the motion in the layer plane is also true for the case $\eta\sim 1$. In particular, for $r\gg 1$ and $l\gg 1$, the surface described by (3) and (4) is a steplike cylinder. The longitudinal velocity v_z of the electrons on such a Fermi surface is zero for an arbitrary value of the corrugation parameter η. The local features of the geometry of the Fermi surface of layered conductors lead to specific singularities in their observed properties in an external magnetic field. This is true, in particular, for cyclotron resonance, which in recent years has been repeatedly observed in organic metals.[21–26]

Suppose that an external magnetic field is directed along the normal to the surface of a semi-infinite conductor and that this surface is parallel to the planes of the conducting layers. When the charge carriers undergo a mirror reflection from the boundary, the surface impedance tensor becomes diagonal in terms of the circular components:

$$Z_\pm=8i\omega\int_0^\infty\frac{dq}{4\pi i\omega\sigma_\pm(\omega,q)-c^2q^2}, \tag{29}$$

where $\sigma_\pm=\sigma_{xx}\pm i\sigma_{yx}$ are the circular components of the transverse conductivity.

The asymptotic expressions for $\sigma_\pm$ for large and small u coincide with the expansions (12) and (14) in which $\tilde{\omega}$ is replaced by $\tilde{\omega}_\pm$ ($\tilde{\omega}_\pm=\omega\mp\Omega+i/\tau$, where Ω is the cyclotron frequency of the charge carriers). The same is true of the expression for $\sigma_a^\pm$. Below we examine the case where the polarization corresponds to cyclotron resonance.

If the Fermi surface of the layered conductor has wide sections resembling cylinders, the impedance for an anomalous skin effect is described by an expression obtained from (22) by substituting $\tilde{\omega}_+$ for $\tilde{\omega}$. This substitution must be carried out everywhere, including the expression for the parameter ξ characterizing the extent to which the skin effect is anomalous. The real part of the impedance in such conditions is

$$\frac{R(H)}{R(0)}=\left(\sqrt{\left(1-\frac{\Omega}{\omega}\right)^2+\frac{1}{(\omega\tau)^2}}\right)^{-\beta/(\beta+3)}\frac{\cos(Y_\beta(\Omega,\omega,\tau))}{\cos(\pi/(\beta+3))}, \tag{30}$$

where

$$Y_\beta(\Omega,\omega,\tau)=\frac{1}{\beta+3}\left\{\pi+\pi\beta\theta(\Omega-\omega)+\beta\ \mathrm{arccot}\left[\omega\tau\left(1-\frac{\Omega}{\omega}\right)\right]\right\},$$

with

$$\theta(x)=\begin{cases}1 & \text{if } x\geq 0,\\ 0 & \text{if } x<0,\end{cases}\quad -1<\beta<0. \tag{31}$$

At a fixed frequency ω the value of $R(H)$ rapidly increases with the magnetic field strength in fields near cyclotron resonance. The discontinuity in $R(H)$ at $H=H_r$ is no smaller in order of magnitude than $R(0)$ at the same frequency. In strong magnetic fields ($H\gg H_r$), where $\omega\ll\Omega$, the value of $R(H)$ increases in proportion to $(\Omega/\omega)^{-\beta/(\beta+3)}$. For moderate (in absolute value) values of β, the increase in $R(H)$ in a strong magnetic field is slow as $\beta\to -1$, when the shape of the effective belts on the Fermi surface is essentially cylin-

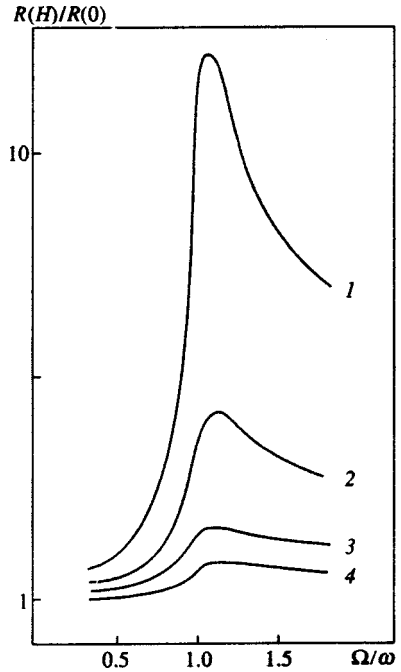


FIG. 2. Dependence of the active part of the surface impedance of a layered organic metal whose Fermi surface has near-cylindrical wide bands, on the magnetic field near cyclotron resonance. The curves are described by Eq. (30) and are plotted for $\omega\tau=10$ with $\beta=-0.2$ (curve *1*), -0.4 (curve *2*), -0.6 (curve *3*), and -0.8 (curve *4*).

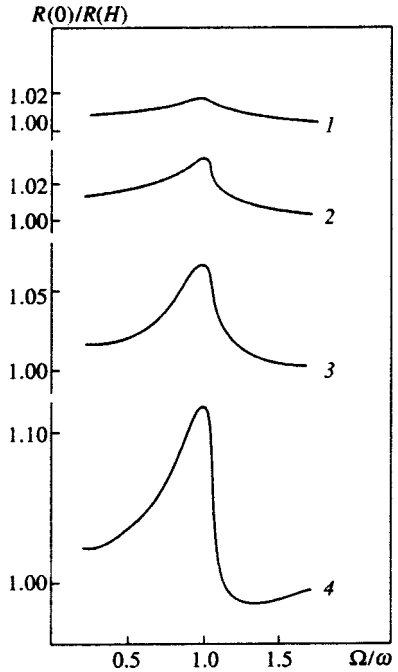


FIG. 3. Dependence of the active part of the surface impedance of a layered organic metal on the magnetic field near cyclotron resonance for the case where one of the extremal sections of the Fermi surface coincides with the line of parabolic points. The curves are described by Eq. (32) and are plotted for $\omega\tau=10$ and $\rho=0.05$ with $\beta=-0.2$ (curve *1*), -0.3 (curve *2*), -0.4 (curve *3*), and -0.5 (curve *4*).

drical, and for $H\gg H_r$ we have $R(H)\sim\sqrt{\Omega/\omega}$. The field dependence of the ratio $R(H)/R(0)$ near cyclotron resonance is depicted in Fig. 2, where the curves are described by (30).

If the cylindrical belts near extremal sections of the Fermi surface are narrow, the related additional term in the conductivity determines the first correction to the leading approximation of impedance for an anomalous skin effect. The leading term in the impedance in this case is independent of the magnetic field. Substituting $\tilde{\omega}_+$ for $\tilde{\omega}$ in (26) and allowing for the small factor ρ, which describes the width of the cylindrical belts on the Fermi surface, we get

$$\frac{R(H)}{R(0)}=1+\tilde{\rho}\tilde{\xi}^{-2\beta/3}\left(\sqrt{\left(1-\frac{\Omega}{\omega}\right)^2+\frac{1}{(\omega\tau)^2}}\right)^{\beta}\times\frac{\cos(Y'_{\beta}(\Omega,\omega,\tau))}{\sin(\pi(1+\beta)/3)},\tag{32}$$

where

$$Y'_{\beta}(\Omega,\omega,\tau)=\frac{\pi}{3}(2-\beta)+\pi\beta\theta(\Omega-\omega)+\beta\ \mathrm{arccot}\left[\omega\tau\left(1-\frac{\Omega}{\omega}\right)\right],\tag{33}$$

$$\tilde{\rho}=\frac{(1+\beta)^2}{2\sqrt{3}\cos(\pi\beta/2)}\rho,\quad \tilde{\xi}=\sqrt{\pi}\,\frac{\omega_p}{\omega}\,\frac{\eta v_0}{c}.$$

This result is valid for $\tilde{\rho}\tilde{\xi}^{-2\beta/3}\ll 1$. At high frequencies the parameter $\tilde{\xi}$ assumes values close to those of the anomaly parameter ξ in the absence of an external magnetic field. Taking for $\tilde{\xi}$ a value on the order of the maximum of ξ ($\tilde{\xi}\sim 10^{-2}$), we see that for a moderate curvature anomaly ($-0.5<\beta<0$) the above inequality holds if we have $\rho<0.1$.

For $\beta<0$ the second term in (32) describes the peak in the field dependence of $R(H)$ related to cyclotron resonance (Fig. 3). The height of this peak depends on ρ. For moderate values of ρ ($\rho<0.1$) the height may amount to 10% of the leading approximation of the real part of the impedance. For moderate values of $\omega\tau$ the top of the peak (it corresponds to the minimum of the resonance term in (32)) is appreciably shifted in relation to H_r:

$$\frac{H_r-H}{H_r}\equiv\frac{\Delta H}{H_r}=\frac{\cot\Phi}{\omega\tau}.\tag{34}$$

The value of Φ is determined by the shape of the effective section of the Fermi surface:

$$\Phi=\frac{\pi}{6}\left(1-\frac{\beta}{1-\beta}\right).\tag{35}$$

Such field behavior near cyclotron resonance has been observed in some layered organic metals of the α-(BEDT-TTF)$_2$MHg(SCN)$_4$ group[25] and in the organic conductor (BEDO-TTF)$_2$ReO$_4$(H$_2$O) (Ref. 26). The experiments of Demishev *et al.*[25,26] were conducted in the 30–100 GHz frequency range in magnetic fields of about 50 kOe with $\omega\tau\sim 10$. The magnitude of the singularities related to cyclotron resonance amounted to several percentage points

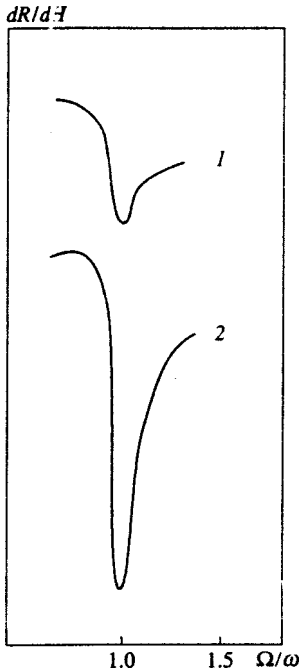


FIG. 4. Cyclotron-resonance induced singularities in the field dependence of dR/dH for a layered conductor in the case where one extremal sections of the Fermi surface coincides with line of parabolic points. The curves are plotted for $\omega\tau=20$ and $\rho=0.05$ with $\beta=-0.25$ (curve *1*) and -0.75 (curve *2*).

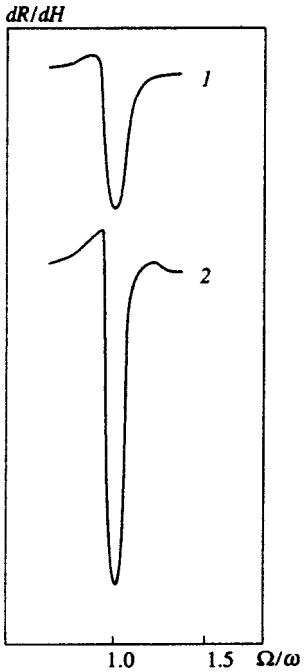


FIG. 5. Cyclotron-resonance induced singularities in the field dependence of dR/dH for a layered conductor whose Fermi-surface curvature has a singularity in one of the extremal sections. The curves are plotted for $\omega\tau=20$ with $\beta=0.25$ (curve *1*) and 0.5 (curve *2*).

of the leading impedance value,[25] which agrees with estimates that follow from (32). The results of Demishev *et al.*[25,26] can serve as proof of the presence of cylindrical sections on the Fermi surface of the layered conductors used in their studies.

The data of Ref. 26 allows estimating the parameter β, which characterizes the shape of the quasicylindrical section of the Fermi surface of $(BEDO\text{-}TTF)_2ReO_4(H_2O)$, via (34) and (35). The experiment was conducted at frequencies correspond to variations in $\omega\tau$ in the interval from 10 to 20 and to variations in $\Delta H/H_r$ in the interval from 0.11 to 0.06. This yields $\beta\approx-(0.25-0.35)$.

The resonance peak in the field dependence of the active part of the impedance must be accompanied by a stronger singularity in the field dependence of dR/dH. If the curvature of the Fermi surface rapidly increases as we move away from the line of parabolic points, the height of the peak in the field dependence of $R(H)$ corresponding to cyclotron resonance may prove to be too small to be observable. However, the resonance singularity in the derivative of the impedance under these conditions may clearly manifest itself. The resonant singularities in the field and frequency dependence of dR/dH can also be observed if $\beta>0$. This corresponds to an anomalously large curvature of the Fermi surface in the vicinity of sections with minimum or maximum diameter. Cyclotron resonance can be observed at moderate values of the parameter β $(0<\beta<1)$. In contrast to the above case for positive β, cyclotron resonance does not manifest itself in the field dependence of the real part of the resonance. The field-dependent term in the expression (32) for $R(H)$ decreases monotonically with increasing magnetic field. The resonance dependence of dR/dH near cyclotron resonance is depicted in Figs. 4 and 5 for several values of the parameter β.

4. CONCLUSION

It is too early to draw any conclusions about the local features of the geometry of the Fermi surface of the majority of layered organic metals, since there is a lot to study in the electron energy spectra of such materials. It can be assumed, however, that here, as in ordinary metals, the Fermi surface contains quasicylindrical bands or sections with an anomalously large curvature. These features of the local geometry of the Fermi surface can be created (if they are absent) or enhanced by applying an agent that changes the shape of the constant-energy surfaces, e.g., by applying external pressure along the normal to the conducting planes.

The above analysis shows that the special features in the profile of the corrugated cylinder, which is the main part of the Fermi surface of layered organic metals, can substantially change the high-frequency properties of these materials. The model developed in this paper makes it possible to study in detail the observable manifestations of the local geometry of the Fermi surface of layered conductors. It resolves some of the difficulties that emerge when one uses the model of tightly bound electrons. For instance, the characteristic features of the observable properties of layered conductors, for which the strong anisotropy of the electrical conductivity is responsible, can be described and analyzed without passing to the limit $\eta\to0$, which corresponds to a conductor with a two-dimensional energy spectrum of the charge carriers.

The model specified by Eqs. (3) and (4) makes it possible to do a detailed study of the frequency dependence of

the surface impedance of layered conductors with different Fermi-surface profiles; with it we can analyze on a unified basis all possible manifestations of cyclotron resonance in a magnetic field that is perpendicular to the surface of the conductor. By using this model, we can investigate the quantum oscillations of the thermodynamic characteristics of layered conductors, and the results of such research could be interesting.

*)E-mail: drig@ugi.rcupi.e-burg.su

[1] V. Z. Kresin, Phys. Rev. B **35**, 8716 (1987).
[2] W. Kang, G. Montambaux, J. R. Cooper, D. Jérome, P. Batail, and C. Lenoir, Phys. Rev. Lett. **62**, 2559 (1989).
[3] M. V. Kartsovnik, P. A. Kononovich, V. N. Laukhin, S. I. Pesotskiĭ, and I. F. Shchegolev, Zh. Éksp. Teor. Fiz. **97**, 1305 (1990) [Sov. Phys. JETP **70**, 735 (1990)].
[4] F. L. Pratt, J. Singleton, M. Doporto, A. J. Fisher, T. J. B. M. Janssen, J. A. A. J. Perenboom, M. Kurmoo, W. Hayes, and P. Day, Phys. Rev. B **45**, 13904 (1992).
[5] N. Harrison, A. House, I. Deckers, J. Caulfield, J. Singleton, F. Herlach, W. Hayes, M. Kurmoo, and P. Day, Phys. Rev. B **52**, 5584 (1995).
[6] J. Wosnitsa, *Fermi Surface of Low-Dimensional Organic Metals and Superconductors*, Springer, Berlin (1996).
[7] V. M. Kontorovich and N. A. Sapogova, JETP Lett. **18**, 165 (1973).
[8] G. T. Avanesyan, M. I. Kaganov, and T. Yu. Lisovskaya, JETP Lett. **25**, 355 (1977).
[9] M. I. Kaganov and Yu. V. Gribkova, Fiz. Nizk. Temp. **17**, 907 (1991) [Sov. J. Low Temp. Phys. **17**, 473 (1991)].
[10] M. I. Kaganov and P. Contreras, Zh. Éksp. Teor. Fiz. **106**, 1814 (1994) [JETP **79**, 360 (1994)].
[11] Ya. M. Blanter, M. I. Kaganov, A. V. Pantsulava, and A. A. Varlamov, Phys. Rep. **245**, 159 (1994).
[12] N. A. Zimbovskaya, Fiz. Nizk. Temp. **20**, 441 (1994) [Low Temp. Phys. **20**, 350 (1994)].
[13] N. A. Zimbovskaya, Zh. Éksp. Teor. Fiz. **107**, 1672 (1995) [JETP **80**, 932 (1995)].
[14] N. A. Zimbovskaya, Fiz. Nizk. Temp. **22**, 1137 (1996) [Low Temp. Phys. **22**, 869 (1996)].
[15] V. M. Gokhfel'd, M. I. Kaganov, and V. G. Peschanskiĭ, Fiz. Nizk. Temp. **12**, 1173 (1986) [Sov. J. Low Temp. Phys. **12**, 661 (1986)].
[16] V. M. Gokhfel'd and V. G. Peschanskiĭ, Ukr. Fiz. Zh. **37**, 1594 (1992).
[17] V. M. Gokhfel'd, O. V. Kirichenko, and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **108**, 2147 (1995) [JETP **81**, 1171 (1995)].
[18] O. B. Kirichenko and V. G. Peschanskiĭ, Fiz. Nizk. Temp. **20**, 574 (1994) [Low Temp. Phys. **20**, 453 (1994)].
[19] N. A. Zimbovskaya, Phys. Low-Dimens. Semicond. Struct. **11/12**, 29 (1996).
[20] N. A. Zimbovskaya, V. I. Okulov, A. Yu. Romanov, and V. P. Silin, Fiz. Met. Metalloved. **62**, 1095 (1986).
[21] J. Singleton, F. L. Pratt, M. Doporto, T. J. B. M. Janssen, M. Kurmoo, J. A. A. J. Perenboom, W. Hayes, and P. Day, Phys. Rev. Lett. **68**, 2500 (1992).
[22] J. Singleton, F. L. Pratt, M. Doporto *et al.*, Physica B **184**, 470 (1993).
[23] A. S. Perel, J. S. Brooks, C. J. G. M. Langerak, T. J. B. M. Janssen, J. Singleton, J. A. A. J. Perenboom, and L. Y. Chiang, Phys. Rev. Lett. **67**, 2072 (1991).
[24] S. V. Demishev, N. R. Sluchanko, A. V. Semeno, and N. A. Samarin, JETP Lett. **61**, 313 (1995).
[25] S. V. Demishev, A. V. Semeno, N. E. Sluchanko, N. A. Samarin, I. N. Voskoboinikov, V. V. Glushkov, J. Singleton, S. J. Blundell, S. O. Hill, W. Hayes, M. V. Kartsovnik, A. E. Kovalev, M. Kurmoo, P. Day, and N. D. Kurshch, Phys. Rev. B **53**, 12794 (1996).
[26] S. V. Demishev, A. V. Semeno, N. E. Sluchanko, N. A. Samarin, I. B. Voskoboĭnikov, M. V. Kartsovnik, A. E. Kovalev, and N. D. Kushch, Zh. Éksp. Teor. Fiz. **111**, 979 (1997) [JETP **84**, 540 (1997)].

Translated by Eugene Yankovsky

Antiferromagnetic resonance in Bi_2CuO_4

L. E. Svistov,*) V. A. Chubarenko, A. Ya. Shapiro, and A. V. Zalesskiĭ

A. V. Shubnikov Institute of Crystallography, Russian Academy of Sciences, 177333 Moscow, Russia

G. A. Petrakovskiĭ

L. V. Kirenskii Institute of Physics, Siberian Department of Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 10 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2244–2255 (June 1998)

Magnetic resonance of the low-frequency spin-wave branch in the Bi_2CuO_4 antiferromagnet with an easy-plane anisotropy has been studied. Angular, frequency, and temperature dependences of the position and width of the antiferromagnetic resonance (AFMR) line have been measured. Our measurements combined with earlier data [H. Ochta, K. Yoshida, T. Matsuya, T. Nanba, M. Motokawa, K. Yamada, Y. Endon, and S. Hosoya, J. Phys. Soc. Jpn. **61**, 2921 (1992); E. W. Ong, G. H. Kwei, R. A. Robinson, B. L. Ramakrishna, and R. B. von Dreele, Phys. Rev. B **42**, 4255 (1990)] have allowed us to determine anisotropy constants of this material and to account for the unusual character of its static susceptibility anisotropy. The AFMR line shifts to the high-field side and broadens in a temperature range of 10–15 K, and the cause of this has remained unclear. In the low-temperature range the line shows a hysteresis corresponding to a static field magnitude several times as large as the spin-flop field. The position and width of the AFMR line depend sensitively on the sample preparation technique. © *1998 American Institute of Physics.* [S1063-7761(98)02406-8]

1. INTRODUCTION

Study of materials containing ions of variable valence is one of rapidly developing branches of modern physics. They attract researchers' attention primarily in connection with the discoveries of high-temperature superconductivity and giant magnetoresistance in such materials.

One representative of this class is Bi_2CuO_4, whose structure belongs to the space group $P4/ncc$.[2–4] The magnetic moment of the Cu^{2+} cation is due to a single uncompensated electron in the d-shell. Neutron diffraction studies[2,4–6] of Bi_2CuO_4 indicate the presence of three-dimensional magnetic ordering at temperatures below $T_N \simeq 42$ K.

The unit cell of Bi_2CuO_4 contains four copper ions. In the conventional notation,[4] the positions of these ions in the lattice are as follows: Cu(1), $(1/4,1/4,z)$; Cu(2), $(1/4,1/4,z+1/2)$; Cu(3), $(3/4,3/4,1/2-z)$; Cu(4), $(3/4,3/4,-z)$. The parameter z is the shift of the parallelepipeds formed by Cu(1) and Cu(2) ions with respect to those of Cu(3) and Cu(4) ions along the $C^{(4)}$ axis. The value of z equals 0.076 of the lattice constant c.[4] The chains of Cu(1), Cu(2) and Cu(3), Cu(4) ions form two magnetic sublattices in the antiferromagnetically ordered state of Bi_2CuO_4. The magnetic anisotropy for Cu^{2+} ions is determined by the anisotropic exchange, since a one-ion anisotropy due to the electric crystal field does not affect ions with spin 1/2. The exchange anisotropy aligns the magnetic moments of the sublattices within the plane perpendicular to the four-fold axis.[6,7] In the absence of magnetic field, the antiferromagnetic vector is directed along a diagonal of the (a,b) square in the easy plane.[8] The effective magnetic moment of the Cu^{2+} ion in antiferromagnetic Bi_2CuO_4 at 4.2 K is $0.65-0.85\mu_B$.[6,7]

The electron spin resonance in the magnetically ordered phase ($T<T_N$) was studied in the submillimeter wave region.[1] From these measurements, the energy gap separating the exchange (high-frequency) branch of the magnon spectrum was derived. The antiferromagnetic resonance of the lower branch of the magnon spectrum in Bi_2CuO_4 single crystals grown from a melt was studied by Pankrats et al.[8] The frequency and angular dependences of the antiferromagnetic resonance field in the microwave and submillimeter wave bands were in good agreement with the magnetic structure of Bi_2CuO_4 determined previously.[6,7]

The static magnetic characteristics of samples depend sensitively on their preparation technique. Crystals grown by the hydrothermal technique[9] have in the ordered state a small ferromagnetic moment m in the basal plane of less than one percent of the nominal Cu^{2+} magnetic moment, which drops as the static magnetic field grows and vanishes for $H>30$ kOe ($T=4.2$ K). Samples grown from a melt do not manifest a weak ferromagnetism.[5,6] In the range of strong static fields, the susceptibilities of both types of samples are equal. The small ferromagnetic moment in samples made by the hydrothermal method was attributed to the Dzyaloshinskii–Moria interaction.[9]

In the present work, we have studied the microwave electron spin resonance of Bi_2CuO_4 single crystals of both types in the temperature range below T_N, and also in samples annealed after growth in an oxygen–helium atmosphere.

2. SAMPLES AND EXPERIMENTAL TECHNIQUES

We have studied Bi_2CuO_4 samples grown both by the flux method and hydrothermal technique.[10,11,13]

The samples were tested by x-ray diffraction and x-ray microanalyzer. x-ray diffraction measurements using a D/max-3C microdiffractometer produced by Rigaku demonstrated identical crystal structures of samples prepared by the two techniques. Nonetheless, the quantitative analysis of the samples performed on a JXA-8600 electron microscope combined with an x-ray analyzer produced by JEOL, in which we used the ZAF-correction program and reference samples of copper and bismuth (the oxygen content was determined by subtraction) revealed differences between compositions of crystals made by the two different methods. Samples grown from a melt had the composition $Bi_{1.9}CuO_{3.87}$, whereas samples prepared by the hydrothermal technique had the formula $Bi_{2.05}CuO_{3.84}$.

The comparison between the formulas of crystals grown by different techniques indicates that the degree of oxidation of cations in crystals grown from the hydrothermal solution is slightly lower than in samples grown from a melt. This difference may be caused by the fact that it is difficult to create a high oxidation potential in an autoclave, therefore hydrothermal crystals are deficient in oxygen, and a small quantity of nonmagnetic Cu^+ cations is present.

Experiments were performed on both as-grown crystals and samples annealed after growth in an atmosphere of oxygen or nitrogen. The annealing was performed at 600°C for 30 h. Note that the annealing in oxygen and nitrogen did not affect the positions and widths of AFMR lines in samples grown from a melt, but radically changed magnetic properties of samples made by the hydrothermal technique.

Typical dimensions of Bi_2CuO_4 samples grown from a melt and hydrothermally were $2\times2\times0.5$ and $1\times1\times0.1$ mm^3, respectively.

The electron spin resonance was measured on a transmission microwave spectrometer. A sample was placed in a rectangular cavity whose resonant modes were in the frequency band of 23–78 GHz. The device was designed to allow sample rotation during an experiment. A magnetic field of 0 to 40 kOe was produced by a superconducting solenoid. The cell with a sample was in a vacuum jacket, so that we could perform our measurements in a temperature range of 1.2 to 110 K. Measurements at a frequency of 9.3 GHz were performed on a commercial ESR spectrometer produced by Brucker.

3. EXPERIMENTAL RESULTS

3.1. Measurements of samples grown from melt solution

Figure 1 shows traces of microwave power transmitted through the cavity versus magnetic field at various temperatures. In the temperature range close to T_N, the ESR line is considerably broadened, and a new line forms at a magnetic field slightly higher than that of ESR. In the low-temperature range, $T<12$ K, the absorption line shape depends on the field scan direction. The solid curves in Fig. 1 were recorded in an increasing magnetic field, and the dashed curves in a decreasing field. The hysteresis behavior is more pronounced at lower magnetic fields, corresponding to lower resonant frequencies.

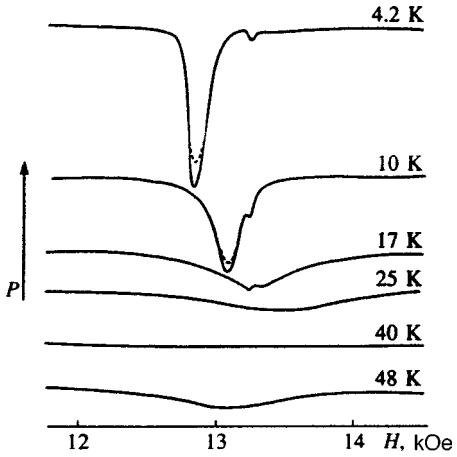


FIG. 1. Typical curves of microwave power transmitted through the cavity loaded with a Bi_2CuO_4 sample grown from a melt versus magnetic field at several temperatures. The narrow resonance line whose position is independent of the temperature is due to a DPPH sample, which is used as a reference for measuring the applied static magnetic field. The solid curves were recorded when the field was scanned in the upward direction, the dashed lines correspond to decreasing magnetic field; $\nu=36$ GHz, $H\|[110]$.

The position of the absorption line is strongly anisotropic. Measurements of angular dependence performed at different microwave frequencies and temperature have demonstrated that resonance conditions are determined by the static field projection on the plane perpendicular to the fourfold symmetry axis. When the static field was rotated in the basal plane, the line shift was well described by the function $A\cos(4\varphi)$. The resonant field had maxima when its direction coincided with the a- or b-axis.

The resonant field versus frequency is close to a linear function. The absorption peak is close to the ESR position of a paramagnet with g-factor $g=2$. Note that resonant fields measured in earlier experiments with the same configuration[1,8,12,14] were also close to a straight line of the same slope. Figure 2 shows resonant fields of samples with respect to the resonant field of diphenylpicrylhydrazyl (DPPH), which was used as a reference ($g=2$ for DPPH). It is clear that the curve of the frequency dependence is steeper than that of the free radical at 25 K, whereas at lower temperatures it has a gentler slope.

The temperature dependence of the resonant field at a microwave frequency of 36 GHz is plotted in Fig. 3. Experiments have been performed with two orientations of the static field in the basal plane, $\varphi=0°$ and 45°. The resonant field shifts to higher values in the temperature range of 10–15 K.

As was noted above, experiments have been performed on both as-grown samples and crystals annealed in the atmosphere of oxygen or nitrogen. Within the experimental uncertainty, the annealing has no effect on the resonances in Bi_2CuO_4 crystals grown from a melt. The microanalysis of chemical composition also has not revealed any changes after annealing.

Figure 4 shows the AFMR line width as a function of

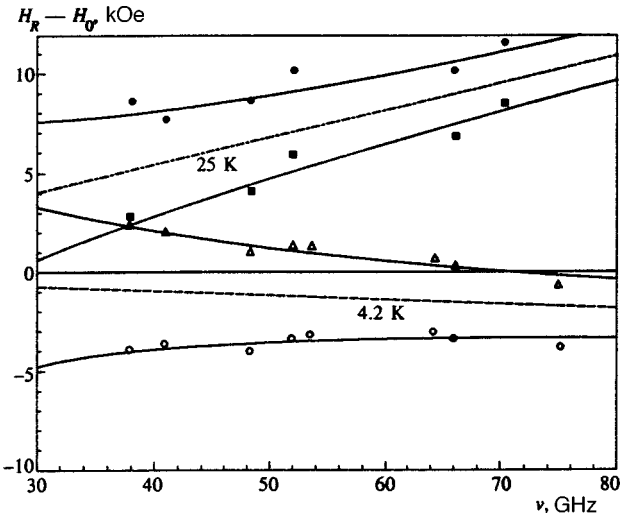


FIG. 2. AFMR field H_R measured with respect to the ESR field H_0 for DPPH as a function of microwave frequency. Open symbols correspond to temperature $T=4.2$ K, filled symbols to $T=23$ K; filled squares and open circles correspond to $H\|[110]$, and filled circles and triangles to $H\|[100]$. The samples were grown from a melt. The solid lines are calculations of the AFMR field as a function of frequency by Eq. (2) with parameters $H_A^{(4)}H_E=10$ and 6.8 kOe^2 and $a/A=-0.02$ and -0.1 at temperatures $T=4.2$ and 23 K, respectively. The dashed and dash-dotted lines show calculations at $\varphi=22.5°$, when the gap in the spectrum due to anisotropy in the plane perpendicular to the four-fold axis vanishes.

temperature for two nonannealed samples. One can see that the temperature-dependent parts of the AFMR line width are similar. The part of the line width independent of temperature is, most probably, due to an inhomogeneous line broadening and depends sensitively on the sample quality. The most plausible factor leading to the inhomogeneous broadening is variation in the alignment of the C^2 axis over the crystal volume. No correlation between the line width and annealing conditions has been detected.

Recordings of the derivative of the absorbed microwave power with respect to magnetic field (at $\nu=9.3$ GHz) as a function of the field are plotted in Fig. 5. Absorption lines recorded for both scanning directions have a rich fine structure, which manifests in the low-temperature range ($T<15$ K). This fine structure is well reproducible. The fine structure is seen in the range of magnetic field below the antiferromagnetic resonance at all static field orientations in the basal plane.

3.2. Discussion

The changes in the resonant absorption with temperature (Fig. 1) and the strong dependence of the resonant field H_R on the static field orientation provide strong evidence in favor of an antiferromagnetic resonance in Bi_2CuO_4. Another argument in favor of the interpretation of the resonant absorption in terms of uniform precession of the magnetic moment is the fine structure in absorption spectra in the field range below the main resonance (Fig. 5). In all probability, the recorded fine structure is due to spin-wave resonances in Bi_2CuO_4. The presence of resonances corresponding to large wave numbers allows us to rule out an interpretation ascribing the absorption line to impurities.

The presence of the gapless branch in the spectrum of magnetic excitations is in accordance with submillimeter wave and microwave measurements,[1,8] and with neutron diffraction experiments,[7] but contradicts other neutronographic measurements.[12]

All resonance properties of Bi_2CuO_4 can be easily described considering it as a two-sublattice antiferromagnet with an easy-plane magnetic anisotropy. A phenomenological theory of the antiferromagnetic resonance taking into account an easy-plane anisotropy was given by Turov.[13] The energy density of such an antiferromagnet is expressed as[13]

$$\mathscr{F}_m=A/2m^2+a/2\ m_z^2+b/2\ l_z^2+f/2\ l_x^2l_y^2-2M_0\ \mathbf{mH}. \tag{1}$$

Here M_0 is the saturation magnetization of one sublattice, **l** and **m** are the normalized vectors of antiferromagnetism and magnetization ($m^2+l^2=1$, $\mathbf{m}\cdot\mathbf{l}=0$). The first term on the right-hand side is the exchange energy, the second and third are responsible for the crystallographic magnetic anisotropy of a uniaxial crystal (in the case of an easy-plane anisotropy $b>0$). The fourth-order term takes into account the anisotropy in the basal plane perpendicular to the [001] axis. The last term describes the magnetic energy due to applied magnetic field.

The two AFMR frequencies corresponding to the acoustic and optic magnon branches are given by

$$\omega_1=\gamma[H^2(1+a/A)-H_EH_A^{(4)}\cos(4\varphi)]^{1/2}, \tag{2}$$

$$\omega_2=\gamma[H_EH_A]^{1/2}[1-H^2/H_E^2]^{1/2}, \tag{3}$$

where $\gamma=g_\perp\mu_B/\hbar=18.15\times10^9$ s^{-1} kOe^{-1} is the gyromagnetic ratio (according to Ref. 1 $g_\perp=2.04$), H is the static magnetic field, $H_E=A/2M_0$, $H_A=b/2M_0$, and $H_A^{(4)}=(\gamma/2M_0)f$. From the static susceptibility measured by Ochta et al.[1] for $H\perp C^{(4)}$, we derive $H_E=2M_0/\chi_{\perp c}=2000$ kOe; $H_A=12$ kOe can be derived from the width of the gap separating the high-frequency AFMR branch.[1] From our measurements, we derive $H_A^{(4)}$ and the ratio a/A. The solid line in Fig. 4 shows calculations of the frequency dependence of the AFMR field by Eq. (2) with $H_EH_A^{(4)}=10$ and 6.8 kOe^2 and $a/A=-0.02$ and -0.1 at temperatures $T=4.2$ and 23 K, respectively. The curves of $H_EH_A^{(4)}$ and a/A versus temperature are plotted in Fig. 3. The minus sign in front of a indicates the easy-axis anisotropy for the ferromagnetic vector **m,** whereas the antiferromagnetic vector **l** has an easy-plane anisotropy ($b>0$). Given a/A, one can determine the ratio between the static susceptibilities $\chi_\perp/\chi_\|=1+a/A$.[13] The subscripts indicate the magnetic field orientation with respect to the four-fold axis $C^{(4)}$. It is clear that, given the negative a for Bi_2CuO_4, $\chi_\perp$ should be about 10% smaller than $\chi_\|$ at $T=20$ K, which is in reasonable agreement with static magnetic measurements.[1,6] It is remarkable that the temperature dependence of the anisotropy constant a (Fig. 3) is strong in the temperature range of 10–20 K, which is far from T_N. The AFMR line also broadens considerably in this temperature interval (Fig. 4). All these effects may be related to a structural transition in Bi_2CuO_4, because the unit cell volume also increases abruptly in this region.[4]

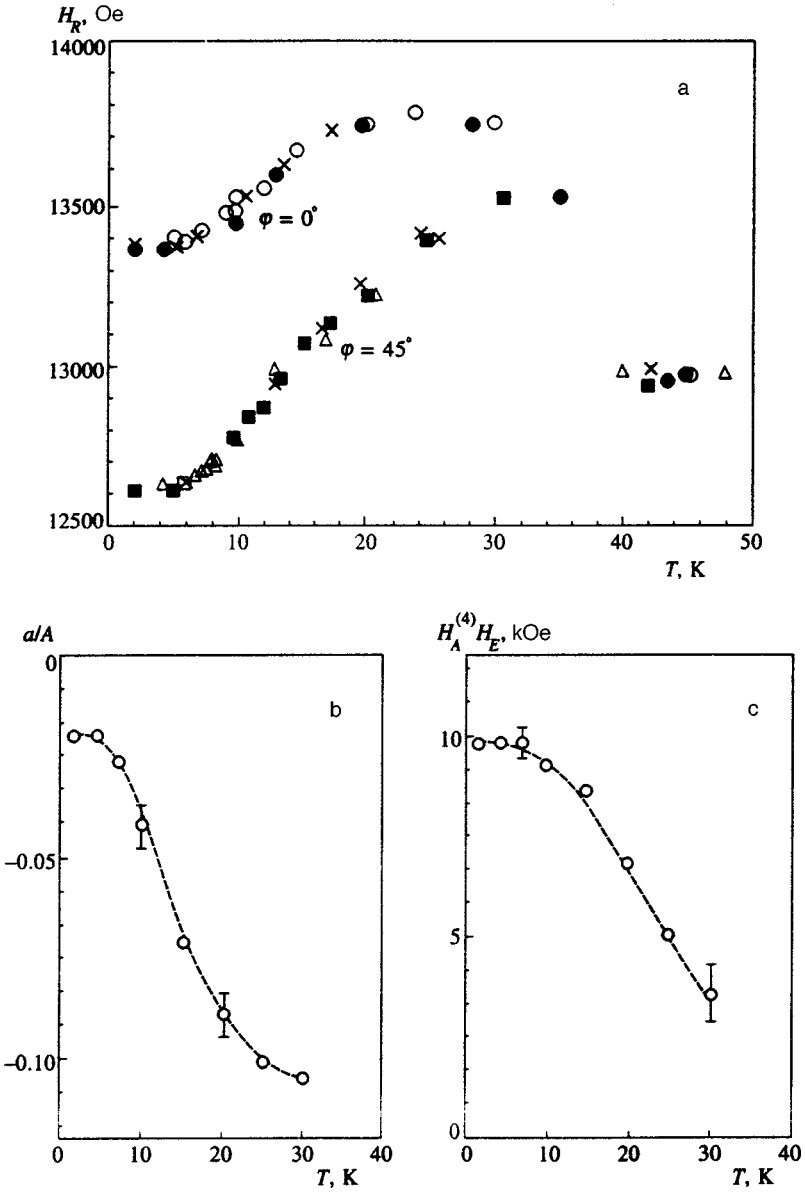


FIG. 3. (a) Resonant field versus temperature at microwave frequency $\nu = 36$ GHz. Measurements were performed for two orientations of the static field in the basal plane, $\varphi = 0$ and 45°. Different symbols show measurements of samples annealed in oxygen or nitrogen, and as-grown samples. Within the experimental uncertainty, curves of $H_R(T)$ for all samples are identical. The samples were grown from a melt. (b) and (c) Temperature dependences of a/A and $H_A^{(4)}H_E$ derived from measurements plotted in Fig. 3a.

In addition to the step on the curve of the AFMR line width versus temperature in the range of 10–15 K (Fig. 4), the width also grows rapidly as the temperature approaches the Néel temperature. This temperature-dependent part of the line width is largely due to the process of three-magnon relaxation. The solid curve in this graph shows calculations of the AFMR line width caused by the above mentioned relaxation process using the equation given in Ref. 15 and parameters of Bi_2CuO_4 given in this section.

The hysteretic behavior of the absorption line up to fields several times larger than spin-flop fields indicates the presence of a highly developed system of antiferromagnetic domains. Unlike the case of a ferromagnet, antiferromagnetic domains are equivalent in energy terms in fields higher than the spin-flop field, so even a slight pinning makes a domain wall stable up to the high fields. It is probable that the shift of the resonance field, step-like growth of the line width, hysteretic behavior of the absorption line, and the abrupt change in the cell volume in the temperature range of 10–15 K are caused by a sharp decrease in the mobility of domain

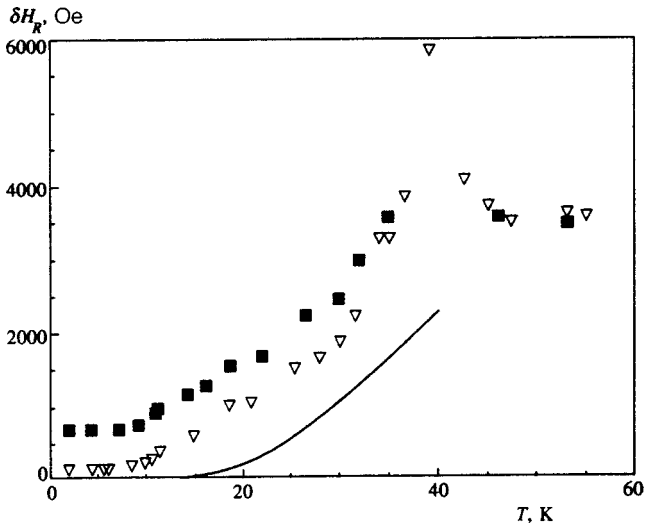


FIG. 4. AFMR line widths versus temperature at $\nu = 36$ GHz. Filled squares and triangles show measurements of two different as-grown samples. The solid curve shows calculations of line widths due to three-magnon relaxation based on formulas of Ref. 15. The samples were grown from a melt.

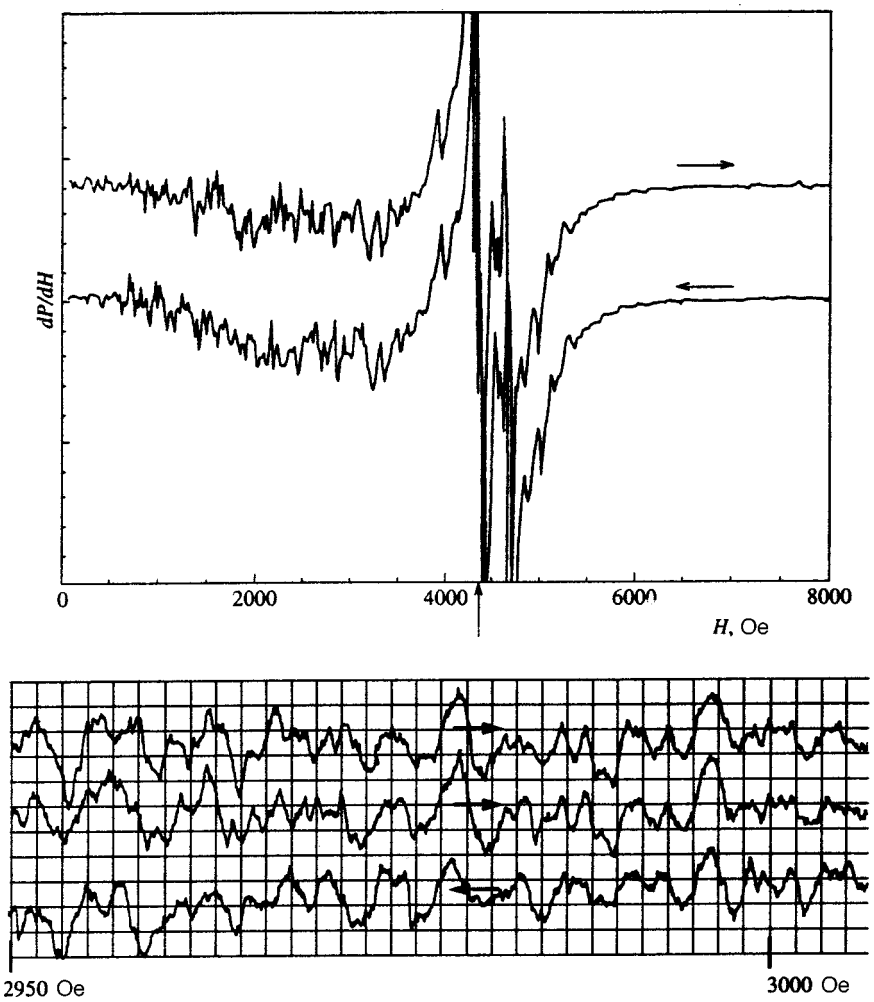


FIG. 5. Measurements of the derivative of the microwave power absorbed by the sample with respect to magnetic field as a function of applied magnetic field. $\nu=9.3$ GHz, $T=10$ K, $\varphi=0°$. The arrow indicates the AFMR position calculated by Eq. (2) with the parameters corresponding to the given temperature. The lower graph shows curves on the extended magnetic field scale.

walls. Information about the nature of antiferromagnetic domains in Bi_2CuO_4 may be derived from measurements of nuclear magnetic resonance and magneto-acoustic experiments.

3.3. Measurements of samples grown by the hydrothermal technique and discussion

In experiments with samples obtained by the hydrothermal technique, we have detected two lines of resonance absorption. Typical curves are given in Fig. 6. The field of resonance $R1$ strongly depends on the direction of the static field H, whereas the position of resonance $R2$ is independent of its orientation. As in the case of samples grown from a melt, the conditions of resonance $R1$ are controlled by the static field projection on the plane perpendicular to the fourfold axis. The position of the low-field resonance $R1$ strongly depends on temperature (Fig. 7). At the same time,

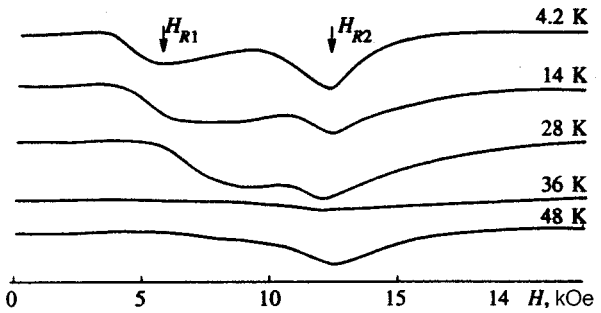


FIG. 6. Typical curves of microwave power transmitted through the cavity loaded with a Bi_2CuO_4 sample grown by the hydrothermal technique at $\nu=36$ GHz and different temperatures.

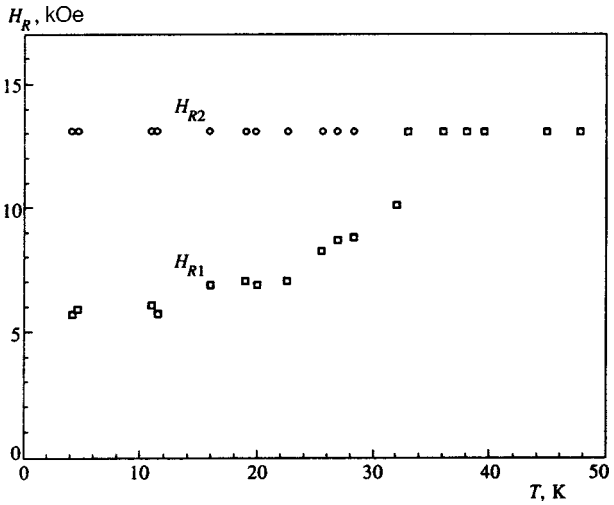


FIG. 7. Resonance fields H_{R1} and H_{R2} versus temperature at $\nu=36$ GHz. The samples were grown by the hydrothermal technique.

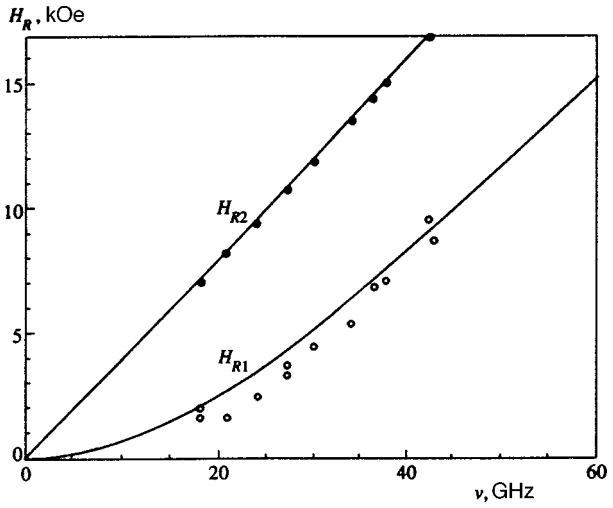


FIG. 8. Resonant fields H_{R1} and H_{R2} versus frequency at $T=4.2$ K. The samples were grown by the hydrothermal technique.

the line width and intensity in the range of 1.2–25 K are almost independent of the temperature. The amplitude of line $R2$ rapidly drops with temperature. The resonant fields corresponding to these lines versus frequency are plotted in Fig. 8. At frequencies higher than 45 GHz (i.e., in fields higher than 9–10 kOe) only line $R2$ can be seen in the spectra. It is highly probable that line $R2$ is due to paramagnetic defects in the sample, whereas line $R1$ is caused by excitations in the magnetically ordered crystal. If we assume that the small ferromagnetic moment is caused by the canting of magnetic sublattices due to the Dzyaloshinskii interaction,[9] the Dzyaloshinskii effective field H_D can be derived from measurements of the resonant field as a function of frequency.[13] The solid curve in Fig. 8 plots a fitting of the resonant field calculated by Eq. (2), where H^2 in the first term on the right-hand side is replaced by $H(H+H_D)$ and the rest of the constants are the same as in samples grown from a melt. The best fitting was obtained at $H_D=20$ kOe. This result is a factor of four larger than the value derived from static measurements of magnetization.[9]

Annealing of crystals grown by the hydrothermal method led to unexpected results. The absorption line $R1$ vanished after annealing in atmospheres of either oxygen or helium, whereas the position and intensity of line $R2$ did not change. The annealing was performed under the same conditions as in the case of the crystals grown by the flux method. Given the oxygen deficiency in the initial samples, we expected that annealing in an oxygen atmosphere should have led to the appearance of an AFMR line similar to that of samples grown from a melt, but it was not detected in our experiments. X-ray microanalysis showed that the crystal composition after the annealing in oxygen atmosphere was $Bi_{2.05}CuO_{3.91}$, i.e., the oxidation degree of cations increased after the annealing, but did not reach the value characteristic to the crystals grown from a melt. It is plausible that the conditions of this annealing are too mild to get rid of the oxygen deficiency in the crystal, but it is sufficient to release local elastic strains in the crystal or make the oxygen concentration uniform over the crystal volume. On the other hand, the difference between the properties of the two types of crystals can be attributed to different ratios between the contents of copper and bismuth.

The elimination of line $R1$ after annealing in oxygen or helium atmosphere under relatively mild conditions casts doubt on the conjecture about a relationship between the small ferromagnetic moment in Bi_2CuO_4 samples manufactured by the hydrothermal technique and canting of sublattices due to the Dzyaloshinskii–Moria interaction. It seems more probable that the resonant absorption in line $R1$ and the small ferromagnetic moment are due to the presence of domain walls in the samples.

The authors are grateful to L. A. Prozorova, A. I. Smirnov, and S. S. Sosin for fruitful discussions. The work was supported by the Russian Fund for Fundamental Research (Projects 96-02-16575 and 95-02-036-a).

*)E-mail: svistov@kapitza.ras.ru

[1] H. Ochta, K. Yoshida, T. Matsuya, T. Nanba, M. Motokawa, K. Yamada, Y. Endon, and S. Hosoya, J. Phys. Soc. Jpn. **61**, 2921 (1992).
[2] E. W. Ong, G. H. Kwei, R. A. Robinson, B. L. Ramakrishna, and R. B. von Dreele, Phys. Rev. B **42**, 4255 (1990).
[3] J. C. Boivin, D. Thomas, and G. Tridot, C. D. Acad. Sci. Paris **276**, 1105 (1973).
[4] J. L. Garcia-Muños, J. Rodriguez-Carvajal, F. Sapina, M. J. Sanchez, R. Ibañez, and D. Beltran-Porter, J. Phys. Condens. Matter **2**, 2205 (1990).
[5] R. Troc, J. Janicki, I. Filatow, P. Fischer, and A. Murasik, J. Phys. Condens. Matter **2**, 6989 (1990).
[6] K. Yamada, K. Takada, S. Hosoya, Y. Watanabe, Y. Endoh, N. Tomonaga, T. Suzuki, T. Ishigaki, T. Kamiyama, H. Asano, and F. Izumi, J. Phys. Soc. Jpn. **60**, 2406 (1991).
[7] M. Ain, G. Dhalenne, O. Guiselin, B. Hennion, and A. Revcolevschi, Phys. Rev. B **47**, 8167 (1992).
[8] A. I. Pankrats, G. A. Petrakovskii, and K. A. Sablina, Solid State Commun. **91**, 121 (1994).
[9] R. Szymczak, H. Szymczak, A. V. Zalessky, and A. A. Bush, J. Magn. Magn. Mater. **140–144**, 1573 (1995).
[10] G. A. Petrakovskii, K. A. Sablina, A. M. Vorotnikov, V. N. Vasiliev, A. I. Kruglik, A. D. Balaev, D. A. Velikanov, and N. I. Kiselev, Solid State Commun. **79**, 317 (1991).
[11] A. V. Zalessky, V. G. Krivenko, T. A. Khimich, N. E. Ainbinder, and A. A. Bush, J. Magn. Magn. Mater. **127**, 281 (1993).
[12] G. A. Petrakovskii, K. A. Sablina, V. V. Val'kov, and B. V. Fedoseev, A. Furer, P. Fisher, and B. Rossli, JETP Lett. **56**, 144 (1992).
[13] E. A. Turov, *Physical Properties of Magnetically Ordered Crystals* [in Russian], Izd. AN SSSR, Moscow (1963), pp. 61–77.
[14] J. P. Attfield, J. Phys. Condens. Matter **1**, 7045 (1989).
[15] V. G. Bar'yakhtar and V. L. Sobolev, Fiz. Tverd. Tela **15**, 2651 (1973) [Sov. Phys. Solid State **15**, 1764 (1973)].

Translation provided by the Russian Editorial office.

Electromagnetic waves in the nonlocal electrodynamics of an oblique Josephson junction of finite thickness

A. I. Lomtev*)

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine

(Submitted 18 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 2256–2262 (June 1998)

The new geometry for an oblique Josephson junction of finite thickness along the magnetic field is considered and a nonlocal integro-differential equation describing the dynamics of the phase difference is derived. The spectrum of low-amplitude electromagnetic excitations for lap-joint junctions is also studied. © *1998 American Institute of Physics.* [S1063-7761(98)02506-2]

1. So far the equations of nonlocal Josephson electrodynamics have been derived and studied in four cases: (1) a tunneling junction at the interface between two ultrathin superconducting films whose thickness is much smaller than the London length; (2) a tunnel junction between bulk superconductors whose thickness is much larger than the London length; (3) a tunnel junction between superconducting layers of finite thickness in the direction perpendicular to the magnetic field; and (4) a tunnel junction at a butt-joint of superconducting plates of finite thickness in the direction of the magnetic field.

For example, in Ref. 1–8 it was shown that nonlocal effects can be large even in junctions with a large thickness d ($d \gg \lambda$, where λ is the London penetration depth) in the direction of the magnetic field (in the direction parallel to the vortices), i.e., in situations previously studied in the local approximation. In the opposite limit of junctions involving thin films, with $d \ll \lambda$, nonlocality becomes a decisive factor. The corresponding equations were derived and studied in Refs. 9–12. A Josephson junction between two superconducting layers of finite thickness in the direction perpendicular to the magnetic field of the vortices was studied in Ref. 13.

Nevertheless, there are still many ways in which the theory can be improved. For one thing, it is necessary to go beyond the limits mentioned above, since experimenters often use junctions whose size in the direction in which the Josephson vortices are oriented is comparable to λ. Such a geometry is realized, e.g., in single-crystal Y–Ba–Cu–O chips with twins and in other ceramic materials with twin boundaries.

Such an approach was adopted in Ref. 14 for a butt-joint junction (the plane of the tunnel junction is perpendicular to the plate plane) for an arbitrary ratio d/λ. There it was found that the relationship between the jump in the phase of the order parameter at the junction and the current density always contains a nonlocal component, whose origin lies in the long-range nature of the field in free space, and that the amplitude of this component is only weakly dependent on the parameter λ/d (the dependence is almost linear) and its shape is independent of λ/d.

The present paper contains a derivation and preliminary analysis of the equations of the electrodynamics of an oblique Josephson junction in a plate with an arbitrary ratio d/λ and an arbitrary angle α between the junction plane and the normal to the plate plane. Such junctions may arise in two types of experiment. First, in granulated polycrystalline high-T_c ceramic materials there can be random weak links with different values of α, from a butt-joint junction ($\alpha \simeq 0$) to a lap-joint junction ($\alpha \simeq \pi/2$). Probably, the main fraction of such junctions consists of general-position junctions, i.e., oblique Josephson junctions with an arbitrary value of α. Second, experimental studies and practical applications ordinarily use lap-joint junctions with a large junction area. The method used in fabricating such junctions can also be employed to produce oblique Josephson junctions. For a lap-joint junction ($\alpha \to \pi/2$), not considered so far, we study, in the nondissipative limit and in the absence of a transport current and an external field, the spectrum of low-amplitude electromagnetic excitations propagating along the junction.

The problem posed above is reduced to that of an oblique Abrikosov vortex in a plane-parallel plate of finite thickness $2d$ in the direction of the magnetic field. This problem of calculating an oblique vortex current averaged over the plate thickness and normal to the junction plane is a new one. The results refer to a general-position junction with an arbitrary angle α. For $\alpha \to \pi/2$ is is possible to approximately integrate the Fourier transform of the integral term over the fast variable $k_y \tan\alpha$ and thus find the spectra of low-amplitude electromagnetic excitations for arbitrary values of the wave vector and the ratio d/λ. It is also possible in this case to find the explicit form of such spectra in the short- and long-wavelength ranges of the spectrum for $d/\lambda \gg 1$ and $d/\lambda \ll 1$.

2. The magnetic field of an arbitrary linear source (e.g., a vortex considered in the London approximation) in a superconducting plate satisfies the equation

$$\lambda^2 \Delta \mathbf{h} - \mathbf{h} = -\boldsymbol{\nu}(\mathbf{r}),$$

$$\boldsymbol{\nu}(\mathbf{r}) = \frac{\Phi_0}{2\pi} \operatorname{curl} \nabla\theta = \frac{\Phi_0}{2\pi} \int \delta(\mathbf{r} - \mathbf{R}(p)) d\mathbf{R}(p), \qquad (1)$$

where Φ_0 is the magnetic flux quantum, θ is the phase of the order parameter, and $\mathbf{R}(p)$ is the parametrically specified radius vector of the points of the vortex core. The core of the Josephson vortex spreads out along the weak-link surface S dividing the superconductor and is a two-dimensional object. Clearly, the corresponding source $\boldsymbol{\nu}(\mathbf{r})$, smeared over this surface, in Eq. (1) can be expressed by the formula

$$\boldsymbol{\nu}(\mathbf{r})=\frac{\Phi_0}{2\pi}\int \delta(\mathbf{r}-\mathbf{R}(a,b))\left(\frac{\partial\varphi}{\partial\mathbf{R}}\times d\mathbf{S}(a,b)\right), \tag{2}$$

where a and b are the arguments of the parametric representation of S, $\mathbf{R}(a,b)$ is the radius vector of the points of S, $d\mathbf{S}(a,b)$ is a vector area element of S, and φ is the phase difference between the edges of the junction. Since we have div $\boldsymbol{\nu}=0$, the arbitrary source of magnetic field can either be represented by a continuous sum over linear cores (and $\mathbf{h}$ can be represented as a linear combination of the Abrikosov vortices) or, conversely, be treated as a three-dimensional vector field. We direct the z axis transverse to and the x axis along the junction. Then for a general junction oriented at an arbitrary angle α and a magnetic field directed at the angle $\pi/2-\alpha$ Eq. (2) yields

$$\nu_x(\mathbf{r})=0,\quad \nu_y(\mathbf{r})=\frac{\Phi_0}{2\pi}\tan\alpha\,\delta(y-z\tan\alpha)\frac{\partial\varphi(x)}{\partial x},$$

$$\nu_z(\mathbf{r})=\frac{\Phi_0}{2\pi}\delta(y-z\tan\alpha)\frac{\partial\varphi(x)}{\partial x}, \tag{3}$$

where $\varphi(\mathbf{r})$ is the jump in the phase of the order parameter at the sides of the junction,

$$\varphi(\mathbf{r})=\theta(x,y\cos\alpha-z\sin\alpha=+0)-\theta(x,y\cos\alpha-z\sin\alpha=-0), \tag{4}$$

which, in view of (3) and the fact that div $\boldsymbol{\nu}=0$, is independent of the coordinates y and z.

The phase jump and hence the source can be found by solving the complete nonlinear equation for the junction. Note that even without solving this equation, we see that due to the linearity of Eq. (1) we can write the magnetic field as $\mathbf{h}=\mathbf{H}_m+\mathbf{H}$, where $\mathbf{H}_m$ is the seed Meissner field generated by the given transport supercurrent and the external magnetic field and is determined by solving the homogeneous analog of Eq. (1) as if there were no weak link and the superconductor were continuous, and the field $\mathbf{H}$ is generated by the source ($\mathbf{H}$ vanishes at $\boldsymbol{\nu}=0$.) Taking a two-dimensional Fourier transform in the plane xy of a plate of thickness $2d$, $|z|<d$, we find that

$$\mathbf{H}=\mathbf{H}_0+\mathbf{a}\exp(\kappa z)+\mathbf{b}\exp(-\kappa z), \tag{5}$$

$$\mathbf{H}_0=-\int_{-d}^{d}\frac{\sinh\kappa|z-z'|}{2\kappa\lambda^2}\boldsymbol{\nu}(k,z')dz',$$

$$\kappa=(\lambda^{-2}+k^2)^{1/2},\quad k^2=k_x^2+k_y^2, \tag{6}$$

where $\mathbf{k}$ is a two-dimensional wave vector. The vector coefficients $\mathbf{a}$ and $\mathbf{b}$ are determined, first, by the condition that the field be divergence-free,

$$\frac{\partial H_z}{\partial z}+i\mathbf{k}\cdot\mathbf{H}_\parallel=0, \tag{7}$$

and continuous at the edges of the plate and, second, by the potential nature of the tangential component of the magnetic field at the boundary of the superconductor (which means that the component of the current normal to the surface vanishes) and by the potential nature of all three components of the field in free space. The latter condition yields

$$\frac{\mathbf{H}_\parallel}{H_z|_{z=\pm d}}=\mp\frac{i\mathbf{k}}{|\mathbf{k}|}, \tag{8}$$

which accounts for the effect of free space on the field and current distributions in the superconductor. Here and below the subscript $\parallel$ denotes the x- and y-projections of vectors.

One consequence of (7) and (8) is

$$\left.\frac{\partial H_z}{\partial z}\right|_{z=\pm d}=\mp H_z|_{z=\pm d}, \tag{9}$$

which determines the coefficients a_z and b_z in terms of the source field:

$$a_z=\Delta^{-1}\left\{\left(\frac{\partial H_{0z}}{\partial z}-kH_{0z}\right)\Big|_{z=-d}(\kappa-k)\exp(-\kappa d)-\left(\frac{\partial H_{0z}}{\partial z}+kH_{0z}\right)\Big|_{z=d}(\kappa+k)\exp(\kappa d)\right\},$$

$$b_z=\Delta^{-1}\left\{\left(\frac{\partial H_{0z}}{\partial z}-kH_{0z}\right)\Big|_{z=-d}(\kappa+k)\exp(\kappa d)-\left(\frac{\partial H_{0z}}{\partial z}+kH_{0z}\right)\Big|_{z=d}(\kappa-k)\exp(-\kappa d)\right\}, \tag{10}$$

where

$$\Delta=4(\kappa\sinh\kappa d+k\cosh\kappa d)(\kappa\cosh\kappa d+k\sinh\kappa d),$$

$$k=|\mathbf{k}|.$$

By combining (8) and (10) we can write the coefficients $\mathbf{a}_\parallel$ and $\mathbf{b}_\parallel$ in terms of the source field:

$$\mathbf{a}_\parallel=\frac{1}{2\sinh(2\kappa d)}\left\{-i\frac{\mathbf{k}}{k}[H_z|_{z=d}\exp(\kappa d)+H_z|_{z=-d}\exp(-\kappa d)]-\mathbf{H}_{0\parallel}|_{z=d}\times\exp(\kappa d)+\mathbf{H}_{0\parallel}|_{z=-d}\exp(-\kappa d)\right\},$$

$$\mathbf{b}_\parallel=\frac{1}{2\sinh(2\kappa d)}\left\{i\frac{\mathbf{k}}{k}[H_z|_{z=d}\times\exp(-\kappa d)+H_z|_{z=-d}\exp(\kappa d)]+\mathbf{H}_{0\parallel}|_{z=d}\exp(-\kappa d)-\mathbf{H}_{0\parallel}|_{z=-d}\exp(\kappa d)\right\}. \tag{11}$$

Formulas (5)–(11) make it possible to find the field and supercurrent inside the plate for an arbitrary source.

Plugging (3) into (6), combining the result with (5), and going over to the coordinate representation, from (5)–(11)

we arrive at the following expression for the density of the current flowing through the junction in the direction normal to the junction plane and averaged over the plate thickness:

$$j(x)=J_m(x)+J(x),$$

where J_m is the seed Meissner current determined by the field H_m, and

$$J(x)=\frac{c\Phi_0}{16\pi^3\lambda^2}\frac{\partial}{\partial x}\int Q(x-x')\frac{\partial}{\partial x'}\varphi(x')dx' \tag{12}$$

is the current due to the source. Here the nonlocal source–current kernel is

$$Q(x)=K_0\left(\frac{|x|}{\lambda}\right)+Q_1(x), \tag{13}$$

where K_0 is a modified Bessel function of order zero,

$$Q_1(x)=\int\frac{dk_x dk_y}{(2\pi)^2}\exp(ik_x x)Q_1(k_x,k_y), \tag{14}$$

and the Fourier transform of the kernel $Q_1(k_x,k_y)$ is

$$\begin{aligned}
Q_1(k_x,k_y)&=\frac{1}{\kappa\cosh\kappa d+k\sinh\kappa d}\,\frac{1}{\kappa\sinh\kappa d+k\cosh\kappa d}\\
&\times\left\{\cos\alpha\left\{\frac{\kappa}{2dk(\kappa^2+\tilde{k}_y^2)}[k\cosh(2\kappa d)+\kappa\sinh(2\kappa d)\right.\right.\\
&-k\cos(2\tilde{k}_y d)+\tilde{k}_y\sin(2\tilde{k}_y d)]\\
&-\frac{1}{\kappa d(\kappa^2+\tilde{k}_y^2)^2}[2k\kappa^2\tilde{k}_y\cos\tilde{k}_y d\sin\tilde{k}_y d+\kappa^2\\
&\times(k^2-\tilde{k}_y^2)(\cosh^2\kappa d\sin^2\tilde{k}_y d+\sinh^2\kappa d\cos^2\tilde{k}_y d)\\
&\left.+k\kappa(\kappa^2-\tilde{k}_y^2)\cosh\kappa d\sinh\kappa d]\right\}\\
&+\frac{\kappa\sin\alpha}{d(\kappa^2+\tilde{k}_y^2)^2\sinh\kappa d\cosh\kappa d}\left\{-\tan\alpha[\cosh^2\kappa d\right.\\
&\times\sin^2\tilde{k}_y d+\sinh^2\kappa d\cos^2\tilde{k}_y d](\kappa\cosh\kappa d\\
&+k\sinh\kappa d)(\kappa\sinh\kappa d+k\cosh\kappa d)\\
&+\frac{k_y}{k}[(\kappa^2-\tilde{k}_y^2)\sinh\kappa d\cos\kappa d\sin\tilde{k}_y d\\
&\times\cos\tilde{k}_y d+\kappa\tilde{k}_y(\sinh^4\kappa d\cos^2\tilde{k}_y d+\cosh^4\kappa d\\
&\left.\left.\times\sin^2\tilde{k}_y d-\sinh^2\kappa d\cosh^2\kappa d)]\right\}\right\},
\end{aligned} \tag{15}$$

with $\tilde{k}_y=k_y\tan\alpha$, $0\leqslant\alpha\leqslant\pi/2$. If in (13)–(15) we set α equal to zero, we arrive at the results of Ref. 14 for a butt-joint Josephson junction. In this case, the first term in (13) corresponds to the limit of two bulk superconductors of thickness $d\gg\lambda$ and is the kernel of the integral term in the equation first obtained in Ref. 1 and used in Refs. 2–8. In the opposite limit of ultrathin films of thickness $d\ll\lambda$, the sum of both terms gives the kernel of the integral term in the equation first studied in Refs. 9, 10, and 11 and equal to

$$Q(x)=\frac{\lambda_{\rm eff}}{\pi}\int_0^\infty\frac{dk}{1+2k\lambda_{\rm eff}}J_0(kx), \tag{16}$$

where J_0 is the Bessel function of order zero, and $\lambda_{\rm eff}=\lambda^2/2d$ is the Peierls penetration depth.

3. A closed equation for the phase difference at the junction can be obtained, as usual, by equating $j(x)$ to the sum of the Josephson supercurrent, the normal current, and the capacitive displacement current, all regarded as internal characteristics of the junctions, and in standard notation has the form

$$\begin{aligned}
&\sin\varphi+\frac{\beta}{\omega_J^2}\frac{\partial\varphi}{\partial t}+\omega_J^{-2}\frac{\partial^2\varphi}{\partial t^2}\\
&\qquad=\frac{J_m(x)}{j_c}+\frac{\lambda_J^2}{\pi\lambda}\frac{\partial}{\partial x}\int Q(x-x')\frac{\partial\varphi(x')}{\partial x'}dx',
\end{aligned} \tag{17}$$

where j_c is the Josephson current density, λ_J and ω_J are the Josephson length and frequency, and β is the dissipation parameter.

The integral kernel $Q(x)$ describes excitations in a Josephson junction in a unified manner, both in a thin film and in a sample that is thick in the direction of the magnetic field. In the general intermediate case with an arbitrary angle α, it is the sum of a well-localized term and a strongly nonlocal nonintegrable term (the second term on the right-hand side of Eq. (13)), which originates in the slowly decreasing tangential component of the magnetic field in free space at the surface of the plate. Here, for a near-butt-joint junction (when the value of the angle α is zero or very small) and a near-lap-joint junction (when the value of the angle α is close to $\pi/2$), there are two physically distinct asymptotic expressions for the kernel $Q_1(x)$ at large values of the argument x. For instance, when α is close to zero, the vortex magnetic flux is almost perpendicular to the plate plane and the magnetic fields on either side of the plate have opposite signs. This leads to a power dependence for $Q_1(x)$ at large distances (see Ref. 14):

$$Q_1(x)\sim\frac{\lambda^2}{d|x|}. \tag{18}$$

On the other hand, for α close to $\pi/2$, the vortex magnetic flux is almost parallel to the plane of the plate and the magnetic field in free space above and below of the plate is almost perfectly balanced. This lowers the degree of nonlocality of $Q_1(x)$, which according to (13)–(15) is characterized by an exponential asymptotic behavior at large x:

$$Q_1(x)\propto\frac{\lambda}{d}K_0\left(\frac{|x|}{d\tan\alpha}\right), \tag{19}$$

where the decay length is equal to the effective thickness of the junction $d\tan\alpha\gg\lambda$.

For general junctions, i.e., oblique junctions with an arbitrary angle α, the difference at large distances is in the nature of the nonlocal behavior of $Q_1(x)$ at large values of x, which according to Eqs. (14) and (15) can be calculated numerically, and approximately $Q_1(x)$ has the following asymptotic behavior:

$$Q_1(x) \propto A \cos\alpha \frac{\lambda^2}{d|x|} + B \sin\alpha \frac{\lambda}{d} K_0\left(\frac{|x|}{d \tan\alpha}\right), \qquad (20)$$

where A and B are constant factors.

4. Let us now examine the spectrum of low-amplitude electromagnetic excitations

$$\varphi(x,t) = \varphi_0 \exp[i(qx - \omega t)], \quad |\varphi_0| \ll 1, \qquad (21)$$

for a lap-joint junction, with the excitations propagating along the Josephson junction.

In the nondissipative limit and in the absence of a transport current and an external field, Eq. (17) yields the following dispersion relation

$$\omega = \omega_J \left[1 + \frac{\lambda_J^2 q^2}{(1+\lambda^2 q^2)^{1/2}} + \frac{\lambda_J^2 q^2}{\pi\lambda} F(q)\right]^{1/2}, \qquad (22)$$

where $F(q) \equiv Q_1(q)$ and, according to (15), is given by the following expression:

$$Q_1(k_x) = \cot\alpha \int \frac{dk_y}{2\pi} Q_1(k_x, k_y; \cot\alpha k_y)$$
$$\approx Q_{1(0)}(k_x) + \cot^2\alpha Q_{1(2)}(k_x) + \cdots, \qquad (23)$$

which amounts to the sum of the first two terms in the expansion of $Q_1(k_x)$ in the small parameter $\cot^2\alpha \ll 1$. In (23),

$$Q_{1(0)}(k_x) = -\frac{1}{4\kappa^2 d \sinh(2\kappa d)} \times [\cosh(2\kappa d) - (1+2\kappa d)\exp(-2\kappa d)],$$

$$Q_{1(2)}(k_x) = \frac{1}{\kappa\cosh\kappa d + k\sinh\kappa d} \frac{1}{\kappa\sinh\kappa d + k\cosh\kappa d} \times \left\{ \frac{\sinh(2\kappa d)}{4d}\left(\frac{3}{2} + \frac{\kappa}{k}\right) + \frac{\exp(-2\kappa d)}{4\kappa k} \times [k(\lambda^{-2} + 2k^2) + 2\kappa\lambda^{-2}] + \frac{\kappa}{4kd\sinh(2\kappa d)} - \frac{k^2}{8d\kappa^2}\sinh^2\kappa d[1 + \exp(-2\kappa d)] + \frac{\kappa\coth(2\kappa d)\exp(-2\kappa d)}{4kd}(2\kappa d - 1)\right\}, \qquad (24)$$

where $k \equiv k_x$, and $\kappa = (\lambda^{-2} + k_x^2)^{1/2}$. Equations (22)–(24) determine the spectrum of low-amplitude electromagnetic excitations for an arbitrary wave vector $q \equiv k_x$ and an arbitrary ratio d/λ. The asymptotic expressions in the short- and long-wavelength ranges are

$$\omega(q \to \infty) = \omega_J \left[1 - \frac{\lambda_J^2}{4\pi\lambda d} + \frac{\lambda_J^2 q^2}{(1+\lambda^2 q^2)^{1/2}}\right]^{1/2}, \qquad (25)$$

$$\omega(q \to 0) = \omega_J \left[1 + \frac{\lambda_J^2}{\pi d \tan^2\alpha} q + \left(1 - \frac{3}{4\pi}\right)\lambda_J^2 q^2\right]^{1/2} \qquad (26)$$

for $d \ll \lambda$ and

$$\omega(q \to 0) = \omega_J \left[1 + \frac{\lambda_J^2}{2\pi d \tan^2\alpha} q + \lambda_J^2 q^2\right]^{1/2} \qquad (27)$$

for $d \gg \lambda$.

The author would like to express his gratitude to Yu. E. Kuzovlev for suggesting the idea of studying an oblique Josephson junction and for stimulating discussions and to Yu. V. Medvedev for his interest and support.

*)E-mail: medvedev@host.dipt.donetsk.ua

[1] Yu. M. Aliev, V. P. Silin, and S. A. Uryupin, Sverkhprovodimost': Fiz., Khim., Tekhn. **5**, 228 (1992) [Supercond., Phys. Chem. Technol. **5**, 230 (1992)].
[2] A. Gurevich, Phys. Rev. B **46**, 3187, 14 329 (1992).
[3] Yu. M. Aliev, V. P. Silin, and S. A. Uryupin, JETP Lett. **57**, 193 (1993).
[4] Yu. M. Aliev and V. P. Silin, Zh. Éksp. Teor. Fiz. **104**, 2526 (1993) [JETP **77**, 142 (1993)].
[5] Yu. M. Aliev and V. P. Silin, Phys. Lett. A **177**, 259 (1993).
[6] V. P. Silin, JETP Lett. **58**, 701 (1993).
[7] G. L. Alfimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **106**, 671 (1994) [JETP **79**, 369 (1994)].
[8] V. P. Silin, JETP Lett. **60**, 460 (1994).
[9] Yu. M. Ivanchenko and T. K. Soboleva, JETP Lett. **51**, 114 (1990).
[10] Yu. M. Ivanchenko and T. K. Soboleva, Phys. Lett. A **147**, 65 (1990).
[11] Yu. M. Ivanchenko and T. K. Soboleva, Fiz. Tverd. Tela (Leningrad) **32**, 2029 (1990) [Sov. Phys. Solid State **32**, 1181 (1990)].
[12] R. G. Mints and I. B. Shapiro, Phys. Rev. B **51**, 3054 (1995).
[13] I. O. Kulik and I. K. Yanson, *The Josephson Effect in Superconducting Tunnelling Structures*, Israel Program for Scientific Translation, Jerusalem (1972) [Russ. orig. Nauka, Moscow (1970)].
[14] Yu. E. Kuzovlev and A. I. Lomtev, Zh. Éksp. Teor. Fiz. **111**, 1803 (1997) [JETP **84**, 986 (1997)]

Translated by Eugene Yankovsky