Identification of extragalactic cosmic-ray sources using data from various detection facilities

A. V. Uryson*)

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 6 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1121–1130 (October 1999)

Showers with energies $E>3.2\times10^{19}$ eV and $E\geqslant10^{20}$ eV detected at the AGASA (Akeno, Japan), Haverah Park, and Yakutsk arrays are investigated. The question of how the identification of sources depends on the error in determining the shower arrival directions is analyzed. Confirmation is obtained for the conclusion in the author's earlier work, that the principal sources of shower-driven particles are Seyfert galaxies with red shifts $z\leqslant0.0092$, which are weak emitters in the x-ray and radio ranges. © *1999 American Institute of Physics.* [S1063-7761(99)00110-9]

1. INTRODUCTION

The origin of cosmic rays with energies $E>4\times10^{19}$ is an object of active study. It follows from an analysis of extensive air showers generated by cosmic rays that particles having such energies are very likely of extragalactic origin[1–4] and come to us from distances of at most ~50 Mparsec (Refs. 4 and 5). [Particles in the form of nuclear fragments with $E\approx(2-3)\times10^{20}$ eV can arrive from distances of ~100 Mparsec (Ref. 6)]. If this is the case, the spectrum of cosmic rays in the range $E\geqslant10^{20}$ eV does not have the blackbody cutoff predicted in Refs. 7 and 8. Indeed, particles with energies $E\geqslant10^{20}$ eV have been detected at various ground stations: Sydney,[9] Yakutsk,[10] Haverah Park,[11] Fly's Eye,[12] and the Akeno Giant Air Shower Array (AGASA).[13] The spectra obtained at these sites are discussed in several papers.[12–15] In previous papers[16,17] we have analyzed the measurement data and submitted arguments arguing that the spectrum does not have a blackbody cutoff.

Given the coordinates of the particle arrival directions and the range of distances up to ~100 Mparsec, we can attempt to identify the cosmic-ray sources.

The sources discussed in the literature are divided into three groups. First of all, the sources can be various astrophysical objects: pulsars, nuclei of active galaxies,[1] hot spots of powerful radio galaxies and quasars,[5] lacertides,[18] and interacting galaxies.[19] Second, ultrahigh energy particles can be generated by cosmic necklaces.[20] Third, they can be generated in the decays of metastable superheavy particles of cold dark matter, which accumulate in the galactic halo.[21] (The possibility of particle acceleration in gamma-ray bursts has also been raised.[22]) Sources can be identified directly only if they belong to the first group. In the second case the incidence of any objects in the region around the particle arrival directions will be random. In the third case the main stream of particles will be from the galactic halo, and a weak stream from the Virgo cluster is possible.[21]

In previous papers[23–25] we have identified cosmic-ray sources from an analysis of 17 AGASA showers with energies $3.2\times10^{19}<E\leqslant2\times10^{20}$ eV, for which the coordinates of the arrival directions have been determined with an rms error $\sigma_{sh}\leqslant3°$ (Refs. 4 and 26).[1)] We have adopted the first model (astrophysical objects as sources) and have looked for possible sources in x-ray pulsars (as the most powerful), Seyfert galaxies, lacertides, and radio galaxies. The field of search for the sources was determined entirely by the error in the shower arrival direction in equatorial coordinates $(\Delta\alpha,\Delta\sigma)$ and was equal to $3\sigma_{sh}=3(\Delta\alpha,\Delta\delta)$. This magnitude of the search field is dictated by two factors. First, we know from statistics[27] that the probability of a proton existing in the rms error field is only 66%, but the same probability in the triple-error field is 99.8%. Second, the error in determining the optical coordinates of astrophysical objects can be disregarded, because it is no more than a few seconds, while the error in the shower coordinates is a matter of degrees. [We have assumed that intergalactic magnetic fields are weak, $B\leqslant8.7\times10^{-10}$ G (Ref. 25), and when a proton is deflected by them, it still remains in the region $3(\Delta\alpha,\Delta\delta)\leqslant9°$.] Calculating the probabilities of the random incidence of possible sources in the $3\sigma_{sh}$ field around the shower axis, we have found that this probability is low, $P>3\sigma$, only for Seyfert galaxies with red shifts $z\leqslant0.0092$, i.e., at a distance of no more than 40 Mparsec from Earth, if the Hubble constant is $H=75$ km/s·Mparsec. The radio waves and x-rays from these galaxies are weak. (Here σ is the spread of a Gaussian distribution in error theory.)

In the present study, in addition to these 17 showers, we consider showers with $E>4\times10^{19}$ eV recorded by the Yakutsk detector array (their errors are calculated in this paper) and showers with $E\geqslant10^{20}$ eV: two recorded by the Haverah Park array,[11] and one by the Fly's Eye array.[12]

2. SHOWERS WITH ENERGIES $3.2\times10^{19}<E<10^{20}$ eV

2.1. Yakutsk showers

Twelve showers with energies $E>4\times10^{19}$ eV have been recorded at the Yakutsk array.[10] The errors in their arrival directions $\Delta\alpha$, $\Delta\delta$ are calculated as follows. In the measurements of Refs. 10 and 28 the horizontal coordinates of the shower axis θ, φ and the errors $\Delta\theta$ and the solid angle $\Delta\Omega$

TABLE I. Arrival directions of Yakutsk showers in equatorial coordinates (Ref. 10), calculated errors $\Delta\alpha$, $\Delta\delta$, and Seyfert galaxies with $z \leqslant 0.0153$ from Refs. 29 and 30 in the shower search field. For showers 3, 4, 6, and 11 the axes have $|b| \leqslant 10°$ and lie well within the ''zones of avoidance'' of the galaxies.

Shower No.	Coordinates		Errors		Seyfert galaxies		
	$\alpha°$	$\delta°$	$\Delta\alpha°$	$\Delta\delta°$	α δ	z	Catalog[a]
1	163.7	52.9	3.0−3.2	1.9	11^h19^m 59.3°	0.0058	[29]
					1129+533	0.0036	[29, 30]
2	270.5	67.6	8.4−8.8	1.4−1.6	1822+665	0.0153	[29]
3	297.8	33.5	2.0−2.1	2.1	-		
4	342.9	65.8	7.7−8.1	1.4−1.6	-		
5	184.0	47.0	3.4−3.6	2.6−2.7	1129+533	0.0036	[29, 30]
					1153+554	0.0036	[29]
					1155+536	0.0038	[29, 30]
					1155+557	0.0041	[29]
					1200+448	0.0023	[29, 30]
					1203+477	0.0019	[29]
					1203+529	0.003	[30]
					1204+433	0.0028	[29]
					1205+434	0.0034	[29, 30]
					1216+475	0.0020	[29, 30]
					1248+413	0.0011	[29]
6	335.2	51.0	3.9−4.3	2.4−2.6		-	
7	118.1	57.0	5.1−5.6	2.0−2.2	0645+609	0.0069	[29, 30]
					0840+503	0.0112	[29, 30]
					0849+515	0.0025	[29, 30]
					0851+589	0.0032	[29]
8	235.4	79.8	10.7−16.9	2.6−3.1	1634+783	0.0046	[29, 30]
9	69.0	74.9	7.5−8.9	1.6−1.9	0609+710	0.0141	[29, 30]
10	92.5	374.0	5.7−6.8	2.1−2.2	0609+710	0.0141	[29, 30]
11	314.8	57.8	5.4−5.9	2.1−2.3		-	
12	21.3	45.7	2.1	3.4	0106+354	0.0006	[29]

[a]Catalog Ref. Nos. are shown in brackets.

are determined. The error of φ is $\Delta\varphi = d\Omega/(\sin\theta\cdot\Delta\theta)$. From the relations between the horizontal and equatorial coordinates (see, e.g., Ref. 2)

$$\sin\delta = \sin\varphi_0\cos\theta - \cos\varphi_0\sin\theta\cos\varphi,$$

$$\alpha = t_l - t, \quad \sin t = \sin\theta\sin\varphi/\cos\delta,$$

$$\cos t = (\cos\varphi_0\cos\theta + \sin\varphi_0\sin\theta\cos\varphi)/\cos\delta,$$

where $\varphi_0 = 61.7°$ N is the geographic latitude of the Yakutsk array, t_l is the local sidereal time, t is the universal time, and the longitude of the Yakutsk site is 129.4° E, we calculate α and δ for (θ,φ), $(\theta+\Delta\theta,\varphi)$, $(\theta-\Delta\theta,\varphi)$, $(\theta,\varphi+\Delta\varphi)$, and $(\theta,\varphi-\Delta\varphi)$. We then find the differences

$$\Delta\delta(\theta_+) = \delta(\theta,\varphi) - \delta(\theta+\Delta\theta,\varphi),$$

$$\Delta\delta(\theta_-) = \delta(\theta,\varphi) - \delta(\theta-\Delta\theta,\varphi),$$

$$\Delta\delta(\varphi_+) = \delta(\theta,\varphi) - \delta(\theta,\varphi+\Delta\varphi),$$

$$\Delta\delta(\varphi_-) = \delta(\theta,\varphi) - \delta(\theta,\varphi-\Delta\varphi),$$

along with the same differences $\Delta\alpha$. The differences $\Delta\alpha$ and $\Delta\delta$ are then the errors of α and δ due to the errors of measurements of θ and φ. (The calculation of the errors from the differences is analogous to the formal approach to the calculation of the corresponding partial derivatives.[27])

Next we find the scatter of the errors of declination and right ascension, combining the errors as follows: From each pair of calculated errors $\Delta\delta(\theta_+)$, $\Delta\delta(\theta_-)$ and $\Delta\delta(\varphi_+)$, $\Delta\delta(\varphi_-)$ we take the maximum values $\Delta\delta(\theta)_{\max}$ and $\Delta\delta(\varphi)_{\max}$ and determine the maximum declination error

$$\Delta\delta_{\max} = [\Delta\delta^2(\theta)_{\max} + \Delta\delta^2(\varphi)_{\max}]^{1/2},$$

and from the minimum values $\Delta\delta(\theta)_{\min}$ and $\Delta\delta(\varphi)_{\min}$ we find the minimum declination error

$$\Delta\delta_{\min} = [\Delta\delta^2(\theta)_{\min} + \Delta\delta^2(\varphi)_{\min}]^{1/2}.$$

In exactly the same way we obtain

$$\Delta\alpha_{\max} = [\Delta\alpha^2(\theta)_{\max} + \Delta\alpha^2(\varphi)_{\max}]^{1/2},$$

$$\Delta\alpha_{\min} = [\Delta\alpha^2(\theta)_{\min} + \Delta\alpha^2(\varphi)_{\min}]^{1/2}.$$

The arrival directions of the showers with the calculated errors and Seyfert galaxies with $z \leqslant 0.0153$ in their search field are summarized in Table I. (The galaxies emit weak streams in the radio and x-ray ranges.) The errors $\Delta\alpha$ and $\Delta\delta$ in the arrival directions of the Yakutsk showers exceed 3°, as a rule, and differ severalfold from one shower to the next.

2.2. Influence of errors in the shower arrival direction on source identification

For our analysis we have chosen AGASA showers[4,26] with $E < 10^{20}$ eV, for which the error in the arrival coordinates is $\sigma_{\rm sh} \leqslant 3°$. Their coordinates and Seyfert galaxies with $z \leqslant 0.0092$ in the search field ($3\Delta\alpha$, $3\Delta\delta$) are shown in Table II. Three of the showers in Table II, Nos. 2–4, for which no such galaxy was found in the search field, arrived from parts

TABLE II. Arrival directions of AGASA showers with $3.2\times10^{19}<E<10^{20}$ eV (Refs. 4 and 26), group assignments, and Seyfert galaxies with $z\leqslant0.0174$ in their search field. Our shower numbering system is used. In the search for galaxies it is assumed that $(\Delta\alpha,\Delta\delta)\leqslant3°$.

Shower No.	Coordinates		Refs.	Group	Galaxy			Refs.
	α	δ			α	δ	z	
1	01^h09^m	20°	4	(1,2)	01^h40^m	13.3°	0.003	30
					0124+133		0.0174	29, 30
2	01 42	71	26	(1)	-		-	-
3	03 30	70	26	(1)	-		-	-
4	05 20	20	26	(1)	-		-	-
5	11 10	24	26	(1,2)	1137+321		0.0092	29, 30
					1137+172		0.0101	29, 30
6	11 12	57.8	4	(1,2)	1122+546		0.0036	29, 30
7	11 27	57.3	4	(1,2)	1119+593		0.0058	29
8	13 25	16	26	(1,2)	1254+219		0.0013	29, 30
					1304+133		0.0091	29
9	13 40	35	26	(1,2)	1308+373		0.0036	29, 30
					1311+368		0.0032	29, 30
					1313+422		0.0020	29
					1351+337		0.0079	29, 30
					1353+407		0.0089	29
10	14 00	50	26	(1,2)	1327+474		0.0018	29, 30
					1327+474		0.0022	29
					1403+539		0.0014	29
11	15 30	41	26	(1,2)	1524+418		0.0083	29, 30
12	18 44	47.4	26	(1,2)	1907+508		0.0080	29, 30
13	20 00	60	26	(1,2)	2036+659		0.006	30
14	21 50	28	26	(1,2)	2205+311		0.0041	29, 30
15	23 20	3	26	(1,2)	2302+120		0.0087	29, 30
					2331+096		0.0067	29, 30

of the celestial sphere with a galactic latitude $|b|\leqslant10°$, i.e., were situated well within the ''zones of avoidance'' of the galaxies. We therefore consider two groups of showers: (1) all the showers; (2) showers arriving from parts of the celestial sphere with $|b|>10°$. The number of showers K in (1) and (2) and the number of showers N for which at least one galaxy with $z\leqslant0.0092$ occurred in the search field are shown in Table III. We consider groups of K artificial showers, for which the arrival coordinates are randomly determined and for which the error in the arrival direction is $(\Delta\alpha,\Delta\delta)=3$, 4, 5, 6, 7°, and we determine the probabilities of the random incidence of galaxies with $z\leqslant0.0092$ in the search field $3(\Delta\alpha,\Delta\delta)$ for a given number of showers N, precisely as in our earlier work.[23] The probabilities are shown in Table III. It follows from the table that the error in the arrival direction of the showers influences source identification. The probability of the random incidence of galaxies in the search field for a given number of showers is found to differ for different catalogs of galaxies. For the catalogs we used, Refs. 29 and 30, this probability is low, $P>3\sigma$, for showers (1) if $(\Delta\alpha,\Delta\delta)<4°$, and for showers (2) if $(\Delta\alpha,\Delta\delta)<6°$. For source identification, therefore, the proposed method must be used to distinguish showers for which the error in the arrival direction $(\Delta\alpha,\Delta\delta)$ is less than 4° or, if the showers have $|b|>10$, is less than 6°.

2.3. Identification of sources from showers with energies $3.2\times10^{19}<E<10^{20}$ eV

Proceeding from the results of Sec. 2.2, we include four Yakutsk showers, Nos. 1, 3, 5, and 12, in the statistics (1) and four Yakutsk showers, Nos. 1, 5, 7, 12, in the statistics (2). The new number of showers K in (1) and (2) and the probabilities of the random incidence of galaxies with $z\leqslant0.0092$ in the search field for N of them are shown in Table IV. These probabilities are low, $P=3\sigma-3.70\sigma$; con-

TABLE III. Probabilities P of random incidence of Seyfert galaxies with $z\leqslant0.0092$ in search zone $(3\Delta\alpha, 3\Delta\delta)$ for N of K showers with $3.2\times10^{19}<E<10^{20}$ eV (Refs. 4 and 26) in groups (1) and (2) [$K=15$ in (1) and $K=12$ in (2)].

Group	Probability P					
	$(\Delta\alpha,\Delta\delta)=3°$	$(\Delta\alpha,\Delta\delta)=4°$	$(\Delta\alpha,\Delta\delta)=5°$	$(\Delta\alpha,\Delta\delta)=6°$	$(\Delta\alpha,\Delta\delta)=7°$	$(\Delta\alpha,\Delta\delta)=8°$
(1)[a]	2.6×10^{-4}	6.7×10^{-3}	0.043	0.13	0.23	0.23
(2)[a]	$<4\times10^{-5}$	8.0×10^{-5}	1.2×10^{-3}	7.8×10^{-3}	0.036	0.11
(1)[b]	1.75×10^{-3}	0.02	0.012	0.048	0.12	0.20
(2)[b]	7.0×10^{-5}	1.8×10^{-3}	2×10^{-4}	1.3×10^{-3}	6.7×10^{-3}	0.023

[a]Galaxy search from Catalogs in Refs. 29 and 30; $N=15$ in (1), $N=12$ in (2).
[b]Galaxy search from Catalog in Ref. 29 only; in both groups $N=10$ for $\sigma_{sh}\leqslant4°$ and $N=12$ for $\sigma_{sh}\geqslant5°$.

TABLE IV. Probabilities P of random incidence of possible sources in the search domain for N of K showers in groups (1) and (2); (I) showers with $3.2\times10^{19}<E<10^{20}$ eV (Yakutsk, Ref. 10; AGASA, Refs. 4 and 26), sources are Seyfert galaxies with $z\leqslant0.0092$; (II) showers with $3.2\times10^{19}<E\leqslant(2-3)\times10^{20}$ eV (Yakutsk, Ref. 10; AGASA, Ref. 4 and 26; Haverah Park, Ref. 11), sources are Seyfert galaxies with $z\leqslant0.0092$; (III) the same showers as in (II), sources are lacertides from Ref. 30; (IV) the same showers as in (II), sources are radio galaxies from Ref. 31.

	(I)			(II)			(III)			(IV)		
Showers	N	K	P	N	K	P	N	K	P	N	K	P
(1)[a]	15	19	2.4×10^{-3}	18	22	7.2×10^{-4}	14	22	0.12	19	22	0.18
(2)[a]	16	16	1.8×10^{-3}	19	19	5.9×10^{-4}	18	19	0.086	19	19	0.65
(1)[b]	15	19	3.6×10^{-4}	17	22	2.5×10^{-4}	-	-	-	-	-	-
(2)[b]	16	16	2.5×10^{-4}	19	19	3.0×10^{-5}	-	-	-	-	-	-

[a]Search for Seyfert galaxies from Refs. 29 and 30.
[b]Search for Seyfert galaxies from Ref. 29 only.

sequently, the Yakutsk detector data confirm the previous source identification based on the data of Refs. 26 and 27.

3. SHOWERS WITH ENERGIES $3.2\times10^{19}<E\leqslant3\times10^{20}$ eV

Data on showers with $E\geqslant10^{20}$ eV are shown in Table V. Also shown in the same table are Seyfert galaxies with red shifts $z\leqslant0.0174$ lying within the search field $(3\Delta\alpha,3\Delta\delta)$ around the shower axis. Shower Nos. 1, 4, and 5 in Table V can be included in the statistics (1) and (2). The number of showers K in the resulting groups and the probabilities P of the random incidence of galaxies with $z\leqslant0.0092$ in the search field for N of them are shown in Table IV. The probability depends on the galaxy catalog used for identification, but it is small in any case: For showers not selected by galactic latitude b we have $P=(2.5-7.2)\times10^{-4}$, and for showers with $|b|>10°$ we have $P=3.0\times10^{-5}-5.9\times10^{-4}$. The probabilities of the random incidence of Seyfert galaxies with different values of z in the search field for N out of K showers are shown in Fig. 1, from which it follows that by selecting galaxies with $z\leqslant0.0174$ in the search field we are in fact selecting those for which random coincidence of the coordinates with the direction of the particles has a low probability.

In addition to Seyfert galaxies, the search field also encompasses lacertides and radio galaxies. The probabilities of the random incidence of these objects in the search field for N out of K showers are shown in Table IV. The probabilities P are high, and according to probability theory, the incidences can be random.

4. DISCUSSION

Our source identifications disagree with the conclusions of Refs. 32 and 33. The authors of Ref. 32 investigated the same showers with $E\geqslant10^{20}$ eV as those in Ref. 25 and in this paper. In Ref. 32, however, they concluded that the sources of the shower-driven particles are quasars with red shifts $0.3\leqslant z\leqslant2.2$. In identifying the possible sources, the authors of Ref. 32 considered objects in the rms error zone around the shower axis and estimated the probability of random incidence of objects in this zone. As mentioned, however, the coordinates of the particles are situated in the rms error field with only 66% probability. Possible sources must also be sought in the triple-error zone around the shower axis, where the coordinates of the particles lie with 99.8% probability. The probability of random incidence of objects in the rms error zone is two to three orders of magnitude lower than in the triple-rms-error zone. Reference 33 deals with the celestial distribution of showers with energies $E\geqslant4\times10^{19}$ eV recorded by the Volcano Ranch Station, Haverah Park, Sydney, and Yakutsk detector arrays. The authors

TABLE V. Showers with $E>10^{20}$ eV and the nearest Seyfert galaxies in their search field. Our shower numbering system is used.

		Coordinates		Seyfert galaxies		
Shower	Energy, eV	α	δ	α δ	z	Catalog[a]
1 [11][b]	$(1.20\pm0.10)\times10^{20}$	$11^h56^m\pm12^m$	$27\pm2.8°$	$11^h37^m+32.1°$	0.0092	[29, 30]
				1217+295	0.0022	[29, 30]
				1223+338	0.001	[30]
				1233+262	0.0037	[29]
2 [11][b]	$(1.05\pm0.08)\times10^{20}$	$13^h24^m\pm34.8^m$	$71\pm2.5°$	1205+654	0.0049	[29]
				1235+744	0.0067	[29]
				1339+679	0.0090	[29, 30]
3 [12]	3×10^{20}	$06^h20^m\pm24^m$	$48(+5.2,-6.3)°$	0645+609	0.0068	[29, 30]
4 [4][c]	1.1×10^{20}	18^h42^m	48°	1907+508	0.0080	[29, 30]
5 [4][c]	2.1×10^{20}	01^h15^m	21.1°	0140+133[d]	0.003	[30]
				0124+189	0.0174	[29, 30]

[a]Catalog Ref. Nos. are shown in brackets.
[b]Errors $\Delta\alpha$ and $\Delta\delta$ are taken from Ref. 32.
[c]Errors $(\Delta\alpha,\Delta\delta)=\sqrt{2}\times1.6°\approx2.3°$.
[d]The galaxy lies in the search field if $\Delta\alpha=2°$ and $\Delta\delta=2.6°$.

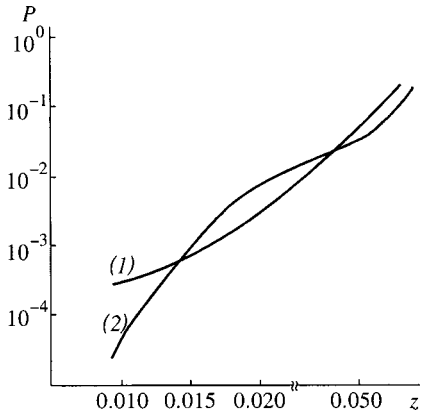


FIG. 1. Probability of random incidence of Seyfert galaxies[29] with different z in the search field for N of K showers. (*1*) $K=22$, $N=17$ for $z\geqslant 0.01$, $N=19$ for $z>0.01$, search field $(3\Delta\alpha, 3\Delta\delta)<12°$; (*2*) $K=N=19$, galactic width of the shower axis $|b|>10°$, search field $(3\Delta\alpha, 3\Delta\delta)<18°$.

were unable to identify any kind of extragalactic cosmic particle sources. In analyzing the celestial distribution of the shower arrival directions, the authors relied on two assumptions. First, for all the showers the source search field was $\Delta\alpha\leqslant 3°$, $\Delta\delta\leqslant 3°$. This field is at least three times smaller than triple the error of the coordinates for all these showers. Second, the nuclei of active galaxies with $z<0.017$, Seyfert galaxies in particular, were assumed to be uniformly distributed in the celestial sphere. However, the distribution of near Seyfert galaxies at a distance less than 100 Mparsec from Earth ($z<0.025$) is nonuniform, as we have demonstrated previously.[24,25]

It follows from the results of Table IV that the principal sources of particles with $3.2\times 10^{19}<E\leqslant(2-3)\times 10^{20}$ eV are Seyfert galaxies with $z\leqslant 0.0092$, which are weak emitters in the x-ray and radio ranges.

In none of our papers have we looked for possible sources in interacting galaxies, in which, according to Ref. 19, conditions can exist for the effective acceleration of particles. Following are the reasons for this decision. In identifying sources in Refs. 23 and 25, we calculated the probabilities of random incidence of objects as possible sources in the shower search field. In all the given shower selections this probability has been found to be low only for Seyfert galaxies: $P>3\sigma$. The majority of normal galaxies are interacting galaxies,[34] and their number is tens of times greater than the number of active nuclei. Consequently, the probability of random incidence of a normal galaxy in the search field will be much higher than for Seyfert galaxies. At the present time it is difficult to discriminate from observational data normal galaxies in which the conditions set forth in Ref. 19 exist for the sufficiently effective acceleration of particles.

5. CONCLUSION

The principal sources of particles driven by showers with energies of $3.2\times 10^{19}<E\leqslant 3\times 10^{20}$ eV are Seyfert galaxies with red shifts $z\leqslant 0.0092$, which are weak emitters in the x-ray and radio ranges. For galactic latitudes $|b|>10°$ the probability of random coincidence of the coordinates of such galaxies with the particle arrival directions is $P=3\times 10^{-5}-5\times 10^{-4}$, depending on the galaxy catalog. [For any galactic latitudes down to $b=0°$ this probability is also low, $P=(2.5-7.2)\times 10^{-4}$.] The error $(\Delta\alpha, \Delta\delta)$ in the shower arrival direction influences the identification of sources. For identification it is necessary to select showers for which the error is less than 4°. If showers with $|b|>10°$ are selected, the error $(\Delta\alpha, \Delta\delta)$ can exceed 6°.

The results obtained here can be tested in studies at the AGASA, Fly's Eye, and Haverah Park ground-based detectors and also at new (under construction and contemplated) giant arrays designed for the investigation of showers with energies $E\geqslant 10^{20}$ eV, which will have far superior angular (to 0.2°) and energy resolution, such as the ShAL-1000, Telescope Array, HiRes, and Pierre Auger Project, as well as satellites.[35]

I would like to thank I. E. Sleptsov for graciously providing the horizontal coordinates of the twelve Yakutsk showers.

*)E-mail: uryson@sci.lebedev.ru

[1)] In an earlier paper[23] we incorrectly stated the coordinates of showers 7 and 8. The nearest Seyfert galaxy in their search field is NGC660, $z=0.003$ (Ref. 31). This oversight did not affect the final conclusions in Ref. 23.

[1] V. S. Berezinskiĭ, S. V. Bulanov, V. L. Ginzburg, V. A. Dogel', and V. S. Ptuskin, in V. L. Ginzburg (Ed.), *Astrophysics of Cosmic Rays*, North–Holland, Amsterdam (1990) [Russian orig., Nauka, Moscow (1990)].
[2] M. N. D'yakonov, T. A. Egorov, N. N. Efimov *et al.*, *Extremely High-Energy Cosmic Rays* [in Russian], Nauka, Sib. Otd., Novosibirsk (1991).
[3] T. Stanev, P. L. Biermann, J. Lloyd-Evans *et al.*, Phys. Rev. Lett. **75**, 3056 (1995).
[4] N. Hayashida, K. Honda, M. Honda *et al.*, Phys. Rev. Lett. **77**, 1000 (1996).
[5] J. Rachen, T. Stanev, and P. Biermann, Astron. Astrophys. **273**, 377 (1993).
[6] F. W. Stecker, Phys. Rev. Lett. **80**, 1816 (1998).
[7] G. T. Zatsepin and V. A. Kuz'min, JETP Lett. **4**, 78 (1966).
[8] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
[9] M. M. Winn, J. Ulrichs, L. S. Peak *et al.*, J. Phys. G Nucl. Phys. **12**, 653 (1986).
[10] B. N. Afanasiev, M. N. Dyakonov, V. P. Egorova *et al.*, in *Proceedings of the International Symposium on Extremely High Energy Cosmic Rays: Astrophysics and Future Observations*, M. Nagano (Ed.), Tanashi, Tokyo (1996), p. 32.
[11] A. Watson, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millennium*, E. W. Kolb and R. D. Peccei (Eds.), World Scientific, Singapore (1995), p. 126.
[12] D. Bird, S. C. Corbato, H. Y. Dai *et al.*, Astrophys. J. **441**, 144 (1995).
[13] M. Takeda, N. Hayashida, K. Honda *et al.*, Phys. Rev. Lett. **81**, 1163 (1998).
[14] B. N. Afanasiev, M. N. Dyakonov, V. P. Egorova *et al.*, in *Proceedings of the 24th International Cosmic Ray Conference (ICRC-25)*, Rome (1995), Vol. 2, p. 756.
[15] N. Hayashida, K. Honda, M. Honda *et al.*, in *Proceedings of the International Symposium on Extremely High Energy Cosmic Rays: Astrophysics and Future Observations*, M. Nagano (ed.), Tanashi, Tokyo (1996). p. 17.
[16] A. V. Uryson, JETP Lett. **65**, 763 (1997).
[17] A. V. Uryson, Zh. Éksp. Teor. Fiz. **113**, 12 (1998) [JETP **86**, 6 (1998)].
[18] N. S. Kardashev and B. V. Komberg, private communication (1998).
[19] C. Cesarsky and V. Ptuskin, in *Proceedings of the 23rd International Cosmic Ray Conference (ICRC-23)*, Calgary (1993), Vol. 2, p. 341.
[20] V. Berezinsky and A. Vilenkin, Phys. Rev. Lett. **79**, 5202 (1997).
[21] V. Berezinsky, M. Kachelrie, and A. Vilenkin, Phys. Rev. Lett. **79**, 4302 (1997).
[22] T. Totani, Astrophys. J. **502**, L13 (1998).
[23] A. V. Uryson, JETP Lett. **64**, 77 (1996).

[24] A. V. Uryson, Zh. Éksp. Teor. Fiz. **113**, 385 (1998) [JETP **86**, 213 (1998)].
[25] A. V. Uryson, Izv. Ross. Akad. Nauk, Ser. Fiz. **63**, 627 (1999).
[26] N. Hayashida, K. Honda, M. Honda *et al.*, in *Proceedings of the 22nd International Cosmic Ray Conference (ICRC-22)*, Dublin (1991), Vol. 2, p. 117.
[27] G. L. Squires, *Practical Physics*, McGraw-Hill, New York (1968).
[28] I. E. Sleptsov, private communication (1998).
[29] V. A. Lipovetskiĭ, S. N. Neizvestnyiĭ, and O. M. Neizvestnaya, Soobshch. SAO, No. 55 (1987).
[30] M. P. Veron-Cetty and P. Veron, ESO Sci. Rep., No. 13 (1993).
[31] H. Kühr, A. Witzel, and I. I. K. Pauliny-Toth, Astron. Astrophys., Suppl. Ser. **45**, 367 (1981).
[32] G. R. Farrar and P. L. Biermann, Phys. Rev. Lett. **80**, 1816 (1998).
[33] G. B. Khristiansen, G. V. Kulikov, G. G. C. Palumbo *et al.*, in *Proceedings of the 25th International Cosmic Ray Conference (ICRC-25)*, Durban, South Africa (1997), Vol. 4, p. 201.
[34] G. S. Wright, R. D. Joseph, N. A. Robertson *et al.*, Mon. Not. R. Astron. Soc. **233**, 1 (1988).
[35] *Proceedings of the 25th International Cosmic Ray Conference (ICRC-25)*, Durban, South Africa (1997), Vol. 5.

Translated by James S. Wood

Topological interpretation of quantum numbers

S. A. Bulgadaev*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 28 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 1131–1147 (October 1999)

The paper shows how vector topological charges for topological excitations in nonlinear σ-models on compact one-dimensional spaces T_G and G/T_G can be defined (here G is a simple compact Lie group and T_G is its maximal commutative subgroup). Explicit solutions, their energies and interactions between different topological charges have been obtained. A possibility of topological interpretation of quantum numbers of groups and particles is discussed.
 [S1063-7761(99)00210-3]

1. INTRODUCTION

The discovery of new, topologically stable solutions,[1–4] has revived an old hypothesis, which dates back to antiquity, that elementary particles have a topological nature. Descartes, for example, suggested a vortex model of magnetism,[5] Helmholtz suggested a vortex model of matter,[6] and Lord Kelvin suggested the concept of atoms as knotted configurations of the ''ether.''[6] Although all these hypotheses proved wrong in the long run, they bore fruit, since they stimulated studies of these phenomena and development of the theories of vortices and knots,[7] as well as their applications to various physical sciences, ranging from hydrodynamics to the theory of polymers.[8–10]

The advantage of this concept over the conventional approach, treating particles as structureless objects, is that the former suggest a visual picture of an elementary particle. Modern physics has much in common with the physics of the second half of the 19th century, given that condensates of various fields play the role of ''ether'' or another hypothetical fluid. For this reason, such concepts have been revived from time to time in the history of physics. In particular, theories based on various topological invariants have achieved considerable success in the physics of condensed state.[9–11] In field theory, the hypothesis of a topological nature of elementary particles and their properties has also been quite popular because it has provided a more visual interpretation of their structure and quantum numbers, in comparison with the previous formal, purely algebraic description. The literature on this topic is enormous, so here we quote only one book for reference.[12] But the set of topological charges generated by only one of the topological invariants is insufficient for a complete description of real particles with their internal symmetries.

The necessary conditions for a topological interpretation are the following:

(1) the corresponding field theory should have a degenerate vacuum state which would form a certain manifold $\mathscr{M}$ with a nontrivial topology;

(2) the theory should have solutions with nontrivial topologies and finite (or logarithmically divergent) energies;

(3) the set of allowed topological charges should be the same as the set of possible quantum numbers.

The first two properties obtain in the cases of such topological solutions as solitons, vortices, instantons, and monopoles, but the third condition is not satisfied, at least for the first three types of excitations.

It is noteworthy that in the recent time adequate techniques for description and investigation of corresponding topological problems have been developed in mathematics.[13]

Usually the quantum numbers of particles are determined by the weights of irreducible representations to which they belong. It is known that the weights of simple compact groups G are related to the maximal commutative subgroup T_G of the group G, i.e., the Cartan maximal Abelian tori.[14]

All possible weights of a simple compact group form an n-dimensional lattice $\mathbb{L}_w$, where n is the rank of the group G. Therefore, topological charges should also belong to this lattice. It was shown previously that solitons with topological isovector charges belonging to the lattice $\mathbb{L}_w$ of the group G can exist in two-dimensional theories generalizing the sine-Gordon (SG) theory. These theories are related to the characters of groups G of rank $n>1$ (Ref. 15). Their vacuum configurations are point-like and form infinite lattices. Charges of respective solitons can be related to the homotopic group π_0 of these lattices.

Generally speaking, feasibility of topological excitations in systems with a degenerate vacuum state depends on the nontriviality of homotopic groups $\pi_i(\mathscr{M})$ of the vacuum space $\mathscr{M}$. For example, $\pi_1(S^1)=\mathbb{Z}$ corresponds to conventional vortices and one-dimensional instantons,[16] whereas $\pi_2(S^2)=\mathbb{Z}$ corresponds to two-dimensional instantons.[4,13] In order to obtain topological charges belonging to a lattice of weights of a simple compact Lie group G, one should consider the following manifolds:[17] a torus T_G in the cases of vortices and one-dimensional instantons, and a more special homogeneous space of group G, namely, the flag space $F_G = G/T_G$, in the case of two-dimensional instantons.

This paper continues the analysis of these manifolds $\mathscr{M}$. We show how vector topological charges can be defined for these manifolds, and explicit solutions with such charges are found. We consider mostly nonlinear σ-models on manifolds $\mathscr{M}$ because they are effective theories for a wide range of models with $\mathscr{M}$ as their vacuum manifolds. We will also

discuss conditions under which quantum numbers of groups G admit a topological interpretation.

2. MANIFOLD WITH A VECTOR GROUP π_1

It is convenient to begin our study of manifolds with vector (i.e., those of rank $n>1$) homotopic groups with manifolds with vector groups $\pi_1(\mathscr{M})$. For our purposes, it is sufficient to consider only manifolds with Abelian π_1, since the corresponding topological charges should be commutative. For simplicity, we limit our analysis to free homotopic groups. Although their properties are simple and well known, we are going to discuss them, focusing attention on possible vector structures. This is helpful in discussing manifolds with vector groups π_2. The simplest generalization of a circle

$$S^1=e^{2\pi i\phi},\quad 0\leqslant\phi\leqslant 1,$$

is the torus T^n, which can be described as a direct product of n circles,

$$T^n=\otimes_{i=1}^n S_i^1. \tag{1}$$

The torus T^n can be also represented as a factor of an n-dimensional Euclidean space $\mathbb{R}^n$:

$$T^n=\mathbb{R}^n/\mathbb{Z}^n, \tag{2}$$

where $\mathbb{Z}^n$ is an n-dimensional simple cubic lattice whose sites have integer coordinates:

$$\mathbb{Z}^n=\oplus_{i=1}^n\mathbb{Z}_i. \tag{3}$$

A more general n-dimensional torus T_L of rank n can be defined as a factor

$$T_L=\mathbb{R}^n/\mathbb{L}, \tag{4}$$

where $\mathbb{L}$ is a certain n-dimensional nondegenerate lattice in $\mathbb{R}^n$:

$$\mathbb{L}=\sum_{i=1}^n n_i\mathbf{e}_i,\quad n_i\in\mathbb{Z}_i,\quad \mathbf{e}_i\in\{\mathbf{e}_i\}_L. \tag{5}$$

Here the set of linearly independent vectors $\{\mathbf{e}_i\}_L$, $i=1,\ldots,n$ is the basis of the lattice $\mathbb{L}$. For T^n, the set $\{\mathbf{e}_i\}_L$ is identical to the canonical orthonormal basis

$$\mathbf{e}_i=(0,\ldots,0,1_i,0,\ldots,0),\quad i=1,\ldots,n.$$

In any lattice the basis can be defined in several different ways, and all these bases are related to each other by modular transformations

$$\{\mathbf{e}_i'\}_L=M\{\mathbf{e}_i\}_L,\quad \det(M)=\pm 1, \tag{6}$$

where M are matrices with integer elements. For bases whose vectors have the same orientation $\det(M)=1$. Usually the most convenient basis is that whose vectors have minimal norms.

The first homotopic group $\pi_1(T^n)$ of a torus, being a group of homotopic classes of mappings of the sphere S^1 into T^n, is usually expressed as follows:

$$\pi_1(T^n)=\oplus_{i=1}^n\mathbb{Z}_i=\mathbb{Z}^n, \tag{7}$$

where the ith component describes the mapping of S^1 onto the ith circle of the torus T^n. It is clear that mappings onto different components cannot cancel out. A similar expression is often used for arbitrary tori T_L. Then this form only signifies that coordinates of different homotopic mapping classes are integer in a certain basis but do not contain information about this basis. This means that a vector basis should be introduced for a complete description of a homotopic group in $\pi_1(T_L)$. In order to obtain this basis explicitly, one can use the factor (coset) nature of the torus T_L. It is desirable to select this basis to be compatible with the vector structure of the covering space $\mathbb{R}^n$. The most natural way is to retain the Euclidean vector structure $\mathbb{R}^n$. In this case, we obtain the following expression for $\pi_1(T_L)$:

$$\pi_1(T_L)=\mathbb{L}, \tag{8}$$

where $\pi_1(T_L)$ contains basis vectors $\mathbf{e}_i$ in explicit form this time. This form contains more information than Eq. (7) because, in addition to the integer property of homotopic classes ($n_i\in\mathbb{Z}_i$), it also describes the geometrical characteristics of the torus T_L. For a full description we also need the metric and the corresponding scalar product in the space of topological charges. They will be of considerable importance in discussing interaction between topological excitations and various topological charges. Note that this metric can differ from that of the covering space (see below).

Since each basis vector $\mathbf{e}_i$ corresponds to a nontrivial elementary homological cycle γ_i of the torus T_L, the corresponding homologic group of cycles $H_1(T_L,\mathbb{Z})$ is isomorphic to the group $\pi_1(T_L)$. This is a consequence of Gurevich's general theorem about isomorphism of the first nontrivial homotopic and homologic groups.[13] Consequently, $H_1(T_L,\mathbb{Z})$ also has a vector structure, which can be expressed as

$$\boldsymbol{\gamma}=\sum n_i\boldsymbol{\gamma}_i,\quad \boldsymbol{\gamma}_i=\gamma_i\mathbf{e}_i,\quad n_i\in\mathbb{Z}_i. \tag{9}$$

A similar vector structure can be introduced to a cohomologic group of integer-valued differential 1-forms $\delta\in H^1(T_L,\mathbb{Z})$, which is dual to a group $H_1(T_L,\mathbb{Z})$,

$$\boldsymbol{\delta}=\sum_k n_k\boldsymbol{\delta}_k,\quad n_k\in\mathbb{Z},$$

with basis $\{\boldsymbol{\delta}_k\}=\{\delta_k\mathbf{s}_k\}=\{d\phi^k\mathbf{s}_k\}$ and convolution

$$(\boldsymbol{\gamma}_i\circ\boldsymbol{\delta}_k)=(\mathbf{e}_i\cdot\mathbf{s}_k)\oint_{\gamma_i}d\phi^k=\delta_{ik}, \tag{10}$$

where $\{\mathbf{s}_k\}$ is the basis of the dual lattice $\mathbb{L}^*$:

$$(\mathbf{s}_i\cdot\mathbf{e}_k)=\delta_{ik}, \tag{11}$$

and $(\mathbf{s}_i\cdot\mathbf{e}_k)$ denotes the Euclidean scalar product of vectors ($\mathbf{s}_i$ and $\mathbf{e}_k$).

Topological charges Q of the torus T^n are usually defined as integers equal to the winding number of elementary cycles γ_i:

$$Q=(q_1,\ldots,q_n),\quad q_i\in\mathbb{Z}_i. \tag{12}$$

The winding numbers can be expressed as integrals in the space of the parameters of the torus T^n of 1-forms δ_k conjugate to cycles γ_i,

$$q_i = \oint_{\gamma_i} d\phi = \oint_{\gamma_i} \sum_{k=1}^{n} n_k d\phi^k = n_i, \quad d\phi = \sum_{k=1}^{n} n_k d\phi^k,$$

or

$$q_i = \oint_{\gamma} d\phi^i = \sum_{k=1}^{n} n_k \oint_{\gamma_k} d\phi^i = n_i, \quad \gamma = \sum_{k=1}^{n} n_k \gamma_k. \tag{13}$$

Here $\gamma \in H_1(T^n, \mathbb{Z})$ and $d\phi \in H^1(T^n, \mathbb{Z})$ are a certain closed path and closed 1-form on T^n, respectively. Similar expressions for Q are also frequently used in the case of arbitrary tori T_L. In this formal representation, all Q's have the same integer structure. All geometrical properties of charges Q are again hidden in abstract bases of elementary cycles and 1-forms.

In order to obtain vector topological charges corresponding to a vector homotopic group, one can use any of the vector structures mentioned above. For example, one can go up to the covering space $\mathbb{R}^n$ and define vector topological charges $\mathbf{Q}$ as contour integrals of vector 1-forms dx^k ($\mathbf{x} \in \mathbb{R}^n$) over the curve γ^*, which is a prototype of the closed path γ on the torus:

$$\mathbf{Q} = \int_{\gamma^*} d\mathbf{x} = \sum_{i=1}^{n} n_i \mathbf{e}_i. \tag{14}$$

The latter equation derives from the condition that both ends of the curve γ^* should belong to the lattice $\mathbb{L}$. A similar expression is obtained if one uses the vector structure from Eq. (9) in the space of cycles γ:

$$\mathbf{Q} = \oint_{\boldsymbol{\gamma}} d\phi = \sum_{i,k}^{n} n_i \mathbf{e}_i \oint_{\gamma_i} d\phi^k = \sum_{i=1}^{n} n_i \mathbf{e}_i. \tag{15}$$

In what follows, we will apply the second approach, which uses the vector structure of the homologic group, to all $\mathscr{M}$ under discussion because they will satisfy the conditions of the Gurevich theorem. Such a selection is both easy to visualize and convenient, especially when the differential forms involved are specified, e.g., by an action functional of the theory or symmetry properties.

Thus, we have introduced vector topological charges of the torus T_L explicitly containing the lattice basis $\{\mathbf{e}_i\}_L$. In order to obtain the metric and the respective scalar product in the space of topological charges, one should consider a specific realization of the torus T_L and a chiral model on this torus. The torus T_L of rank n can be realized in the form of a diagonal n-dimensional matrix $\mathbf{t}$:

$$\mathbf{t} = \mathrm{diag}[\exp(2\pi i \mathbf{s}_1 \cdot \boldsymbol{\phi}), \ldots, \exp(2\pi i \mathbf{s}_n \cdot \boldsymbol{\phi})], \tag{16}$$

where the set of vectors $\{\mathbf{s}_i\}$ is the basis of the dual lattice $\mathbb{L}^*$ from Eqs. (10) and (11). It is clear that one of the exponential functions in the matrix $\mathbf{t}$ with dual vector $\mathbf{s}_i$ corresponds to each elementary cycle $\mathbf{e}_i$.

The lattice $\mathbb{L}^*$ is a lattice of weights for the torus T_L treated as a group. The necessary condition for the topological interpretation of its weights is $\mathbb{L} \supseteq \mathbb{L}^*$. For example, the lattice $\mathbb{Z}^n$ is identical to its dual lattice, hence all weights of the group T^n have a topological interpretation.

For this reason, tori corresponding to the self-dual lattices $\mathbb{L} = \mathbb{L}^*$ play an important role in the case of general tori T_L, since all their weights are in a one-to-one correspondence with possible topological charges.

Now let us proceed to compact chiral models related to tori T_L and present an explicit realization of the vector topological charges introduced previously. A first model is a generalization of the two-dimensional σ-model on circle S^1 (Ref. 19). We will use for T_L the realization in the form (16). A two-dimensional (Euclidean) nonlinear σ-model on T_L can be described in terms of the following action $\mathscr{S}$ (or energy E):

$$\mathscr{S} = \frac{1}{2\alpha} \int d^2x \, \mathrm{Tr}(\mathbf{t}_\mu^{-1} \mathbf{t}_\mu) = \frac{(2\pi)^2}{2\alpha} \int d^2x (\boldsymbol{\phi}_\mu \cdot \boldsymbol{\phi}_\mu), \tag{17}$$

where $\mathbf{t}_\mu = \partial_\mu \mathbf{t}$, $\boldsymbol{\phi}_\mu = \partial_\mu \boldsymbol{\phi}$, $\mu = 1,2$. Here the new scalar product (*) is determined by the effective n-dimensional metric g_{ik} on the space of the torus

$$(\boldsymbol{\phi} \cdot \boldsymbol{\phi}) = \sum_{p=1}^{n} (\mathbf{s}_p \cdot \boldsymbol{\phi})(\mathbf{s}_p \cdot \boldsymbol{\phi}) = \sum g_{ik} \phi^i \phi^k,$$

$$g_{ik} = \sum_{p=1}^{n} s_p^i s_p^k. \tag{18}$$

It is clear that metric g_{ik} for a general lattice $\mathbb{L}$ differs from the canonical Euclidean metric on $\mathbb{R}^n$. This is the metric which defines interaction among various topological charges on the torus T_L.

Such chiral models can be treated as effective theories corresponding in the long-wave limit to theories of the Ginzburg–Landau type with vacuum manifold $\mathscr{M} = T_L$ or lattice models with variables that belong to T_L. Since $\pi_1(T_L)$ is nontrivial, the theory yields vortex-like solutions. These vortex solutions are ill-defined at small distances where the full action should be used to determine the vortex core structure. In describing topological properties, however, only their behavior at large distances is important. For this reason, we will use in our analysis a cut-off parameter a at small distances, which is an analogue of the vortex core radius or lattice constant.

The corresponding equations

$$g_{ik} \partial^2 \phi^k = 0 \tag{19}$$

have N-vortex solutions in plane $\mathbb{R}^2$ with N punctures at the points (x_i, y_i):

$$\boldsymbol{\phi}(\mathbf{x}) = \sum_{i=1}^{N} \mathbf{q}_i \frac{1}{\pi} \tan^{-1}\left(\frac{y - y_i}{x - x_i}\right), \quad \mathbf{q}_i \in \mathbb{L}, \quad (\mathbf{q}_i \cdot \mathbf{s}_k) \in \mathbb{Z},$$

$$(x,y) \in \mathbb{R}^2. \tag{20}$$

The energy of one vortex with topological charge

$$\mathbf{q} = \sum_{i=1}^{n} n_i \mathbf{e}_i$$

is logarithmically divergent:

$$E_q=\frac{(\mathbf{q}\cdot\mathbf{q})}{2\alpha}2\pi\ln\frac{R}{a},\quad (\mathbf{q}\cdot\mathbf{q})=\sum_{i=1}^{n} n_i^2, \tag{21}$$

where R is the space radius. The energy of the N-vortex solution with the total topological charge

$$\mathbf{Q}=\sum_{i=1}^{N}\mathbf{q}_i=0$$

is finite and equals

$$E_N=\frac{2\pi}{2\alpha}\sum_{i\neq k}^{N}(\mathbf{q}_i\cdot\mathbf{q}_k)\ln\frac{|x_i-x_k|}{a}+C(a)\sum_{i}^{N}(\mathbf{q}_i\cdot\mathbf{q}_i). \tag{22}$$

Here the scalar product of topological charges is given by the formula

$$(\mathbf{q}_i\cdot\mathbf{q}_k)=\sum_{p=1}^{n} n_{ip}n_{kp},\quad \mathbf{q}_i=\sum_{p=1}^{n} n_{ip}\mathbf{e}_p, \tag{23}$$

and $C(a)$ is a specific (not universal) constant which determines the ''self-energy'' (or core energy) of vortices and depends on how the core is regularized. It follows from Eqs. (21) and (22) that, irrespective of the lattice $\mathbb{L}$, topological charges corresponding to different cycles do not interact with one another, as in the case of the torus T^n! This property is a consequence of the occurrence of the effective metric g_{ik} on the torus space. We can see that properties of vortex-like excitations in nonlinear σ-models on all tori T_L (with torus dimensionalities equal to their ranks, including complex Abelian tori) are similar to those of a nonlinear σ-model on the torus T^n. It is easily understandable, since initial cycles on any torus can be deformed so as to obtain canonical cycles. At the same time, this result supports the correctness of the vector structure which we have introduced in the space of topological charges.

But this is not the end of the story. There are tori different from those like T^n and T_L. They can be called degenerate because their ranks n are smaller than their dimensionalities p. They will be discussed in the next section.

Other models to be discussed in this paper are one-dimensional conformal σ-models on T_L, which are generalizations of similar σ-models on S^1. Their action is expressed as

$$\mathcal{S}=\frac{1}{2\alpha}\int dx dx'\frac{|\mathbf{t}(x)-\mathbf{t}(x')|^2}{(x-x')^2} \tag{24}$$

$$=\frac{1}{2\alpha}\int dx\,dx'2\sum_{i=1}^{n}\{1-\cos[2\pi(\mathbf{s}_i\cdot(\boldsymbol{\phi}(x)-\boldsymbol{\phi}(x')))]\}, \tag{25}$$

where $x\in\mathbb{R}^1$. By using the technique suggested previously[16] one can prove that these models have N-instanton solutions like that given by Eq. (20) with vector topological charges $\mathbf{Q}=\Sigma_i^N\mathbf{q_i}\in\mathbb{L}$:

$$\boldsymbol{\phi}(x)=\sum_{i=1}^{N}\mathbf{q}_i\frac{1}{\pi}\tan^{-1}\frac{x-a_{1i}}{a_{2i}}. \tag{26}$$

Here a_{1i} and a_{2i} are arbitrary constants characterizing positions and widths of the corresponding instantons. All topological charges $\mathbf{q}_i$ in Eq. (26) should satisfy the condition

$$(\mathbf{q}_i\cdot\mathbf{s}_k)\leqslant 0\ \text{ or }\ (\mathbf{q}_i\cdot\mathbf{s}_k)\geqslant 0 \tag{27}$$

for all $\mathbf{s}_k$, $k=1,\ldots,n$. This condition is a generalization of the similar condition in the σ-model on S^1, which is, in its turn, an analogue of the analytical (or anti-analytical) property of two-dimensional instantons. The corresponding action is

$$\mathcal{S}_N=\frac{(2\pi)^2}{2\alpha}\sum_{k=1}^{n}\sum_{i}^{N}|(\mathbf{s}_k\cdot\mathbf{q}_i)|,\quad (\mathbf{s}_k\cdot\mathbf{q}_i)\in\mathbb{Z}. \tag{28}$$

Such an additive form of the action linear in $|q_i|$ holds only if the set of charges satisfies condition (27). Interaction between different charges turns up only for superpositions of instantons with arbitrary charges (or if fluctuations of the instanton background are taken into account).

3. CARTAN TORI

This section is dedicated to degenerate tori associated with the Cartan maximal Abelian tori T_G of simple compact Lie groups G. We will consider some of their representations, the corresponding homotopic groups $\pi_1(T_G)$, and their topological charges.

The Cartan torus T_G consists of elements

$$\mathbf{g}=\exp 2\pi i(\mathbf{H}\cdot\boldsymbol{\phi}),\quad \mathbf{H}=\{H_1,\ldots,H_n\}\in\mathscr{C},$$
$$[H_i,H_j]=0, \tag{29}$$

where n is the rank of the group G, and $\mathscr{C}$ is the Cartan maximal commutative subalgebra of the Lie algebra $\mathscr{G}$ of the group G. In deriving Eq. (29) we have assumed that the product $(\mathbf{H}\cdot\boldsymbol{\phi})$ is the conventional scalar product in the Euclidean space. All the H_i can be diagonalized at the same time because they are commutative. Their eigenvalues $\mathbf{w}$, which are named weights, depend on the specific representation of G and $\mathscr{C}$. The weights $\{\mathbf{w}_a\}_\tau$, $a=1,\ldots,p$, which belong to a p-dimensional representation $\tau(G)$, form a set of ''quantum numbers'' for this representation.

All possible weights $\mathbf{w}$ of a simply connected group G (or a universal covering group $\tilde{G}$ in the case of a multiply connected group G) form a lattice of weights $\mathbb{L}_w$. Its basis can be defined in several different ways. From the viewpoint of representation theory, the most convenient basis is that of fundamental weights $\overline{\mathbf{w}}_i$, $i=1,\ldots,n$. Any weight $\mathbf{w}$ can be expressed as

$$\mathbf{w}=\sum_{1}^{n} n_i\overline{\mathbf{w}}_i, \tag{30}$$

where all n_i are integers. In a general case, not all $\overline{\mathbf{w}}_i$ have a minimal norm. In some cases a more convenient basis is that composed of vectors with minimal norms.

The lattice of weights $\mathbb{L}_w$, the lattice of roots $\mathbb{L}_r$, and the lattice of dual roots $\mathbb{L}_v$ (the last two are related with $\mathbb{L}_w$) are necessary to calculate $\pi_1(T_G)$, which can depend on the specific representation $\tau(G)$. Group $\pi_1(T_G)$ can be obtained in a general form by the method of group theory and Lie

algebras.[14] Here we use a more simple and easily understandable technique which was applied in the previous section to tori T_L. Since one can select eigenvectors $|a\rangle$ of operators $\mathbf{H}$ as a basis in each irreducible representation $\tau(G)$ of dimensionality p,

$$\mathbf{H}|a\rangle = \mathbf{w}_a|a\rangle, \quad a=1,\ldots,p, \tag{31}$$

all H_i in this basis (and also all elements $\mathbf{g}\in T_G$) have a diagonal form:

$$\mathbf{g}_\tau = \mathrm{diag}[\exp(2\pi i\mathbf{w}_1\cdot\boldsymbol{\phi}),\ldots,\exp(2\pi i\mathbf{w}_p\cdot\boldsymbol{\phi})]. \tag{32}$$

In the general case some weights in Eq. (32) can be identical and even equal to zero. The trace of element $\mathbf{g}_\tau$ is a character of τ-representation of G:

$$\chi(\mathbf{g})_\tau = \sum_{a=1}^{p} \exp 2\pi i(\mathbf{w}_a\cdot\boldsymbol{\phi}).$$

The main differences between this form of tori T_G and those of type T_L in the form (16) are the following:

(1) The dimensionality of diagonal matrices in Eq. (32) equals the dimensionality p of the representation τ, which is usually greater than the rank n of the group G;

(2) The set of weights $\{\mathbf{w}\}_\tau$ has discrete Weyl symmetry, which implies Weyl invariance of the torus T_G and two properties:

$$\sum_{a=1}^{p}\mathbf{w}_a = 0, \quad g_{ik} = \sum_{a=1}^{p} w_i^a w_k^a = B_\tau\delta_{ik}, \tag{33}$$

where the constant B_τ depends on the representation.

It follows from Eq. (33) that, in the case of the torus T_G, the effective metric g_{ik} over the space of its parameters is proportional to the Euclidean metric. This property is very important for the interaction between different topological charges and the difference between properties of nonlinear σ-models on tori T_L and T_G.[18]

From Eq. (32) it is obvious that in this representation all $\mathbf{g}\in T_G$ are periodic with the lattice of periods $\mathbb{L}_\tau^{-1}$, which is reciprocal to the lattice $\mathbb{L}_\tau$ generated by the weights $\mathbf{w}_a$ ($a=1,\ldots,p$) of the representation τ. This means that $\mathbb{L}_\tau^{-1}$ forms a set of all topological charges of the τ-representation of the torus T_G. Let us denote this set by $\mathbb{L}_\tau^Q$. It clearly follows from the definition of the lattice $\mathbb{L}_\tau$ that these lattices are identical for conjugated representations (whose weight sets are opposite to one another). Generally speaking, the lattice $\mathbb{L}_\tau$ can be a sublattice of the lattice $\mathbb{L}_w$: $\mathbb{L}_w\supseteq\mathbb{L}_\tau$. If $\mathbb{L}_w\supset\mathbb{L}_\tau$, then we can define a factor-group $\pi_\tau = \mathbb{L}_w/\mathbb{L}_\tau$, which should be a finite Abelian group.

The lattice of dual roots $\mathbb{L}_v$ is reciprocal to the lattice $\mathbb{L}_w$ and has its own basis of dual roots $\mathbf{r}_i^v$ ($i=1,\ldots,n$):

$$(\mathbf{r}_i^v\cdot\overline{\mathbf{w}}_k) = \delta_{ik}. \tag{34}$$

Now one can see that, for those representations τ for which $\mathbb{L}_\tau = \mathbb{L}_w$ holds, the lattice $\mathbb{L}_\tau^Q$ is isomorphic to the n-dimensional lattice $\mathbb{L}_v$ of dual roots $\mathbf{r}^v$:

$$\mathbb{L}_\tau^Q = \mathbb{L}_v.$$

For those representations for which $\mathbb{L}_w\supset\mathbb{L}_\tau$ holds, the lattice $\mathbb{L}_\tau^Q$ is multiplied by the factor π_τ:

$$\mathbb{L}_\tau^Q = \mathbb{L}_v\times\pi_\tau,$$

where the symbol $\times$ denotes semidirect product. This means that $\mathbb{L}_v$ is a sublattice of the lattice $\mathbb{L}_\tau^Q$. This structure $\mathbb{L}_\tau^Q$ is similar to the structure of the group $\pi_1(T_G)$ for multiply connected groups G.[17] Again the maximal factor π_τ corresponds to an adjoint (ad) representation

$$\pi_{ad} = Z_G,$$

where Z_G is the center of the group G.

In order to obtain a topological interpretation of all weights $\{\mathbf{w}_a\}_\tau$ in terms of topological charges $\mathbf{Q}\in\mathbb{L}_\tau^Q$, we have to satisfy the following condition:

$$\mathbb{L}_\tau^Q = \mathbb{L}_v\times\pi_\tau \supseteq \mathbb{L}_\tau = \mathbb{L}_w/\pi_\tau.$$

In the case of exact equality we have

$$\mathbb{L}_v\times\pi_\tau = \mathbb{L}_w/\pi_\tau.$$

Structures of all lattices $\mathbb{L}_w$, $\mathbb{L}_r$, and $\mathbb{L}_v$, of simple compact groups are known.[14] For example, the lattices $\mathbb{L}_r$ and $\mathbb{L}_v$ belong to four series of integer lattices of types A, D, E, and Z. For simply constructed groups G [i.e., for groups of series $A_n=SU(n+1)$, $D_n=SO(2n)$, $E=E_{6,7,8}$], the lattice $\mathbb{L}_v$ is identical to the root lattice $\mathbb{L}_r$, which is a sublattice of the weight lattice, $\mathbb{L}_r\subseteq\mathbb{L}_w$:

$$\mathbb{L}_w/\mathbb{L}_r = Z_G.$$

In what follows, we will discuss only the minimal and adjoint representations. The weights of minimal representations usually generate a lattice of weights $\mathbb{L}_w$ (except the case of orthogonal groups from series B and D). In this case, the sublattice of topological charges is identical to that of dual roots $\mathbb{L}_v$. The weights of adjoint representations are roots. They generate a root lattice $\mathbb{L}_r$. In this case, topological charges belong to the lattice $\mathbb{L}_r^{-1}=\mathbb{L}_{w*}$, which is a weight lattice of the dual group G^* (or a lattice of the dual weights $\mathbf{w}^*$). Hence it follows that, for all simple compact groups G, topological charges of the torus T_{G^*} of the adjoint dual group G^* reproduce all weights of the group G.

Usually one attempts to describe all weights of the group G in terms of the group G itself. As was shown in Ref. 17, for groups $G=G_2, E_8, C_n, adA_n$, adB_n, adD_n, and $adE_{6,7}$, the lattice of all possible topological charges satisfies $\mathbb{L}_t\supseteq\mathbb{L}_w$. Consequently, these groups allow a topological interpretation of all their ''quantum numbers'' in terms of these groups. For groups G and G' such that $\mathbb{L}_t^G=\mathbb{L}_w^{G'}$ there is also a possibility of topologically interpreting all ''quantum numbers'' of group G' in terms of topological charges of the group G. For example, $\mathbb{L}_{\min}^{G_2}=\mathbb{L}_w^{A_2}$. For other groups G such that $\mathbb{L}_t^{adG}=\mathbb{L}_w^G$ (for example, $G=A_n, B_n, D_n$, and $E_{6,7}$), not all sets of topological charges corresponding to a specific representation (for example, quark representations of the groups A_n or spinor representations of the groups B_n and D_n) correspond to exact (single-valued) representations.

Now consider a two-dimensional chiral model on the torus T_G. The corresponding action $\mathscr{S}$ has the form

$$\mathscr{S}=\frac{1}{2\alpha}\int d^2x\,\mathrm{tr}_\tau(\mathbf{g}_\nu^{-1}\mathbf{g}_\nu)=\frac{(2\pi)^2}{2\alpha}\int d^2x\,\mathrm{tr}_\tau(\mathbf{H}\cdot\boldsymbol{\phi}_\nu)^2$$
$$=\frac{(2\pi)^2}{2\alpha}B_\tau\int d^2x(\boldsymbol{\phi}_\nu)^2, \tag{35}$$

where

$$\mathbf{g}=\exp 2i\pi(\mathbf{H}\cdot\boldsymbol{\phi})\in T_G,\quad \boldsymbol{\phi}=(\phi^1,\ldots,\phi^n), \tag{36}$$

n is the rank of the group G. Here turns up the effective metric g_{ik} from Eq. (33) generated by the system of weights $\{\mathbf{w}\}_\tau$ of the τ-representation.

It is convenient to introduce to this model a normalized trace in order to eliminate the constant B_τ from the effective metric:

$$\mathrm{Tr}=\frac{\mathrm{tr}_\tau}{B_\tau}. \tag{37}$$

Then $\mathscr{S}$ takes its usual form without the factor B_τ. The corresponding equations

$$(\partial_\nu)^2(\mathbf{H}\cdot\boldsymbol{\phi})=0 \tag{38}$$

have classical vortex-like solutions in the region $R>r>a$ similar to solutions of Eq. (20):

$$\boldsymbol{\phi}(\mathbf{x})=\frac{1}{\pi}\mathbf{Q}\tan^{-1}\left(\frac{y-y_0}{x-x_0}\right),\quad \mathbf{Q}\in\mathbb{L}_\tau^{-1}. \tag{39}$$

The topological interpretation of all quantum numbers of groups G such that $\mathbb{L}_\tau^{-1}\supseteq\mathbb{L}_w$ is based on these solutions.[17] The energy of these vortices is also logarithmically divergent:

$$E=\frac{(2\pi)^2}{2\alpha}\int(\partial_\mu\boldsymbol{\phi})^2d^2x=\frac{2\pi}{2\alpha}(\mathbf{Q})^2\ln\frac{R}{a}. \tag{40}$$

This leads to a logarithmic interaction between vortices with different noncollinear topological charges, in contrast to the case of tori of type T_L:

$$E=\frac{1}{2\alpha}(\mathbf{Q}_1\cdot\mathbf{Q}_2)2\pi\ln\left(\frac{|\mathbf{x}_1-\mathbf{x}_2|}{a}\right). \tag{41}$$

This interaction depends on the mutual orientation of vector topological charges, therefore, configurations of charge lattices are of great importance. The energy of an N-vortex solution is expressed in the same form as Eq. (22), but with the usual scalar product of topological charges. Now it is clear that properties of σ-models on Cartan tori depends on the geometry of sets of weights which define a torus and on the lattice $\mathbb{L}_\tau^{-1}$. As was noted above, all root lattices related to the simple compact groups belong to four series A, D, E, and Z, which are integer-valued, given a proper selection of the scale.[20] How the critical properties of topological phase transitions in nonlinear σ-models on Cartan tori T_G depend on the latter was described in Ref. 18 (see also Ref. 15).

One-dimensional conformal σ-models on tori T_G can be considered similarly to the models on tori T_L. The only difference is that now the sum in the trace is performed over all weights of a given representation. Since instantons (or antiinstantons) do not interact with one another, there is no great difference between properties of σ-models on tori T_L and T_G so long as we consider only nonmixed configurations. But they will be different if we take into account all possible configurations.

It might be interesting to list all compact spaces besides tori for which the homotopic group π_1 can be an integer lattice of rank $n\geqslant 2$. In such spaces, topological excitations may have new properties. For example, in nonlinear σ-models on the torus T^2, there exist knotted vortex configurations corresponding to so-called ''toric'' knots.[13] In particular, a vortex with topological charge $\mathbf{q}=(2,3)$ corresponds to a ''trefoil'' knot.

But such knotted vortices can exist only in d-dimensional ($d\geqslant 3$) physical spaces. In spaces $\mathbb{R}^d$ with $d\leqslant 2$ they can exist only in the isotopic space $\mathscr{M}$, since the torus T^2 cannot be embedded in $\mathbb{R}^{1,2}$. For this reason, Lord Kelvin's conjecture can be realized in the case of low dimensionality only in the isotopic space.

It is noteworthy that there are spaces with π_1 equal to finite cyclic groups. They can also generate vortex excitations; their topological charges will be defined modulo these groups. These cases, however, are not discussed in this paper.

4. MANIFOLDS WITH $\pi_2=\mathbb{L}$

Now let us proceed to compact manifolds $\mathscr{M}$ with vector homotopic group π_2. Manifolds of this type that are most widely known can be divided into two large classes:

(1) Homogeneous spaces of simple Lie groups;

(2) Hodge submanifolds of complex projective spaces $\mathbb{CP}^n$ differing from $\mathscr{M}$ in the previously discussed class. We are more interested in homogeneous spaces because they can have group π_2 with topological charges belonging to weight lattices (or their sublattices) of the corresponding Lie groups. In finding out solutions of the instanton type with topological charges $\mathbf{q}\in\mathbb{L}_r$, the most important spaces are complex homogeneous spaces. The basic classification theorem about such manifolds asserts that all of them can be treated as fiber bundle spaces over some flag spaces F:

$$\mathscr{M}\overset{T}{\to}F,$$

where fiber T is a parallelizable space. Moreover, the space T should be a complex torus if $\mathscr{M}$ is simply connected or Kahler. But if both conditions are satisfied, then $\mathscr{M}$ should be a flag space F (Refs. 21 and 22). Since $\pi_2(T)=0$, flag spaces F are most important from the viewpoint of the existence of instanton solutions. Among these spaces, there are maximal flag spaces $F_G=G/T_G$ with $\pi_2(F_G)=\mathbb{L}_v$ (Refs. 17 and 23).

Chiral models on F_G were analyzed by Perelomov and Prati,[23,24] who showed that, since $\pi_2(G/T_G)=\pi_1(\tilde{T}_G)\neq 0$ and there is a complex structure on G/T_G, the corresponding equations have holomorphic instanton solutions. But the analysis of topological charges did not take account of the possible vector structure of the corresponding groups $\pi_2(F_G)$. The underlying assumption of the authors was that corresponding topological charges, being convolutions of certain 2-forms with different 2-cycles, were scalar integers. Since $\pi_2(G/T_G)=\mathbb{L}_v$, one can treat the corresponding topo-

logical charges as isovectors. In order to see how they can be realized, one should turn to the σ-model on G/T_G. Its action has the form[23]

$$\mathcal{A}[u]=\frac{1}{2g^2}\int d^2x g_{a\bar{b}}(u,\bar{u})\partial_\mu u^a \partial_\mu \bar{u}^b,$$

$$a,b=1,\ldots,p, \quad 2p+n=\dim G. \tag{42}$$

Here u^a are local complex coordinates on G/T_G (these are also field variables of the theory), and metric $g_{a\bar{b}}(u,\bar{u})$ is an integer 2-form from the homological class ϱ, where ϱ is an isovector equal to half the sum of all positive roots of the Lie algebra $\mathcal{G}$:

$$\varrho=\frac{1}{2}\sum_{a>0}\mathbf{r}_a=\sum_{i=1}^{n}\bar{\mathbf{w}}_i,$$

and $\bar{\mathbf{w}}_i$ are the fundamental weights of the group G. This is the metric which is Einsteinian on G/T_G (Refs. 24 and 25), i.e.,

$$g_{a\bar{b}}(u,\bar{u})=kR_{a\bar{b}}(u,\bar{u}), \quad k=1/2, \tag{43}$$

where $R_{a\bar{b}}(u,\bar{u})$ is the Ricci tensor of space F_G. This selection of the invariant metric guarantees that the action (42) is renormalizable.[24,26]

On holomorphic fields $u^a(z)$, $z\in\mathbb{C}=\mathbb{R}^2$, the action $\mathcal{S}$ is formally identical (up to a numerical factor) with the topological invariant

$$Q=\frac{1}{2}\int d^2x g_{a\bar{b}}(u,\bar{u})\varepsilon_{\mu\nu}\partial_\mu u^a \partial_\nu \bar{u}^b=\frac{1}{2}\int g_{a\bar{b}}du^a\wedge d\bar{u}^b. \tag{44}$$

As is well known,[21,24] all integer 2-forms on $F_G=G/T_G$

$$\Omega=\Omega_{a\bar{b}}du^a\wedge d\bar{u}^b$$

can be expanded (modulo exact forms) in the basis of 2-forms ω_i, $i=1,\ldots,n$, parametrized by the fundamental weights $\bar{\mathbf{w}}_i$:

$$\Omega=\sum_{1}^{n}c_i\omega_i, \quad c_i\in\mathbb{Z}. \tag{45}$$

The 2-form corresponding to the metric $g_{a\bar{b}}(u,\bar{u})$ has a similar expansion with coefficients $c_i=1$. On the other hand, there is a space of 2-cycles γ on G/T_G which is dual to the space of 2-forms. All 2-cycles can be presented as linear combinations of Schubert cells of the minimal dimensionality (real dimensionality 2 or complex dimensionality 1), whose number equals the rank of the group G (Ref. 21). Since $\pi_1(F_G)=0$ and, according to the Gurevich theorem, $\pi_2(F_G)=H_2(F_G,\mathbb{Z})=\mathbb{L}_v$, one can introduce a vector structure on the space of 2-cycles similar to the structure in Eq. (9) and select a parametrization of 2-cycles by the lattice $\mathbb{L}_v$. In this case, basis cycles $\{\gamma_i\}$ are parametrized by simple dual roots $\mathbf{r}_i^v$:

$$\gamma\to\boldsymbol{\gamma}=\sum_{1}^{n}n_i\gamma_i\mathbf{r}_i^v, \quad n_i\in\mathbb{Z}. \tag{46}$$

The convolution of 2-forms ω_i and 2-cycles γ_k dual to them, i.e., the integral of the 2-form ω_i over the 2-cycle γ_k, equals

$$(\omega_i\circ\gamma_k)=\int_{\gamma_k}\omega_i=\delta_{ik}.$$

Since the 2-form ϱ, which determines the action and the corresponding topological charges, is fixed, the vector nature of the space of 2-cycles implies that the space of topological charges $\mathbf{Q}$ has a vector nature too:

$$\mathbf{Q}=\sum_{i=1}^{n}q_i\mathbf{r}_i^v=\left(\varrho\circ\sum_{1}^{n}q_i\gamma_i\right)\mathbf{r}_i^v, \quad q_i\in\mathbb{Z}. \tag{47}$$

Hence we conclude that topological charges corresponding to various independent cycles cannot be added as scalars, in particular, they cannot cancel one another. In calculations, it is advisable to note the similarity between the topological cell structure of torus T^n and flag F_G in their first nontrivial cell complexes: for the torus T^n this complex equals a bouquet of circles $T^n=S_1^1\vee\ldots\vee S_n^1$ (Ref. 13), whereas for the space of flags it looks like a bouquet of 2-spheres $M=S_1^2\vee\ldots\vee S_n^2$.

For example, for $G=SU(n+1)$, one can select elements of the upper diagonal line closest to the main diagonal on the 2-spheres as local coordinates. Then the instantons corresponding to these 2-spheres are simple Belavin–Polyakov instantons.[4] For the action $\mathcal{S}$ of n instantons with topological charges $q_1,\ldots,q_n$ corresponding to n different 2-cycles, we obtain

$$\mathcal{S}=\frac{1}{2g^2}\sum_{1}^{n}|q_i|. \tag{48}$$

This expression looks like the similar expression for one-dimensional instantons in Eq. (28).

Thus, we have shown that instanton solutions of two-dimensional chiral models on maximal flag spaces $F_G=G/T_G$ can realize isovector topological charges belonging to $\pi_2(G/T_G)=\mathbb{L}_v$, so they can realize all quantum numbers of groups G such that $\mathbb{L}_w\subseteq\mathbb{L}_v$. Other groups require additional factorization of the space F_G.

5. THREE-DIMENSIONAL TOPOLOGICAL EXCITATIONS

Topological excitations of vortex and instanton types, i.e., those related to the groups $\pi_1(T_G)$ and $\pi_2(F_G)$, can also exist in a three-dimensional space. Vortex excitations can form closed or open lines. In the first case, they can form knotted configurations in both physical and isotopic spaces. Their energy in the case of open vortex lines is $E\sim L\ln R/a$, where L is the length of the line. Topological excitations associated with topological charges $\mathbf{Q}\in\pi_2(G/T_G)$ can exist in the generalized Higgs–Salam–Weinberg (HSW) models in the form of three-dimensional particle-like monopoles. They should have a finite energy if gauge field are included, or their energy can diverge without these fields. It is noteworthy that, since $\pi_2(F_G)=\mathbb{L}_v$, which does not contain weights of fundamental quark representations in the case of $G=SU(n)$, the latter cannot exist in the form of topological excitations on F_G. Probably, this fact is

related to the quark confinement. Additional factorization of F_G or additional spontaneous symmetry breaking is needed for a topological interpretation of quarks in the HSW model.[27]

Consider as an example of three-dimensional topological excitations related to the homotopic group π_2 a three-dimensional chiral model on the sphere S^2 with action[19]

$$S=\frac{1}{2\alpha}\int d^3x(\partial_\mu \mathbf{n})^2, \quad \mathbf{n}\in S^2. \tag{49}$$

The following equations have solutions of the instanton type, and the latter can be expressed in the form

$$n_i(x)=\frac{x_i}{r}\theta_r(r-a), \tag{50}$$

where $\theta_r(r)$ is the regularized step function. Its topological charge is

$$Q=\frac{1}{4\pi}\int dS, \tag{51}$$

where the integral $\int dS$ is performed over the sphere S^2 in the isotopic space, and the energy

$$E=\frac{1}{2\alpha}\int d^3x(\partial_\mu \mathbf{n})^2=\frac{1}{\alpha}4\pi\int_a^R dr\simeq\frac{4\pi}{\alpha}R \tag{52}$$

grows linearly with R. This means that the instantons in this model can be bound in neutral complexes, as quarks in quantum chromodynamics, and can be detected only at small separations between them. All infrared (long-wave) properties should not be affected by their existence. The σ-models, defined on other manifolds with $\pi_2=\mathbb{L}$, have similar properties.

Using an exact homotopic sequence for stratification $G\overset{T_G}{\rightarrow}F_G$, one can prove[14] that

$$\pi_k(G)=\pi_k(G/T_G)=\pi_k(G,T_G), \quad k=3,4,\ldots,$$

where $\pi_k(G,T_G)$ are the corresponding relative homotopic groups. This means that for $k>2$ the flag space F_G has the same homotopic properties as the group G. For this reason, the nonlinear σ-model on the flag space F_G may have knotted solutions associated with the group $\pi_3(F_G)=\pi_3(G)=\mathbb{Z}$ similar to those suggested recently for the four-dimensional Yang–Mills theory.[28]

6. DISCUSSION

We have shown that quantum numbers of groups can be represented in the form of topological charges of vortices and instantons. But instantons, in the strict sense of this term, do not interact with one another in general. Thus, instantons alone cannot be used as a model of real interacting particles. Their realization requires topological excitations interacting as two-dimensional vortices, but through three-dimensional potentials of the Coulomb and Yukawa types (or logarithmically):

$$E=\mathbf{q}_1\cdot\mathbf{q}_2V(r), \quad V_{\mathrm{C}}(r)=\frac{1}{r}, \quad V_{\mathrm{Y}}(r)=\frac{e^{-mr}}{r}.$$

In a three-dimensional space, such excitations can be generated in a conformal (hence nonlocal) nonlinear σ-model on manifolds with $\pi_2=\mathbb{L}$ (Ref. 29). With this end in view, theories like the quantum chromodynamics or σ-models with additional gauge fields are applied to local theories. The latter procedure can be executed by various means. For example, in one version of this procedure nonlinear gauge σ-models can be considered on fixed F_G. In another, more sophisticated version, F_G is treated as one element of the family of such conjugate subspaces in the group G. In this case, all these subspaces are equivalent if there is no fixing potential for F_G. This means that the fields of σ-models at different points can assume values in different subspaces F_G, moreover, the gauge fields can occur as conjugation transformations relating these subspaces. This approach demands further development.

In conclusion, note that the manifolds T_G and F_G should be viewed as indispensable components of any theory like the Grand Unification Theory or the theory of strings, which attempts to develop a physically transparent (in particular, one based on topological concepts) interpretation of elementary particles.

The author is grateful to G. Volovik for helpful discussions of some topics considered in the paper and to V. Gurarii for information on the literature about complex uniform spaces.

The work was supported by the Russian Fund for Fundamental Research (Grants 96-02-17331-a and 96-15-96861).

*)E-mail: bulgad@itp.ac.ru

[1] A. M. Polyakov, JETP Lett. **20**, 194 (1974).
[2] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[3] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[4] A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 114 (1975).
[5] D. Mattis, *The Theory of Magnetism*, Harper & Row Publishers (1965).
[6] F. Klein, *Vorlesungen über die Entwicklung der Mathematik im 19-zten Jahrhundert*, Chelsea Publ. Co., New York, 1967 [Russian transl., Vol. I, Nauka, Moscow, 1989].
[7] W. H. Thomson, Trans. Roy. Soc. Edin. **25**, 217 (1869); P. G. Tait, *On Knots*, I, II, III, Scientific Papers, Cambridge University Press, (1900); M. F. Atiyah, *The Geometry and Physics of Knots*, Cambridge University Press (1990).
[8] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press (1989).
[9] P. G. De Gennes, *Superconductivity of Metals and Alloys*, W. A. Benjamin, Inc. (1966).
[10] M. D. Frank-Kamenetskii, A. V. Lukashin, and A. V. Vologodskii, Nature (London) **258**, 398 (1975); M. D. Frank-Kamenetskii and A. V. Vologidskii, Usp. Fiz. Nauk **134**, 641 (1981) [Sov. Phys. Usp. **24**, 679 (1981)].
[11] G. E. Volovik, *Exotic Properties of Superfluid ^3He*, World Scientific, Singapore (1992).
[12] R. Rajaraman, *Solitons and Instantons*, North-Holland, Amsterdam–New York–Oxford (1982).
[13] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry—Methods and Applications*, parts I–III, Springer–Verlag, New York, 1984, 1985, 1990 [Russ. original, Nauka, Moscow (1979)].
[14] N. Bourbaki, *Groupes et algebres de Lie*. Chapters IV–VI, Hermann, Paris (1968); Chapters VII, VIII, Hermann, Paris (1975).
[15] S. A. Bulgadaev, Nucl. Phys. B **224**, 349 (1983); Phys. Lett. B **166**, 88 (1986).
[16] S. A. Bulgadaev, Phys. Lett. A **125**, 299 (1987).

[17] S. A. Bulgadaev, JETP Lett. **63**, 796 (1996).
[18] S. A. Bulgadaev, JETP Lett. **63**, 780 (1996); E-prints archives, hep-th/990691, hep-th/9909031.
[19] A. M. Polyakov, *Gauge Fields and Strings*, Harwood Academic Publishers (1987).
[20] G. H. Conway and N. J. A. Sloane, *Sphere Packing, Lattices and Groups*, Vol. I and II, Springer-Verlag (1988).
[21] N. E. Hurt, *Geometric Quantization in Action*, D. Reidel Publ. Company (1983).
[22] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. II, Interscience Publishers (1969).
[23] A. M. Perelomov, Usp. Fiz. Nauk **134**, 577 (1981) [Sov. Phys. Usp. **24**, 649 (1981)].
[24] A. M. Perelomov and M. C. Prati, Nucl. Phys. B **258**, 647 (1985).
[25] D. S. Freed, *Flag Manifolds and Infinite Dimensional Kahler Geometry*, in *Proc. of Intern. Conf.* (1985).
[26] E. Witten, Phys. Rev. D **16**, 2991 (1977).
[27] T. Vachaspati, Phys. Rev. Lett. **76**, 188 (1996).
[28] L. D. Faddeev and A. Niemi, E-print archives, hep-th/9705176.
[29] S. A. Bulgadaev, Landau Institute Preprint 29/05/97 (1997).

Translation provided by the Russian Editorial office.

Temperature dependence of superradiance intensity

E. V. Orlenko*) and B. G. Matisov

St. Petersburg State Technical University, 195251 St. Petersburg, Russia
(Submitted 8 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1148–1160 (October 1999)

Using the idea of exchange interaction in a system of two-level atoms participating in a superradiance process, we derive from first principles the superradiance Hamiltonian of such a system, which is found to be analogous to the Heisenberg Hamiltonian. We consistently calculate the coupling constant of the interaction that leads to the emergence of a superradiance state in the system. We also predict the existence of isospin excitations in the superradiance state, whose presence reduces the intensity of the corresponding superradiance pulse. Finally, we calculate the temperature dependence of the intensity of the superradiance pulse and find it be analogous to the Bloch $T^{3/2}$-law for spin systems. © *1999 American Institute of Physics.* [S1063-7761(99)00310-8]

1. INTRODUCTION

Many light-induced phase transitions can take place in the simplest quantum systems, such as ensembles of two-level atoms interacting through radiation, the electrostatic field, the dipole–dipole interaction, photons, etc.[1] The study of such nonequilibrium phase transitions has shown that there is a profound analogy between such processes and second-order phase transitions[2,3] that occur in a spin system in the presence of an interaction between the spins whose coupling constant (exchange Coulomb interaction) exceeds the energy of thermal motion that leads to disorientation of the spins. In this case, spontaneous alignment of the spins emerges in the system, which macroscopically manifests itself in remanent magnetization. Phase transitions in quantum-optics systems are also caused by an interaction that leads to the emergence of order in the orientation of so-called energy spins (isospins). Dicke superradiance is one example of the manifestation of such transitions.[4] Naturally, this brings phase transitions and the superradiance effect closer together. Furthermore, as shown by Andreev *et al.*[1] in their analysis of the state of polariton generation in an open medium of two-level atoms interacting via the Stokes field in Raman scattering of light, the onset of a superradiance regime is actually a phase transition.

Interaction via a reradiation field is one of the most universal types of interaction in such systems. However, in a medium consisting of two-level atoms with constant dipole moments, the collectivization of the ensemble of atoms can occur due to static dipole–dipole interaction. Additional interaction via phonons emerges in crystals.

The analogy with equilibrium second-order phase transitions in magnetic systems is due to the fact that the Hamiltonian describing the behavior of two-level atoms in a radiation field is similar, if we allow for interatomic interaction, to the Heisenberg Hamiltonian for spin systems. Many attempts have also been made in quantum optics to reduce the Hamiltonian directly to the Heisenberg Hamiltonian.[5] However, the coupling constant, which is a coefficient in the scalar product of isospins, is assumed to be either the coupling constant of the interaction via the reradiation field or simply the dipole–dipole coupling constant.

When drawing an analogy with magnetic systems, it must be noted that the direct magnetic dipole interaction exists by itself in magnetic systems, but there it does not lead to spontaneous spin alignment, since its coupling constant is small compared to the energy of thermal motion. Hence it is the exchange Coulomb interaction of bound d-electrons, and to a lesser extent, of free electrons that leads to the emergence of magnetic order. The coupling constant of this interaction can be calculated from first principles,[6] and it coincides, to order of magnitude, with the chemical binding energy,[7] so that allowing for this constant can obstruct the process of spin disorientation. It is therefore unlikely that the interaction of atoms via the reradiation field can by itself lead to alignment of energy spins, since the coupling constant of this interaction is small[4] compared to the equilibrium temperature of the system of atoms, to the equilibrium thermal radiation emitted by the system, and finally to the intensity of the pump field, which in radiative systems acts as a variable parameter and is the analog of temperature.

The electric dipole interaction of atoms plays an important role in generating a superradiance state even when the atoms have only dipole moments of transitions. Zaĭtsev *et al.*[8] used a semiclassical approach to study the effect of the Coulomb interaction on the superradiance of a system of two-level atoms and found that the Coulomb interaction leads to coherent transfer of excitation from atom to atom, which results in approximate spatial homogeneity of inversion in a chain of atoms. Thus, it can be stated (and this statement is later proved) that the Coulomb dipole–dipole interaction not only does not destroy the superradiance state, but instead facilitates the generation of such a state, contrary to common belief. It is also shown that the Coulomb interaction must be taken into account in all systems with a small Fresnel number, since in such systems the time of ''exchange'' of excitations, τ_c, is much shorter than τ_R, the superradiance time.

Attempts to account for the Coulomb interaction for a small number of atoms were made in earlier work,[9,10] where it was found that the electrostatic interaction in the semiclassical approximation has no effect on the dynamics of superradiance, but leads to phase modulation. However, a more thorough analysis of the dynamics of superradiance with allowance for the Coulomb interaction[8] has revealed that the luminescence intensity exhibits distinct oscillations in its time dependence, which might be related to the propagation of wavelike excitations in a system.

In the present paper we derive, from first principles, the system Hamiltonian analogous to the Heisenberg Hamiltonian in the theory of magnetism. In doing so we allow for the dipole Coulomb interaction of the atoms comprising the system and the interaction via the reradiation field. We then use the Hamiltonian to study wave excitations in the system analogous to spin waves in a ferromagnet or antiferromagnet. We find that precisely these excitations lead to a characteristic temperature dependence of the superradiance intensity. We also calculate the critical temperature at which superradiance disappears and a second-order phase transition occurs in the system. Long-range order forces, which emerge in a Boltzmann gas when long-wavelength isospin excitations (predicted in the present paper) propagate in it, must be examined by using quantum mechanical principles. Taking such an *ab initio* approach in this paper, we are not only able to establish the isospin contributions to the interaction between the atoms, but also to understand the physics of these contributions. The mechanism of formation of isospin excitations and the dispersion laws for them are obtained in a natural manner on the basis of operator models.

2. HAMILTONIAN OF A COOPERATIVE SYSTEM

We examine the energy states of a system consisting of two noninteracting atoms in each of which the active levels have energies E_1 and E_2. The state of each atom is described by a spinor (see Ref. 4), where $\chi_{(\uparrow)}=\binom{1}{0}$ means that the atom is in the energy state with $E=E_2$ and $\chi_{(\downarrow)}=\binom{0}{1}$ means that the atom is in the state with $E=E_1$. Such a state of the system of two noninteracting atoms can be written as a simple product of the spinors $\chi_{\rm I}$ and $\chi_{\rm II}$ of the corresponding atoms I and II:

$$\chi_{\rm I,II}=\chi_{\rm I}(i)\chi_{\rm II}(j), \tag{1}$$

where the indices i and j run through the values 1 and 2 corresponding to the states $\downarrow$ and $\uparrow$.

If, following Dicke,[4] we introduce the operator

$$\hat{S}_z=\frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

the Hamiltonian of the noninteracting system can be written

$$\hat{H}_0=E(\hat{S}_{\rm Iz}+\hat{S}_{\rm IIz}), \tag{2}$$

where $E=E_2-E_1$. Then $\hat{H}_0\chi_{\rm I,II}=E(\hat{S}_{\rm Iz}+\hat{S}_{\rm IIz})\chi_{\rm I,II}$. To some of the energy states of the system, say with $E=E_1+E_2$, there can correspond two functions, $\chi_{\rm I,II}=\chi_{\rm I}(\uparrow)\chi_{\rm II}(\downarrow)$ and $\chi'_{\rm I,II}=\chi_{\rm I}(\downarrow)\chi_{\rm II}(\uparrow)$.

The system is thus degenerate. The functions $\chi_{\rm I,II}$ and $\chi'_{\rm I,II}$ are orthogonal to one another (as can be easily verified).

We now allow for interaction between the atoms. The mechanism of this interaction can be any one of those described above, provided that it is consistent with the experimental situation. To describe this interaction we introduce an operator $\hat{V}_{\rm I,II}$. The coupling constant of any of the types of interaction discussed above is small compared to the difference of the energies E_2 and E_1, so that ordinary degenerate perturbation theory is applicable. The perturbation introduces a well-known correction to the total energy of the system:[7] $\varepsilon^{(1)}=K\pm A$, where in our case $K=\langle\chi_{\rm I,II}|\hat{V}_{\rm I,II}|\chi_{\rm I,II}\rangle$ and $A=\langle\chi'_{\rm I,II}|\hat{V}_{\rm I,II}|\chi_{\rm I,II}\rangle$, with the corresponding correct wave functions being

$$\chi_+=\frac{1}{2}(\chi_{\rm I}(i)\chi_{\rm II}(j)+\chi_{\rm I}(j)\chi_{\rm II}(i)),$$

$$\chi_-=\frac{1}{2}(\chi_{\rm I}(i)\chi_{\rm II}(j)-\chi_{\rm I}(j)\chi_{\rm II}(i)).$$

Clearly, if initially the system is in a state with $E=E_1+E_2$, the state that is advantageous when the perturbation is accounted for is

$$\chi_-=\frac{1}{2}(\chi_{\rm I}(\uparrow)\chi_{\rm II}(\downarrow)-\chi_{\rm I}(\downarrow)\chi_{\rm II}(\uparrow))$$

if $A>0$, or χ_+ if $A<0$. Initially, the energy states of the diatomic system with $E=2E_1$ and $E=2E_2$ are not degenerate, and of these states only χ_+, the symmetric state, is realized, since $\chi_-\equiv 0$.

We introduce the operator $\hat{P}_{\rm I,II}=\frac{1}{2}(1+4\hat{\mathbf{S}}_{\rm I}\cdot\hat{\mathbf{S}}_{\rm II})$, where $\hat{S}_x$, $\hat{S}_y$, and $\hat{S}_z$ are, as in Ref. 4, equivalent to the Pauli matrices with a factor of 1/2 It can easily be verified that for the symmetric state the eigenvalue of the operator $\hat{S}^2=(\hat{\mathbf{S}}_{\rm I}+\hat{\mathbf{S}}_{\rm II})^2$ corresponds to the total isospin of the system, which is equal to 1. For the antisymmetric state of the system the total isospin is 0. Similarly, the eigenvalues of the operator $\hat{P}_{\rm I,II}$ are $+1$ and -1 for the χ_+ and χ_- states, respectively. Then the operator of the interaction energy of two atoms that explicitly allows for the isospin states of the atoms can be written

$$\hat{H}_{\rm int}=K+A\hat{P}_{\rm I,II}=K+\frac{A}{2}(1+4\hat{\mathbf{S}}_{\rm I}\cdot\hat{\mathbf{S}}_{\rm II})$$

$$=K+\frac{A}{2}+2A\hat{\mathbf{S}}_{\rm I}\cdot\hat{\mathbf{S}}_{\rm II}. \tag{3}$$

The total Hamiltonian of a system consisting of N atoms with allowance solely for pairwise interactions can be written

$$\hat{H}=\sum_{k,l}(\hat{H}_{0,k}+\hat{H}_{{\rm int},k,l})$$

$$=E\sum_k\hat{S}_{kz}+\sum_{k<l}J_{kl}\hat{\mathbf{S}}_k\cdot\hat{\mathbf{S}}_l+\frac{N}{2}E_0, \tag{4}$$

where $E_0=K+A/2$, and $J_{kl}=2A_{kl}$, with k and l the numbers of the atoms.

We have obtained a Hamiltonian analogous to the Heisenberg Hamiltonian of the theory of spin systems. The Hamiltonian (4) contains an exchange parameter J_{kl} (just like the Heisenberg Hamiltonian), since it is the exchange interaction that is the result of the symmetry of the wave functions of the initial system and therefore is closely related to alignment (in our case the alignment of isospins). However, in contrast to the Heisenberg Hamiltonian, the term responsible for the exchange interaction is positive. The reason is that the sign of of the ''exchange'' term is directly linked to the symmetry of χ_+. The corresponding sign in the analogous expression in the Heitler–London problem is related just to the symmetry of the space part of the complete wave function, while the symmetry of the spin parts, in accordance with Young tableaux,[7] is supplementary.

Now we discuss the relationship between the Hamiltonian (4) and Dicke states. These states,[4] which are characterized by the total isospin of the system (the so-called cooperative quantum number r) and the projection of the total spin on the z axis (the polarization quantum number m), are eigenstates of the operator (4), so the matrix of this operator calculated for the Dicke states is diagonal. In our problem, states with a definite polarization number m cease to be degenerate, since degeneracy in the total isospin of the system is lifted due to the interaction described by the second term in (4). Dicke states $|r,m\rangle$ with different values of r but equal values of m now correspond to different sublevels of the energy:

$$\hat{H}|r,m\rangle=\left\{Em+\frac{J}{2}\left[r(r+1)-\frac{N}{2}\left(\frac{1}{2}+1\right)\right]\right\}|r,m\rangle, \quad (5)$$

where the operator $J\Sigma_{k<l}\hat{\mathbf{S}}_k\cdot\hat{\mathbf{S}}_l$, $J=J_{kl}$, in (4) is written

$$\frac{J}{2}\left[\left(\sum_k \hat{\mathbf{S}}_k\right)^2-\sum_k \hat{\mathbf{S}}_k^2\right].$$

In contrast to magnetic spin systems, the main contribution to the energy of the system is provided by the first term, since $|E|\gg|J|$, and in the absence of preliminary pumping the system of the atoms is in a state with energy E_1, which corresponds to the Dicke state $|r=N/2, m=-N/2\rangle$. However, of all the possible states with $m=0$ the most energetically advantageous is $|r=0, m=0\rangle$, as Eq. (5) clearly shows.

This state corresponds to a shallow well into which all the paired atoms fall and land on a single level, irrespective of whether they were initially bosons or fermions. Dicke found that such a completely antisymmetric state is stable, since in this case the system does not emit radiation, being in a kind of ''frozen'' state.

An avalanche discharge of photons can be initiated by interactions with an external pump field, as described by the operator

$$\hat{W}=-\mathbf{P}\cdot(\mathbf{e}_x\hat{S}_x+\mathbf{e}_y\hat{S}_y)=-\mathbf{P}\cdot\sum_k(\mathbf{e}_x\hat{S}_{kx}+\mathbf{e}_y\hat{S}_{ky}), \quad (6)$$

where $\hat{S}_x=\Sigma_k\hat{S}_{kx}$, $\hat{S}_y=\Sigma_k\hat{S}_{ky}$, and $\mathbf{P}$ is the polarization vector. The nonvanishing matrix elements of this operator in Dicke states correspond to $m\to m\pm1$ transitions. Hence this interaction enables the system to run through all possible values of m, and the system, after being pumped to the state with $m=N/2$, can easily find itself in a state with $m=0$. But this latter state can be realized as a linear combination of states with different values of $r(|m\rangle=\Sigma_r C_{rm}|r,m\rangle)$, so that the probability of the system being taken out of the ''frozen'' state and emitting radiation with intensity $I\propto N^2$, as predicted by Dicke, is $P_{\rm rad}=1-|C_{00}|^2$. The presence of such a ''frozen'' state can explain the time lag preceding the superradiance regime.

Thus, the Hamiltonian (4) derived in the present paper describes both the general case of cooperative systems and the special case of a superradiance state.

3. ISOSPIN EXCITATIONS. TEMPERATURE DEPENDENCE OF SUPERRADIANCE

We now turn to a system of atoms cooled to low temperature, with the mean free path that the atoms travel in the time τ_R of superradiance being much less than the characteristic interatomic distance: $v_T\tau_R\ll n^{-1/3}$, where n is the concentration of the atoms, and v_T is the thermal velocity of the atoms. Then we can assume that the ''neighbors'' of these atoms are essentially unchanged, and following Ref. 8, the system can be considered a chain of interacting atoms. In a solid such a model is more than justified.

We attribute the spontaneous deviation of the polarization vector $\mathbf{P}$ from its quasiequilibrium position at the moment at which avalanche discharge of superradiance begins to excitations that exist in the system. We also assume that any thermal effects are the cause of relaxation in the system of isospins, when both ''neighbors'' go to the state with $E=E_1$. We denote this state by

$$|\xi_n\rangle=|\chi_{(\downarrow)n-1}\chi_{(\downarrow)n}+\chi_{(\downarrow)n}\chi_{(\downarrow)n+1}\rangle. \quad (7)$$

The Hamiltonian describing the excitation is chosen in the form (4) with allowance for (3), from which we subtract the energy of the coherent ground state,

$$\hat{V}_{\rm exc}=\sum_i \hat{H}_{{\rm int},i,i+1}-\varepsilon_0=A\sum_i(\hat{P}_{i,i+1}-1), \quad (8)$$

where $\varepsilon_0=(N/2)E_1+(N/2)E_2$. We can easily show that

$$\hat{P}_{k,k+1}|\chi_{k(\uparrow)}\chi_{k+1(\uparrow)}\rangle=|\chi_{k(\uparrow)}\chi_{k+1(\uparrow)}\rangle,$$

$$\hat{P}_{k,k+1}|\chi_{k(\uparrow)}\chi_{k+1(\downarrow)}\rangle=|\chi_{k(\downarrow)}\chi_{k+1(\uparrow)}\rangle,$$

and so forth. A possible state with one transferred spin in the chain of atoms can be written as a linear combination of the states $|\xi_n\rangle$:

$$|\xi\rangle=\sum_n C_n|\xi_n\rangle. \quad (9)$$

If we denote the energy of such an excitation by $\hbar\omega$, we can write $\hat{V}_{\rm exc}|\xi\rangle=\hbar\omega|\xi\rangle$. Applying the operator (8) directly to the function (9) and allowing for the properties of the operator $\hat{P}_{ij}$, we obtain

$$\hbar\omega\sum_n C_n|\xi_n\rangle=A\sum_n(P_{n-2,n-1}-1+P_{n,n-1}-1)C_n|\xi_n\rangle,$$

$$\hbar\omega\sum_n C_n|\xi_n\rangle = A\sum_n C_n|(\xi_{n-1}+\xi_{n+1}-2\xi_n)\rangle.$$

Equating the coefficients of the ξ_n, we obtain $\hbar\omega C_n = A(C_{n-1}+C_{n+1}-2C_n)$. Now we seek a solution for the C_n as a linear combination of traveling waves:

$$\sum_k q_k \exp[i(kx_n-\omega_k t)], \quad x_n = n\rho,$$

where **k** is the wave vector. Then $\hbar\omega = 4A\sin^2(k\rho/2)$, where ρ is the mean separation of atoms in a gas or the lattice constant in a solid.

In the long-wavelength approximation $k\rho \ll 1$ we obtain

$$\hbar\omega_k = 4A\left(\frac{k\rho}{2}\right)^2 = A\rho^2k^2,$$

or the dispersion relation for isospin excitations,

$$\omega_k = (A\rho^2/\hbar)k^2,$$

which has exactly the same form as for spin waves in a solid. The Hamiltonian form of the energy of these excitations with various numbers of transferred spins can be written in terms of generalized coordinates and momenta as the sum of energies of harmonic oscillators with distinct ω_k,

$$\hat{V}_{\text{exc}} = \sum_k \left(\frac{\hat{P}_k^2}{2}+\frac{\omega_k^2\hat{x}_k^2}{2}\right), \tag{10}$$

and the energy of the state of oscillatory systems averaged over the ensemble can be written

$$u = \sum_k \hbar\omega_k\left(\overline{n_k}+\frac{1}{2}\right), \tag{11}$$

where $\overline{n_k}$ is the Planck distribution function for the number of excitation quanta in the system (pseudomagnons). The sum over k ($k = 2\pi\nu/\rho$, with ν an integer) goes from $k=0$ to a $k_{\max}$, which is related to the existence of a minimum wavelength $\lambda_{\min}$. Clearly, the relative deviation of the polarization vector from the optimum value $p_\infty = |d|N/2$ at $T=0$ is related to the propagation of excitations in the system. Thus, the extent of this deviation can be calculated by determining the total number of thermal excitations of pseudomagnon type:

$$\frac{|P_\infty - P(T)|}{P_\infty} \sim V\int \frac{d^3k}{(2\pi)^3(e^{\hbar\omega_k/T}-1)} = \frac{V}{4\pi^2}\int \frac{\sqrt{k^2}dk^2}{e^{\hbar\omega_k/T}-1}. \tag{12}$$

Let us find $\omega_{\max}$. The number of modes is equal to the number of atoms. Hence, using the dispersion relation $\omega_k = A\rho^2k^2/\hbar$, we obtain

$$(6\pi^2 n)^{2/3}\frac{A\rho^2}{\hbar} = \omega_{\max}. \tag{13}$$

We calculate the integral in (12) for two limits.

1) $T \leqslant \hbar\omega_{\max}$, the low-temperature limit:

$$\frac{|P_\infty - P(T)|}{P_\infty} \sim \frac{V}{4\pi^2}\int_0^\infty \frac{\sqrt{k^2}dk^2}{e^{\hbar\omega_k/T}-1} = \frac{3}{2}\left(\frac{T}{\hbar\omega_{\max}}\right)^{3/2}\zeta\left(\frac{3}{2}\right)\frac{\sqrt{\pi}}{2}, \tag{14}$$

where $\zeta(x)$ is the Riemann zeta function.

2) $T \geqslant \hbar\omega_{\max}$, the high-temperature limit:

$$\frac{|P_\infty - P(T)|}{P_\infty} \sim \frac{V}{4\pi^2}\int_0^{\omega_{\max}} \frac{\sqrt{\hbar\omega_k/\rho^2 A}\,d(\hbar\omega_k/\rho^2 A)}{\hbar\omega_k/T} = \frac{V}{4\pi^2}\frac{1}{(A\rho^2)^{3/2}}2\sqrt{(6\pi^2 n)^{2/3}A\rho^2}T = \frac{3T}{\hbar\omega_{\max}}. \tag{15}$$

Thus,

$$\frac{|P_\infty - P(T)|}{P_\infty} \sim \Theta(T) = \begin{cases} \dfrac{3}{2}\left(\dfrac{T}{\hbar\omega_{\max}}\right)^{3/2}\zeta\left(\dfrac{3}{2}\right)\dfrac{\sqrt{\pi}}{2}, & T \ll \hbar\omega_{\max}, \\ 3\dfrac{T}{\hbar\omega_{\max}}, & T \geqslant \hbar\omega_{\max}. \end{cases} \tag{16}$$

The characteristic critical temperature at which superradiance disappears is

$$T_C = \hbar\omega_{\max} \approx A. \tag{17}$$

The superradiance intensity is also temperature dependent, $I(T) \sim P^2(T)$. Hence, since $P(T) = P_\infty[1-\Theta(T)]$, for the temperature dependence of the superradiance intensity we have $I(T) = I_\infty[1-\Theta(T)]^2$. At low temperatures ($\Theta(T) \ll 1$), $I(T)$ can be expanded in series:

$$I(T) = I_\infty[1-2\Theta(T)] = I_\infty\left[1-3\zeta\left(\frac{3}{2}\right)\frac{\sqrt{\pi}}{2}\left(\frac{T}{\hbar\omega_{\max}}\right)^{3/2}\right]. \tag{18}$$

At high temperatures ($\Theta(T) \sim 1$) we have

$$I(T) = I_\infty\left(1-\frac{T}{A}\right)^2. \tag{19}$$

The existence of a critical temperature T_C suggests that the system can undergo a phase transition with ordering, analogous to spin ordering in antiferromagnets, where the equilibrium state is one with systematic antiparallel orientation of the spins. We believe that such ordering of isospins emerges in superradiance systems just before the discharge of the coherent pulse. Existing thermal excitations spontaneously degrade this emergent order, and thereby reduce the luminescence intensity.

4. COUPLING CONSTANT

Here we discuss the problem of allowing for the overlap of the coordinate functions of interacting identical atoms. As

is known, allowance for the symmetry properties of the function of a system of identical particles leads to a splitting of energy terms. Note that this splitting is not the result of direct interaction of atoms (dipole–dipole, etc.). According to exchange perturbation theory,[6] the correction to the interaction energy due to the permutation of atoms is given by the formula

$$\varepsilon(R)=(\Phi'|\hat{V}(R)|\Phi), \tag{20}$$

where the prime indicates the wave function of the atoms that differs from the initial wave function (unprimed) by a permutation of the atoms. The function $V(R)$, where R is the nuclei separation, is the potential energy of the interaction of the atoms in the dipole approximation. In other words, the function $V(R)$ describes the unsplit energy term of a system of two atoms.

Our numerical model, analogous to the one used in Ref. 11, is based on the following ideas. To the direct interaction of atoms we assign a function $A(R)$, present in (3), that is written as a Sutherland potential:

$$V(R)=A(R)=\begin{cases} -\dfrac{C}{R^3}, & R>\alpha, \\ \infty, & R<\alpha, \end{cases}$$

where C is a constant defined as the product of the dipole moments of the transitions between the active levels of a two-level system (see Appendix A), and α is the range of the repulsive forces between the atoms. Let T be the temperature of the gas in the adopted atomic units. Then the probability of the distance between neighboring atoms being R is given by the Boltzmann function

$$W(R)=\frac{1}{Z}\exp\left\{-\frac{V(R)}{T}\right\},$$

where Z is the normalization factor, which can be found from the condition $\int_0^\rho W(R)dR=1$, with ρ the mean separation of atoms. We find the averaged interaction energy with a modified version of Eq. (20):

$$\overline{\varepsilon}(R_0)=(\Phi'|\hat{V}(R)W(R)|\Phi). \tag{21}$$

Note that this procedure is equivalent to averaging with a density matrix in the coordinate representation, with the exchange overlap not accounted for in the normalization factors, since the corresponding contributions contain integrals of rapidly oscillating functions. The state vector of the relative motion of atoms can be written

$$|\Phi)=e^{ikz}+f(\Theta)\frac{e^{ikR}}{R},$$

where $z=R\cos\Theta$ is the position measured along the axis connecting the atoms, $f(\Theta)$ is the scattering amplitude determined by the direct interaction of atoms, i.e., by the potential V, and $k=p/\hbar$ is the wave number (p is the momentum of relative motion of the atoms). The state vector in which the permutation of atoms is taken into account can be written

$$|\Phi')=e^{-ikz}+f(\pi-\Theta)\frac{e^{ikR}}{R}.$$

Using these definitions, we obtain an expression for the desired exchange energy (see Appendix B):

$$\overline{\varepsilon}=\frac{bT}{2kZ}\mathrm{Im}(I_2(k)+2\overline{f}I_3(k)), \tag{22}$$

where $\overline{f}$ is the scattering amplitude averaged over the angular variables (it is of order α), $b=C/T$, and

$$Z=\int_\alpha^\rho \exp\left\{\frac{b}{R^3}\right\}R^2dR.$$

If $ka>1$, $\rho\gg\alpha$, and $\alpha>a$ (here a is the Bohr radius, which is unity in atomic units), using the approximate values of the integrals, we obtain

$$\overline{\varepsilon}=\pm\frac{3b^2T}{k\alpha^6}\frac{1}{R_0}\exp\left\{\frac{4}{3}kR_0-\frac{b}{\alpha^3}\right\}\left(1+\frac{2\overline{f}}{R_0}-\frac{1.5}{\alpha}\right)=\pm J, \tag{23}$$

where J is coupling constant in (4).

For atomic hydrogen at close to room temperature $T=10^{-14}$ erg ($\sim$100 K), the constant $b=5\times10^3c$ (where $c=0.3$) and $k\approx1$ (the values are given in atomic units). The radius of the potential barrier in the Sutherland potential is usually chosen within the range of the van der Waals minimum and is approximately $\alpha\approx5$, and $R_0\approx10$. For such values Eq. (23) yields $J=3C/R_0^3\approx3T$. When the temperature increases by a factor of 10, the saddle point in the integrals winds up inside the barrier, i.e., $R_0<\alpha$, and the calculations become invalid.

The increase of the coupling constant for the atomic interaction due to exchange effects is a consequence of the interference redistribution of the concentration of atoms in the gas that takes place in such a way that the probability of the atoms that are coupled via the dipole interaction coming into ''coherent'' contact rises sharply.

5. CONCLUSION

At present there is no doubt that an increase in the superradiance temperature of a system degrades the characteristics of emission by reducing the luminescence intensity, but a consistent theoretical description of this phenomenon has yet to be found.[12] We have proposed a model that encompasses, at least in principle, both the ways in which the atoms interact in a superradiance system and the ideas about the thermal excitations of such a system, which resemble spin waves. We called these excitations isospin waves. The presence of such excitations can be verified experimentally: they can be identified by the characteristic k-dependence of the fine structure of a superradiance line. Furthermore, we believe that the existence of such excitations leads to a characteristic temperature dependence of the intensity of a superradiance pulse,

$$I(T)=I_\infty\left[1-\left(\frac{T}{T_C}\right)^{3/2}\right].$$

Using these ideas of isospin waves, we have calculated the critical temperature above which superradiance disappears. For sodium this temperature is about 700 K (superradiance disappears in sodium in Rydberg states at higher temperatures), while for pure hydrogen the temperature is roughly 40 K.

This work was sponsored by the Russian Fund for Fundamental Research (Grant No. 99-02-17076).

APPENDIX A: EXCHANGE ENERGY FOR THE DIPOLE INTERACTION OF ATOMS

1. System of hydrogen atoms

For the two-level system we take the states $\psi_2=\psi_{210}(\mathbf{r})$ and $\psi_1=\psi_{100}(\mathbf{r})$. This leads to an expression for the exchange contribution to the energy in the first approximation in the form

$$A=\frac{2d_{21}^z d_{12}^z}{R^3}=\left\langle \psi_{210}(\bar{r}_{\rm I})\psi_{100}(\bar{r}_{\rm II}) \times\left|\frac{3d_{\rm Iz}d_{\rm IIz}-\mathbf{d}_{\rm I}\mathbf{d}_{\rm II}}{R^3}\right|\psi_{100}(\bar{r}_{\rm I})\psi_{210}(\bar{r}_{\rm II})\right\rangle,$$

where

$$d_{12}^z=d_{21}^z=\int \psi_{210}(\bar{r})r\cos\theta\psi_{100}(\bar{r})d^3r=\frac{2^8}{3^5\sqrt{2}}a\,.$$

Then

$$A=\frac{C}{R^3},\quad C=\frac{2^{15}}{3^{10}}a^2, \tag{A1}$$

where a is the Bohr radius.

2. System of sodium atoms

For this system the states are

$$\psi_2=A_{n_2}\left(\frac{r}{2a}\right)^4 e^{-1.68r/2a}\sqrt{\frac{3}{4\pi}}\cos\theta,$$

$$\psi_1=A_{n_1}\left(\frac{r}{a}\right)^3 e^{-1.68r/a}\frac{1}{\sqrt{4\pi}}.$$

The normalization factors are

$$A_{n_2}=\frac{(1.68)^5\sqrt{1.68}}{\sqrt{2a^3}\sqrt{10!}}2^4,\quad A_{n_1}=\frac{(1.68)^4\sqrt{1.68}}{\sqrt{a^3}\sqrt{8!}}.$$

For sodium the dipole moment of the transition is

$$d_{12}^z=d_{21}^z=\frac{2^{10}\sqrt{2^5}}{3^{11}}\frac{\sqrt{30}}{1.68}a. \tag{A2}$$

Then for the exchange contribution to the energy of the dipole interaction of sodium atoms we have

$$A=\frac{C}{R^3},\quad C=\frac{2^{26}}{3^{21}1.68}10a^2. \tag{A3}$$

APPENDIX B

Integrals of the type

$$I_n(k)=\int_\alpha^\rho \exp\left\{2ikR+\frac{b}{R^3}\right\}\frac{dR}{R^n} \tag{B1}$$

can be calculated in the following way.

If $k\ll 1$, $n>0$, and $b>0$, the integral can be calculated by the method of steepest descent.[13] The saddle point in the complex plane is

$$z_0=R_0\exp\left\{-\frac{\pi i}{8}-\frac{in}{8k}\right\},\quad \alpha<R_0=\left(\frac{3b}{kT}\right)^{1/4}.$$

The integral converges rapidly if $kR_0>1$, and is given by

$$I_n(k)=\frac{1}{R_0^{n-1}}\left(\frac{\pi}{4kR_0}\right)^{1/2}\exp\left\{\frac{4}{3}kR_0+\frac{n+3}{8}\pi i\right\}. \tag{B2}$$

If $k=0$ and $n=-2$, the integral can be calculated by the asymptotic expansion method, and

$$Z=I_{-2}(0)=\frac{b^{1/2}}{3x_\alpha^{3/2}}e^{x_\alpha}\left[1+\frac{3}{2x_\alpha}+O\left(\frac{x_\alpha}{x_\rho}\right)\right]. \tag{B3}$$

Here $x_\alpha=b/\alpha^3$ and $x_\rho=b/\rho^3$.

*)E-mail: quark@citadel.stu.neva.ru

[1] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Effects in Optics*, IOPP, Bristol (1993).
[2] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics* [in Russian], Nauka, Moscow (1981).
[3] Yu. L. Klimontovich, *The Kinetic Theory of Electromagnetic Processes*, Springer-Verlag, New York (1983).
[4] R. H. Dicke, Phys. Rev. **93**, 99 (1954).
[5] V. I. Yukalov, Acta Phys. Pol. A **57**, 295 (1980).
[6] E. V. Orlenko and A. A. Rumyantsev, Zh. Éksp. Teor. Fiz. **97**, 439 (1990) [Sov. Phys. JETP **70**, 244 (1990)]; E. V. Orlenko and T. Yu. Latyshevskaya, Zh. Éksp. Teor. Fiz. **113**, 2129 (1998) [JETP **86**, 1167 (1998)].
[7] D. I. Blokhintev, *Quantum Mechanics*, Gordon & Breach, New York (1964); L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[8] A. I. Zaĭtsev, V. A. Malyshev, and E. D. Trifonov, Zh. Éksp. Teor. Fiz. **84**, 475 (1983) [Sov. Phys. JETP **57**, 275 (1983)].
[9] B. Coffey and R. Friedberg, Phys. Rev. A **17**, 1033 (1978); K. Nakamura and S. J. Washimiga, J. Phys. C **13**, 3483 (1980); H. Steudel, Ann. Phys. (Leipzig) **37**, 57 (1980).
[10] C. R. Stroud, J. H. Eberly, W. L. Lama, and L. Mandel, Phys. Rev. A **5**, 1094 (1972).
[11] E. V. Orlenko and A. A. Rumyantsev, Fiz. Nizk. Temp. **15**, 485 (1989) [Sov. J. Low Temp. Phys. **15**, 272 (1989)].
[12] L. Moi, P. Goy, P. Gross, J. M. Raimond, C. Fabre, and S. Haroche, Phys. Rev. A **27**, 2043 (1983); P. Goy, L. Moi, P. Gross, J. M. Raimond, C. Fabre, and S. Haroche, Phys. Rev. A **27**, 2065 (1983).
[13] J. Mathews and R. L. Walker, *Mathematical Methods of Physics*, 2nd ed., Addison-Wesley, Reading, Massachusetts (1970).

Translated by Eugene Yankovsky

Radiative properties of diamagnetic levels in atoms: dependence of transition probability on magnetic field strength

V. D. Ovsyannikov*) and V. V. Chernushin

Voronezh State University, 394693 Voronezh, Russia
(Submitted 4 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1161–1183 (October 1999)

We study the effect of diamagnetic interaction on the probability of radiative transitions of an atom from states split by a field. We write the analytic expressions for the diamagnetic corrections to the matrix elements of transitions belonging to the Lyman and Balmer series and of transitions between arbitrary nondegenerate states in hydrogen. We also discuss the perturbation theory for transitions from degenerate diamagnetic states. The theory is based on expanding in powers of the field strength the eigenfunctions and eigenvalues of the matrix of diamagnetic interaction in the subspace of states with given principal and magnetic quantum numbers. The field changes the coefficients in both the superposition and the degenerate basis. To derive the analytic expressions for the higher-order matrix elements, we use the Sturm expansion of the reduced Coulomb Green's function. We also elaborate on the features of the frequency dependence of the corrections to the radiative matrix elements, which correlate with the structure of the diamagnetic spectrum of excited levels. Finally, we establish that the magnetic field acts selectively on the diamagnetic components of emission lines: as the field strength increases, an increase in the intensity of certain lines is accompanied by a decrease in the intensity of the other lines.
[S1063-7761(99)00410-2]

1. INTRODUCTION

Establishing the changes in the characteristics of an atomic spectrum caused by a magnetic field constitutes a fundamental problem of experimental and theoretical physics dating back to the discovery of the Zeeman effect.[1] The splitting, polarization, and intensity of the Zeeman lines have been thoroughly described by the classical and quantum mechanical theories. However, investigations into the nonlinear effects of a magnetic field on an atom are of more recent vintage. The research in this field deals primarily with the shift and splitting of atomic lines in the lowest order in the diamagnetic interaction (see, e.g., Braun's review[2]). Insufficient attention to the quadratic interaction in the field strength can be explained by the very small value of the corrections introduced by this interaction into the energies of the atomic levels under conditions typical of laboratory experiment, in which the paramagnetic interaction linear in the field strength essentially completely determines the position of the atomic lines in the spectrum, while the diamagnetic interaction produces a relatively small correction to the lowest bound states.

The situation changes dramatically as the magnetic field strength and/or the energy of the excited state increase. Hence interest in the dynamic part of the spectrum of the interaction operator was encouraged by the discovery of astronomical objects with very high magnetic fields (white dwarfs and neutron stars), as well as by the development of a new branch of atomic and molecular physics, the spectroscopy of highly excited (Rydberg) states.[3] By the 1980s, extensive data had been gathered on the splitting, in the first order in the diamagnetic interaction, of the Zeeman states of the atoms of hydrogen and alkali metals (a review of the corresponding literature is given in Ref. 2). Grozdanov and Taylor[4] calculated second-order corrections by employing an effective Hamiltonian that used the additional symmetry of the excited states of the hydrogen atom.[5]

The splitting into diamagnetic sublevels is accompanied by a redistribution of the intensity of the emitted lines across the entire split set. Finding the intensity of the individual components of this set amounted to calculating the distribution of the dipole oscillator strengths, which is proportional to the contribution of vectors with given angular momenta to the initial and final states.[2–4] Since the wave functions of the atom in the field were determined by the eigenvectors of the diamagnetic matrix in the lowest order, neither the structure of these states (the coefficients in the expansion in spherical harmonics) nor the corresponding matrix elements of the radiative transitions depended on the field strength. The dependence of these characteristics on the field strength can be detected only if higher-order corrections are taken into account. Knowing this dependence not only augments the information about the properties of the atom and its interaction with the field, but also makes it possible to use this information in developing new methods to determine the force with which a field acts on the atom, and for magnetic induction monitoring of the radiative intensity.

In addition to an existing line becoming more (or less) intense, new lines can emerge that are not present in the spectrum of the free atom. The diamagnetic part of the interaction operator mixes states of the same parity and different

orbital quantum numbers, which leads to the emergence in a magnetic field **B** of lines that obey the selection rules for orbital angular momentum, $|\Delta l|=2L+1$, with $L=1,2\ldots$. The corresponding dipole matrix elements are proportional to B^{2L}. Thus, in addition to emission lines corresponding, for example, to dipole $p-s$ transitions, magnetically induced lines emerge in the spectrum. These new lines correspond to $f-s$, $h-s$, etc. transitions, whose matrix elements are proportional to B^2, B^4, etc.

A specific situation emerges for the transitions between hydrogen-like states. The degeneracy that remains if we allow only for the paramagnetic splitting is lifted completely if we take into account the diamagnetic interaction. Here the wave function of each diamagnetic state is a superposition of wave functions with given orbital angular momentum of well-defined parity. The coefficients of such a superposition are, in the zeroth approximation, independent of the field. Nevertheless, the contribution of each term to the matrix element of the dipole transition between the hydrogen-like levels in the field emerges in a specific diagram of the perturbation theory for degenerate states,[6] so that, for example, the $f-s$ transitions produce a contribution proportional to B^2, as they do for nondegenerate states of many-electron atoms.

In the present paper we calculate the corrections to the matrix elements of the dipole radiative transitions between hydrogen-like states, which are induced by the interaction of the atom and a constant magnetic field. In Sec. 2 we derive general formulas for the corrections of the perturbation theory for nondegenerate atomic levels, which can be used, for example, to calculate the transitions between states corresponding to ''circular'' orbits of an electron in a Coulomb field. The first-order corrections are written in terms of the radial matrix elements of the operators of the dipole moment and the diamagnetic interaction. The matrix elements are calculated analytically via a Sturm expansion for the Coulomb Green's function.

Magnetically induced corrections to the matrix elements of the radiative transitions from the degenerate states of a hydrogen-like atom are discussed in Sec. 3. There we allow for the effect of the magnetic field on the basis functions and the coefficients in the superposition of these functions. In Sec. 4 we give the results of numerical calculations of the corrections to the intensity of the Zeeman components belonging to the Lyman and Balmer series. There we also analyze the magnetically induced corrections to the distant components of both series, which correspond to radiative decay of highly excited (Rydberg) states.

2. DIAMAGNETIC CORRECTIONS TO THE DIPOLE MATRIX ELEMENTS OF TRANSITIONS BETWEEN NONDEGENERATE STATES

The probability W_{if} (intensity I_{if}) of emission of electric dipole radiation corresponding to the transition of an atom between states $|i\rangle$ and $|f\rangle$ with energies E_i and E_f is proportional to the third (fourth) power of the transition frequency $\omega_{if}=E_i-E_f$ and the square of the matrix element of the dipole moment operator $d_{if}=\langle i|d_z|f\rangle$ (see, e.g., Ref. 7):

$$I_{if}\propto\omega_{if}W_{if}\propto\omega_{if}^4|d_{if}|^2. \tag{1}$$

The introduction of a magnetic field changes both the frequency and the transition matrix element, so that the dependence of the line intensity on the magnetic field strength B can be found if one knows the dependence on B of the two factors in (1). For a field $B\ll B_0/n^3$ ($B_0=2.35\times10^5$ T, and n is the principal quantum number of an atomic level), the magnetically induced corrections to frequencies and matrix elements can be found by perturbation-theory techniques. Here for the central (unshifted) Zeeman line with polarization parallel to the magnetic field (π-polarization), both the frequency shift and the change in the matrix element are quadratic in the field strength. For the deviated lines with polarization perpendicular to the magnetic field (σ-polarization), the main contribution to the change in the intensity in weak fields is provided by linear corrections to the line frequency. Nevertheless, knowing the quadratic corrections for such lines can be important, especially for a state with a large value of the principal quantum number participating in the transition. Since corrections to the frequencies of atomic lines have been thoroughly studied (see, e.g., Refs. 2, 4, and 6), we limit ourselves to corrections to matrix elements, since so far there have been no studies of such corrections.

The operator representing the interaction of an atom and a constant magnetic field **B** contains two terms, $V_B=V_m+V_D$, where[1)]

$$V_m=-\mathbf{m}\cdot\mathbf{B} \tag{2}$$

is the magnetic dipole interaction, with $\mathbf{m}=-(\mathbf{L}+2\mathbf{S})/2$ the operator of the magnetic dipole moment (the fine structure constant $\alpha\approx1/137$ is included in the definition of the unit magnetic field B_0), and

$$V_D=\frac{B^2}{12}r^2\left[1-\mathbf{C}_{20}\left(\frac{\mathbf{r}}{r}\right)\right] \tag{3}$$

is the diamagnetic interaction, with r and $\mathbf{r}/r$ the radial and angular variables of the atom. In the one-electron approximation, the radius vector **r** determines the position of an outer electron with respect to the atomic nucleus ($\mathbf{C}_{20}(\mathbf{r}/r)$ is a modified spherical harmonic), and $\mathbf{L}=\mathbf{l}$ and $\mathbf{S}=\mathbf{s}$ are the orbital and spin angular momenta of the valence electron.

If we neglect the effect of spin on the energy spectrum E_n of the unperturbed atom and allow for the fact that radiative transitions are spin-independent, we can incorporate the operator (2) into the main Hamiltonian and leave only the diamagnetic interaction (3) as the perturbation. Here the wave functions of the unperturbed basis do not change, while the energies split according to the magnetic quantum number, $E_{nm}=E_n+mB$. Since the action of the diamagnetic interaction operator on the wave function does not change the value of the magnetic quantum number, we can set up a perturbation theory for this operator in the subspace of states with a given m (below we assume m to be positive).

Suppose that in zero magnetic field the nondegenerate initial and final states in (1) coincide with the states $|1\rangle$ and $|0\rangle$ (with energies E_1 and E_0). Then the equations for the wave function and energy of the atom in a finite field can be

written in integral form, convenient for iterations, via the Green's function $G_E(\mathbf{r},\mathbf{r}')$ of the unperturbed atom:[8]

$$E_{i(f)}=E_{1(0)}+\langle 1(0)|V_D|i(f)\rangle,$$

$$|i(f)\rangle=|1(0)\rangle-G'_{E_{i(f)}}V_D|i(f)\rangle, \tag{4}$$

where

$$G'_{E_{i(f)}}(\mathbf{r},\mathbf{r}')=G_{E_{i(f)}}(\mathbf{r},\mathbf{r}')-\frac{\langle\mathbf{r}|1(0)\rangle\langle 1(0)|\mathbf{r}'\rangle}{E_{1(0)}-E_{i(f)}} \tag{5}$$

is the reduced Green's function.

Using the standard iterative procedure of perturbation theory to solve Eqs. (4), we obtain series expansions in powers of the small parameter $B^2\ll 1$ for the wave functions and energies. These expansions are then used to obtain power series for the frequency and the transition matrix element:

$$\omega_{if}(B)=\omega_{10}\left(1+\sum_{s=1}^{\infty}w_{10}^{(s)}B^{2s}\right),$$

$$d_{if}(B)=d_{10}\left(1+\sum_{s=1}^{\infty}q_{10}^{(s)}B^{2s}\right). \tag{6}$$

The coefficient $w_{10}^{(s)}$ in the expansion for the frequency is the ratio of the difference of the diamagnetic susceptibilities, $\chi_0^{(s)}-\chi_1^{(s)}$, which determine the s-order corrections to the energy of the lower and upper level, $\Delta E_{f(i)}^{(s)}=-\chi_{0(1)}^{(s)}B^{2s}/2s!$, to the unperturbed frequency ω_{10} (see Ref. 9):

$$w_{10}^{(s)}=\frac{\chi_0^{(s)}-\chi_1^{(s)}}{2s!\,\omega_{10}}. \tag{7}$$

We can easily obtain analytic expressions for the first-order susceptibilities $\chi_{nlm}^{(1)}$ of unperturbed hydrogen-like states (states with $m\leqslant l<n\leqslant m+4$, the lowest in the Zeeman set of states with a given magnetic quantum number, which are called states with circular or nearly circular orbits) from the diagonal part of the matrix element of the operator V_D (see, e.g., Refs. 2–4):

$$V_{ll'}=\langle nlm|V_D|nl'm\rangle=-\frac{B^2}{2}v_{ll'}^{(1)},$$

$$\begin{aligned}v_{ll'}^{(1)}=&-\frac{(l^2+l-1+m^2)n^2}{4Z^2(2l-1)(2l+3)}[5n^2+1-3l(l+1)]\delta_{l'l}\\&+\left\{\frac{[(l+1)^2-m^2][(l+2)^2-m^2]}{(2l+1)(2l+3)^2(2l+5)}[n^2-(l+1)^2]\right.\\&\left.\times[n^2-(l+2)^2]\right\}^{1/2}\frac{5n^2}{8Z^2}\delta_{l'l+2}\\&+\left\{\frac{[(l-1)^2-m^2][l^2-m^2]}{(2l-3)(2l-1)^2(2l+1)}[n^2-(l-1)^2]\right.\\&\left.\times[n^2-l^2]\right\}^{1/2}\frac{5n^2}{8Z^2}\delta_{l'l-2},\end{aligned} \tag{8}$$

where $\delta_{ll'}$ is the Kronecker delta.

An expansion of the tensor $\chi^{(2)}$ into irreducible parts for the unperturbed levels of many-electron atoms was done in Ref. 9, where the results of numerical calculations of the irreducible components of this tensor for the s, p, and d states of alkali elements were also given. A method for calculating $\chi^{(2)}$ of highly excited (Rydberg) states was proposed in Ref. 6. Numerical calculations of susceptibilities $\chi^{(s)}$ of very high orders (up to $s=75$) of hydrogen levels with the principal quantum numbers $n\leqslant 3$ were done by Vaĭnberg *et al.*[10] Analytic expressions for third-order susceptibilities in nondegenerate states of hydrogen with arbitrary values of n were derived in Ref. 11.

In addition to perturbation-theory techniques, other methods have been developed to calculate the energies of atomic levels, making it possible to determine the changes in the frequencies of radiative transitions in magnetic fields stronger than those that perturbation theory can handle (see, e.g., Refs. 12–15). Thus, we now possess a broad array of analytic and numerical methods to calculate the changes in the frequencies of radiative transitions of atoms in magnetic fields, and consequently to determine the corrections (7). There have as yet been no studies of the changes induced by a magnetic field in the matrix element $d_{if}(B)$.

The solution of this problem can be obtained on the basis of the integral equations (4). Expanding the wave functions of the initial and final states in powers of the operator (3), we can write the coefficients $q_{10}^{(s)}$ in (6) as a ratio of the s-ordermatrix element to the unperturbed matrix element. In particular, at $s=1$ (below we limit ourselves to the first-order correction $q_{10}^{(1)}\equiv q_{10}$ and drop the superscript (s) indicating the order) we have

$$q_{10}=q_{10}(1)+q_{10}(0), \tag{9}$$

where

$$q_{10}(1)=-\frac{\langle 1|V_DG'_{E_1}d_z|0\rangle}{B^2d_{10}} \tag{10}$$

determines the contribution related to changes in the wave function of the upper level, $|1\rangle$, induced by the field, and the term

$$q_{10}(0)=-\frac{\langle 1|d_zG'_{E_0}V_D|0\rangle}{B^2d_{10}} \tag{11}$$

is related to the variation of the wave function of the lower level, $|0\rangle$.

After integration over the angular variables of the valence electron is performed, the quantities (10) and (11) can be written as a ratio of the radial matrix elements of the first and orders of the operators of the electric dipole and diamagnetic interactions (r and r^2). In particular, for the π-transition from the $|nlm\rangle$ state to the $|n'l'm\rangle$ state with $m=l'=l-1$ we have

$$q^{\pi}_{nl,n'l'}(nl)=-\frac{l}{4(2l+3)}\frac{\langle nl|r^2 g_l^{(n)}(r,r')r'|n'l'\rangle}{\langle nl|r|n'l'\rangle},$$

$$q^{\pi}_{nl,n'l'}(n'l')=-\frac{l}{4(2l+1)}\frac{(2l+3)\langle nl|rg_{l-1}^{(n')}(r,r')r'^2|n'l'\rangle-2\langle nl|rg_{l+1}^{(n')}(r,r')r'^2|n'l'\rangle}{(2l+3)\langle nl|r|n'l'\rangle}. \tag{12}$$

Similar expressions for the σ-transition from the $|nlm\rangle$ state to the $|n'l'm'\rangle$ state with $m'=l'=m-1=l-1$ have the form

$$q^{\sigma}_{nl,n'l'}(nl)=-\frac{l+1}{4(2l+3)}\frac{\langle nl|r^2 g_l^{(n)}(r,r')r'|n'l'\rangle}{\langle nl|r|n'l'\rangle},$$

$$q^{\sigma}_{nl,n'l'}(n'l')=-\frac{l}{4(2l+1)}\frac{\langle nl|rg_{l-1}^{(n')}(r,r')r'^2|n'l'\rangle}{\langle nl|r|n'l'\rangle}-\frac{\langle nl|rg_{l+1}^{(n')}(r,r')r'^2|n'l'\rangle}{4(2l+1)(2l+3)\langle nl|r|n'l'\rangle}. \tag{13}$$

In these expressions, $g_{l_1}^{(n)}(r,r')$ is the reduced radial Green's function in the subspace of states of the valence electron with the orbital angular momentum l_1 (Ref. 8).

For transitions between hydrogen-like states in atoms, all the radial matrix elements in the expressions (12) and (13) can be calculated analytically and represented as functions of principal and orbital quantum numbers. For the first-order matrix elements, these expressions are given by the Gordon formulas (see, e.g., Ref. 16). Using the Sturm expansion of the reduced Green's function of the Coulomb potential of a charge Z,[6,9,11]

$$g_l^{(n)}(r,r')=\frac{4Z}{n}\left\{\sum_{k\neq n_r}^{\infty}\frac{k!}{(k+2l+1)!}\frac{f_{kl}(2Zr/n)f_{kl}(2Zr'/n)}{k+l+1-n}+\frac{n_r!}{(n+l)!\,n}\left[\frac{5}{2}f_{n_r l}\left(\frac{2Zr}{n}\right)f_{n_r l}\left(\frac{2Zr'}{n}\right)+r\frac{df_{n_r l}(2Zr/n)}{dr}f_{n_r l}\left(\frac{2Zr'}{n}\right)+f_{n_r l}\left(\frac{2Zr}{n}\right)r'\frac{df_{n_r l}(2Zr'/n)}{dr'}\right]\right\}, \tag{14}$$

we obtain expressions for the second-order matrix elements that are similar to the Gordon formulas. The reason is the orthogonality of the Sturm function

$$f_{kl}(x)=e^{-x/2}x^l L_k^{2l+1}(x) \tag{15}$$

and the radial wave function

$$R_{nl}(r)=\frac{2Z^{3/2}}{n^2}\sqrt{\frac{n_r!}{(n+l)!}}f_{n_r l}\left(\frac{2Zr}{n}\right), \tag{16}$$

which follows from the orthogonality of the Laguerre polynomials $L_k^{\alpha}(x)$ (see Ref. 17). This property truncates the infinite series in (14), leaving only a finite number of terms. In particular,

$$\begin{aligned}\left\langle f_{kl}\left(\frac{2Zr}{n}\right)|r^2|f_{k'l}\left(\frac{2Zr}{n}\right)\right\rangle&=\left(\frac{n}{2Z}\right)^5\int_0^{\infty}e^{-x}x^{2l+4}L_k^{2l+1}(x)L_{k'}^{2l+1}(x)dx\\&=\left(\frac{n}{2Z}\right)^5\frac{(k+2l+1)!}{k!}\{-(k-2)_3\delta_{k'k-3}+6(k+l)\\&\times(k-1)_2\delta_{k'k-2}-3k[5k(k+2l+1)+4l(l+1)+2]\\&\times\delta_{k'k-1}+[(k+2l+2)_3+9k(k+2l+2)_2\\&+9(k-1)_2(k+2l+2)+(k-2)_3]\delta_{k'k}\\&-3(k+2l+2)[5k(k+2l+3)+4l^2+14l+12]\\&\times\delta_{k'k+1}+6(k+l+2)(k+2l+2)_2\delta_{k'k+2}\\&-(k+2l+2)_3\delta_{k'k+3}\},\end{aligned} \tag{17}$$

where we have used Pochhammer's symbol $(a)_s=a(a+1)\times\cdots(a+s-1)$. Note that the orthogonality condition holds only if the energy of the initial (or final) state coincides with the energy of the Green's function.

Thus, after integrating the term with two Sturm functions (15) with respect to the radial variable of the operator V_D, we are left with six terms in (14), each of which is similar to the radial matrix element of the first-order dipole radiative transition and has an index k that differs from the radial quantum number $n_r=n-l-1$ by at most 3. In contrast to (17), the arguments of the Sturm function in such a matrix element are different, since the energies of the initial (final) and intermediate (the Green's function) states do not coincide. It is convenient to express such a matrix element in terms of the hypergeometric function of two variables, the Appell function F_2 (Ref. 17), which is represented by a double sum with the number of terms depending on the radial quantum numbers of the upper and lower levels:

$$\begin{aligned}\left\langle f_{kl}\left(\frac{2Zr}{n}\right)\middle|r\middle|f_{k'l'}\left(\frac{2Zr}{n'}\right)\right\rangle&=\left(\frac{nn'}{4Z^2}\right)^2\frac{(k+b)!(k'+b')!}{k!\,k'!}\frac{a!}{b!\,b'!}x^{l+2}(x')^{l'+2}F_2\\&\times(a+1;-k,-k';b+1,b'+1;x,x'),\end{aligned} \tag{18}$$

where we have introduced the notation

$$a=l+l'+3,\quad b=2l+1,\quad b'=2l'+1,$$

$$x=\frac{2n'}{n+n'},\quad x'=\frac{2n}{n+n'}.$$

The second-order matrix elements are represented in the form of a superposition of expressions of type (18), in which one of the integer parameters k or k' coincides with the radial quantum number of the upper, $n_r=n-l-1$, or lower, $n_r'=n'-l'-1$, level, while the other coincides with the summation index of the series in (14). In practice, the radial quantum number of the lower level is usually not very large, so that it is convenient to write the series F_2 as a sum of $n_r'+1$ Gauss hypergeometric series:

$$F_2(a+1;-k,-n_r';b+1,b'+1;x,x') = \sum_{s=0}^{n_r'} \frac{(a+1)_s(-n_r')_s}{s!\,(b'+1)_s}(x')^s\,{}_2F_1(a+1+s,-k;b+1;x).$$

Since $N_s=a+s-b=s+l'-l+2$ is a positive integer ($l'=l\pm1$), ${}_2F_1$ can be written as the sum of N_s+1 power-law expressions in terms of the argument x (see Ref. 17):

$${}_2F_1(a+1+s,-k;b+1;x)=(1-x)^{k-N_s}\sum_{s'=0}^{N_s}\times\frac{(b+k+1)_{s'}(-N_s)_{s'}}{s'!(b'+1)_{s'}}(x)^{s'}.$$

As a result, the first-order matrix element $\langle nl|r|n'l'\rangle$ can be written as a Gordon formula, and for the correction factors we can obtain fractional rational expressions that depend on the principal and orbital quantum numbers of the initial and final states. For lines of the Lyman series ($n'=1$ and $l'=0$) these expressions have the form

$$q^{\pi}_{np,1s}(np)=\frac{1}{2}q^{\sigma}_{np,1s}(np)=\frac{n^4(15n^8-80n^6+152n^4-84n^2+189)}{60Z^4(n^2-1)^3},$$

$$q^{\pi}_{np,1s}(1s)=\frac{79n^6-677n^4+269n^2-55}{120Z^4(n^2-1)^3},$$

$$q^{\sigma}_{np,1s}(1s)=\frac{113n^6-1089n^4+263n^2-55}{120Z^4(n^2-1)^3},\qquad(19)$$

$$q^{\pi}_{np,1s}=q^{\pi}_{np,1s}(np)+q^{\pi}_{np,1s}(1s)=\frac{30n^8-100n^6+74n^4+159n^2-55}{120Z^4(n^2-1)},$$

$$q^{\sigma}_{np,1s}=q^{\sigma}_{np,1s}(np)+q^{\sigma}_{np,1s}(1s)=\frac{60n^{10}-260n^8+348n^6+125n^4-208n^2+55}{120Z^4(n^2-1)^2}.$$

The analogous expressions for the transitions $|np\rangle\to|2s\rangle$ of the Balmer series ($n'=2$ and $l'=0$) are

$$q^{\pi}_{np,2s}(np)=\frac{1}{2}q^{\sigma}_{np,2s}(np)=\frac{n^4(15n^8-280n^6+1684n^4-3168n^2+12\,096)}{60Z^4(n^2-4)^3},$$

$$q^{\pi}_{np,2s}(2s)=\frac{2(451n^6-6316n^4+15\,184n^2-13\,120)}{15Z^4(n^2-4)^3},$$

$$q^{\sigma}_{np,2s}(2s)=\frac{2(507n^6-8172n^4+15\,568n^2-13\,120)}{15Z^4(n^2-4)^3},\qquad(20)$$

$$q^{\pi}_{np,2s}=q^{\pi}_{np,2s}(np)+q^{\pi}_{np,2s}(2s)=\frac{15n^8-160n^6+164n^4+4312n^2-6560}{60Z^4(n^2-4)},$$

$$q^{\sigma}_{np,2s}=q^{\sigma}_{np,2s}(np)+q^{\sigma}_{np,2s}(2s)=\frac{15n^{10}-220n^8+804n^6+2076n^4-12\,288n^2+13\,120}{30Z^4(n^2-4)^2}.$$

For the transitions $|ns\rangle\to|2p\rangle$ we have (below we give only the overall coefficient, which takes into account the contributions of the upper and lower levels, $q^{\pi,\sigma}=q^{\pi,\sigma}(n)+q^{\pi,\sigma}(n')$)

$$q^{\pi}_{ns,2p}=-\frac{100n^8-675n^6+1028n^4-5088n^2+7680}{120Z^4(n^2-4)},$$

$$q^{\sigma}_{ns,2p}=\frac{125n^{10}-1630n^8+5224n^6+9280n^4-56\,064n^2+61\,440}{120Z^4(n^2-4)^2}.\qquad(21)$$

For the transitions $|ndm\rangle\to|2pm'\rangle$ the overall coefficients are

$$q^{\pi}_{nd0,2p0}=\frac{1565n^8-20\,040n^6+50\,512n^4+196\,608n^2-344\,064}{5376Z^4(n^2-4)},$$

$$q^{\sigma}_{nd1,2p0}=\frac{15n^{10}-268n^8+1392n^6-128n^4-8448n^2+10\,752}{42Z^4(n^2-4)^2},\qquad(22)$$

$$q^{\pi}_{nd1,2p1}=\frac{15n^8-208n^6+560n^4+3008n^2-5376}{42Z^4(n^2-4)},$$

$$q^{\sigma}_{nd2,2p1}=\frac{45n^{10}-804n^8+4176n^6+512n^4-34\,304n^2+43\,008}{84Z^4(n^2-4)^2},$$

$$q^{\sigma}_{nd0,2p1}=\frac{835n^{10}-16\,580n^8+92\,048n^6+70\,976n^4-1\,081\,344n^2+1\,376\,256}{2688Z^4(n^2-4)^2}.$$

TABLE I. The factors $q^{\pi,\sigma}$ for magnetically induced corrections to the matrix elements of the radiative π- and σ-transitions from the np to the $1s$ and $2s$ states and from the ns and nd_m to the $2p_{m'}$ states in the hydrogen atom; $(k)=10^k$.

q	$n=2$	3	4	5	6	7	8
$q^{\pi}_{np,1s}(n)$	9.926	136.7	876.2	3.543(3)	1.091(4)	2.801(4)	6.315(4)
$q^{\pi}_{np,1s}(1s)$	−1.468	0.0833	0.3815	0.4931	0.5477	0.5788	0.5983
$q^{\sigma}_{np,1s}(1s)$	−2.838	−0.0573	0.4647	0.6580	0.7522	0.8057	0.8391
$q^{\pi}_{np,2s}(n)$		154.3	564.6	2.778(3)	9.368(3)	2.522(4)	5.846(4)
$q^{\pi}_{np,2s}(2s)$		−63.23	35.51	49.90	54.48	56.52	57.60
$q^{\sigma}_{np,2s}(2s)$		−176.4	17.02	45.93	55.38	59.66	61.99
$q^{\pi}_{ns,2p0}(n)$		−137.7	−2.706(3)	−1.149(4)	−3.582(4)	−9.245(4)	−2.090(5)
$q^{\pi}_{ns,2p0}(2s)$		−210.9	−56.89	−28.37	−17.75	−12.50	−9.482
$q^{\sigma}_{ns,2p1}(2s)$		−421.8	−113.8	−56.75	−35.50	−25.01	−18.96
$q^{\pi}_{nd0,2p0}(n)$		55.39	538.2	2.934(3)	1.023(4)	2.806(4)	6.576(4)
$q^{\pi}_{nd0,2p0}(2p)$		−11.65	23.11	28.00	29.50	30.15	30.49
$q^{\sigma}_{nd1,2p0}(n)$		117.6	568.9	3.360(3)	1.206(4)	3.351(4)	7.913(4)
$q^{\sigma}_{nd1,2p0}(2p)$		−54.27	44.44	55.82	58.75	59.82	60.30
$q^{\pi}_{nd1,2p1}(n)$		117.6	568.9	3.360(3)	1.206(4)	3.351(4)	7.913(4)
$q^{\pi}_{nd1,2p1}(2p)$		−51.71	60.44	74.54	78.50	80.08	80.85
$q^{\sigma}_{nd2,2p1}(n)$		176.4	853.3	5.040(3)	1.809(4)	5.027(4)	1.187(5)
$q^{\sigma}_{nd2,2p1}(2p)$		−80.13	74.67	93.09	98.00	99.86	110.72
$q^{\sigma}_{nd0,2p1}(n)$		183.3	345.8	2.533(3)	9.688(3)	2.766(4)	6.630(4)
$q^{\sigma}_{nd0,2p1}(2p)$		−108.5	88.89	111.6	117.5	119.6	120.6

With allowance for the symmetry relation for the coefficients $q_{nlm,n'l'm'}=q_{nl-m,n'l'-m'}$, which follows from the invariance of the operator V_D under inversion of the vector **B**, the expressions (19)–(22) fully determine the magnetically induced corrections to the matrix elements of the transitions belonging to the Lyman and Balmer series. The overall coefficients have an asymptotic dependence, common to all transitions, on the principal quantum number n of the upper level: $q_{nl,n'l'}\sim n^6$. Here the main contribution is provided by corrections due to the action of the magnetic field on the upper state. The corrections to the wave function of the lower level, which determine the term $q_{nl,n'l'}(n'l')$, are essentially independent of n for $n\gg 1$, and are negligible compared to $q_{nl,n'l'}(nl)$.

One interesting feature of the quantities q is that the same power of the factor $n^2-n'^2$ appears in the denominators of all the expressions (19)–(22): the third power for the individual components $q^{\pi,\sigma}(nl)$ and $q^{\pi,\sigma}(n'l')$, the second power for the total quantity $q^{\sigma}=q^{\sigma}(nl)+q^{\sigma}(n'l')$, and the first for $q^{\pi}=q^{\pi}(nl)+q^{\pi}(n'l')$. This pattern can serve as a check on the validity of such expressions for other transitions as well.

Table I lists the numerical values of the components $q^{\pi,\sigma}_{nlm,n'l'm'}(nl)$ and $q^{\pi,\sigma}_{nlm,n'l'm'}(n'l')$, which correspond to the contributions of the magnetically induced corrections to the upper and lower levels involved in the dipole transitions of the hydrogen atom from the np to the $1s$ and $2s$ states and from the ns and nd_m to the $2p_{m'}$ states ($m=0,1,2$ and $m'=0,1$). By virtue of the simple relationship between the corrections for the π- and σ-transitions between the p and s states, $q^{\sigma}_{np,n's}(np)=2q^{\pi}_{np,n's}(np)$, which follows from the general expressions (12) and (13), we list only the values of $q^{\pi}_{np,n's}(np)$. The data in Table I graphically show the relationship between the contributions to the overall correction to the matrix element of the corrections to the wave functions of the upper and lower levels, with the overall correction being obtained by adding the numbers from two neighboring rows. Clearly, when $n\gg n'$, the contribution of the correction to the lower level can be neglectd. The data can be used directly in the matrix elements of the transitions between nondegenerate states ($|2p\rangle,|3p\rangle\rightarrow|1s\rangle$, $|3p\rangle\rightarrow|2s\rangle$, and $|3d_{1,2}\rangle,|4d_{1,2}\rangle\rightarrow|2p_{0,1}\rangle$). For other states the data determine the correction only partially, since a magnetic field mixes highly excited states of the hydrogen atom in such a way that the angular momentum ceases to be a constant of the motion even in the zeroth approximation (see Sec. 3). Note that the magnetically induced corrections to the matrix element are positive for all the transitions listed in Table I, with the exception of $q^{\pi}_{ns,2p}$. Thus, for transitions from the unperturbed states in the Lyman and Balmer series, the presence of a magnetic field increases the absolute value of the radiative matrix element.

The general formulas for the coefficient q determining the corrections to the matrix elements of transitions between nondegenerate $|nlm\rangle$ states with arbitrary principal quantum numbers can also be easily derived from the expressions for the corresponding radial integrals. All these transitions reduce the orbital quantum number by 1, $|nlm\rangle\rightarrow|n'l'-1m'\rangle$, with the magnetic quantum number remaining unchanged, $m'=m$, in the case of π-transitions, and they reduce the magnetic quantum number by 1, $m'=m-1$, in the case of σ-transitions. As in (19)–(22), the salient characteristic of these transitions is the principal quantum number of the upper level; the orbital (l) and magnetic (m) quantum numbers differ by at most 3: $l=n-1,n-2$ and $m=l,l-1$. Since $l-l'=1$ and $m-m'=0,1$, the variation of the principal quantum number in such transitions is limited by the condition $n-n'=1,2$. Thus, there are generally three types of π-transitions and six types of σ-transitions between non-

TABLE II. Analytic expressions for the factors $q^{\pi,\sigma}_{\nu_l\nu_m\nu_{n'}}$ and $w^{\pi,\sigma}_{\nu_l\nu_m\nu_{n'}}$ for the magnetically induced corrections to the matrix elements and frequencies of the radiative π- and σ-transitions between nondegenerate states of the hydrogen atom, $nlm\to n'l-1m'$, where $l=n-\nu_l$, $m=n-\nu_m$, and $n'=n-\nu_{n'}$; the upper part of the table corresponds to π-transitions ($m'=m$) and the lower (larger) part, to σ-transitions ($m'=m-1$).

$$q^{\pi}_{121}=-\frac{n(n-1)}{48(2n-1)}(18n^5-112n^4+76n^3+4n^2-9n+1),$$

$$w^{\pi}_{121}=\frac{(n-1)^3n^3(3n-1)}{4(2n-1)},$$

$$q^{\pi}_{231}=-\frac{n-2}{48(2n-1)^2}(36n^7-502n^6+1272n^5-2350n^4+2698n^3-1677n^2+521n-64),$$

$$w^{\pi}_{231}=\frac{(n-1)^2n^2(n-2)}{4(2n-1)}(3n^2+7n-4),$$

$$q^{\pi}_{232}=-\frac{n-2}{24(n-1)}(25n^6-210n^5+460n^4-524n^3+400n^2-168n+32),$$

$$w^{\pi}_{232}=\frac{n^2(n-2)^3}{8(n-1)}(5n^2-4n+2).$$

$$q^{\sigma}_{111}=-\frac{n}{48(2n-1)^2}(48n^7-318n^6+445n^5-271n^4+80n^3+n^2-8n+1),$$

$$w^{\sigma}_{111}=\frac{n^3(n-1)^2}{4(2n-1)}(4n^2-3n+1),$$

$$q^{\sigma}_{121}=-\frac{n(n-1)}{48(2n-1)^2}(48n^6-306n^5+381n^4-154n^3-6n^2+17n-2),$$

$$w^{\sigma}_{121}=\frac{n^3(n-1)^3}{2},$$

$$q^{\sigma}_{221}=-\frac{n-1}{48(2n-1)^2}(48n^7-486n^6+1385n^5-3170n^4+4204n^3-2879n^2+978n-128),$$

$$w^{\sigma}_{221}=\frac{n^2(n-1)^3}{2(2n-1)}(2n^2+5n-4),$$

$$q^{\sigma}_{231}=-\frac{48n^8-570n^7+2373n^6-5827n^5+9724n^4-9799n^3+5552n^2-1627n+192}{48(2n-1)^2},$$

$$w^{\sigma}_{231}=\frac{n^2(n-1)^2}{4(2n-1)}(4n^3+3n^2-25n+12),$$

$$q^{\sigma}_{222}=-\frac{57n^8-604n^7+2233n^6-4469n^5+5708n^4-4864n^3+2672n^2-864n+128}{48(n-1)^2},$$

$$w^{\sigma}_{222}=\frac{n^2(n-2)^2}{16}(11n^2-12n+8),$$

$$q^{\sigma}_{232}=-\frac{n-2}{48(n-1)^2}(57n^7-540n^6+1623n^5-2563n^4+2550n^3-1612n^2+584n-96),$$

$$w^{\sigma}_{232}=\frac{n^2(n-2)^3}{16(n-1)}(11n^2-11n+6).$$

degenerate states, and each of these types can be characterized by a set of three numbers,

$$\nu_l=n-l,\quad \nu_m=n-m,\quad \nu_{n'}=n-n'.$$

Table II lists the analytic expressions for the factors $q^{\pi,\sigma}\equiv q^{(1)}$, which determine the B^2-corrections in (6) to the matrix elements of the transitions between nondegenerate states, and which are characterized by certain sets of the numbers ν_l, ν_m, and $\nu_{n'}$ (lower indices). For all these expressions, the contributions of the corrections to the wave functions of the upper and lower levels are of the same order, but have opposite signs (this property of having opposite signs is clearly seen in the α lines of the Lyman and Balmer series in Table I).

Since to have all the information about the changes in the transition probability we must know the corrections to both factors in (1), in Table II we also list the analytic expressions for the corresponding factors $w^{\pi,\sigma}\equiv w^{(1)}$, which determine the first-order corrections to the dipole-transition frequencies in (6).

Table III lists the numerical values of the factors $q^{\pi,\sigma}_{nlm,n'l'm'}$ calculated by the formulas of Table II for various values of the principal quantum number n of the initial state. When $n\leqslant 5$, the positive contribution of the upper level is dominant; when $n\geqslant 6$, the factor q^{π} becomes negative (for the π-transition with $l=n-2$, $m=n-3$, and $n'=n-1$ this happens when $n\geqslant 12$), since in this range of values of n the negative contribution of the correction to the wave function

TABLE III. The factors $q^{\pi,\sigma}$ for magnetically induced corrections to the matrix elements of the radiative π- and σ-transitions between nondegenerate states of the hydrogen atom, $nlm \to n'l-1m'$, where $l=n-\nu_l$, $m=n-\nu_m$, and $n'=n-\nu_{n'}$; the first three rows of the table correspond to π-transitions ($m'=m$) and the remaining rows, to σ-transitions ($m'=m-1$); $(k)=10^k$.

ν_l	ν_m	$\nu_{n'}$	$n=4$	5	6	7	8	9	10
1	2	1	1.910(2)	1.942(2)	−6.434(2)	−4.026(3)	−1.324(4)	−3.386(4)	−7.464(4)
2	3	1	5.252(2)	1.718(3)	4.113(3)	7.908(3)	1.265(4)	1.669(4)	1.651(4)
2	3	2	6.293(2)	1.076(3)	−9.947(2)	−1.232(4)	−4.687(4)	−1.295(5)	−3.005(5)
1	1	1	2.121(2)	−1.606(1)	−1.784(3)	−7.743(3)	−2.283(4)	−5.524(4)	−1.177(5)
1	2	1	1.484(2)	−6.460(1)	−1.655(3)	−7.072(3)	−2.093(4)	−5.101(4)	−1.094(5)
2	2	1	4.732(2)	1.069(3)	1.521(3)	4.986(2)	−4.949(3)	−2.026(4)	−5.439(4)
2	3	1	2.106(2)	4.961(2)	4.928(2)	−1.083(3)	−7.057(3)	−2.264(4)	−5.646(4)
2	2	2	9.280(2)	1.269(3)	−2.106(3)	−1.790(4)	−6.361(4)	−1.698(5)	−3.854(5)
2	3	2	6.133(2)	9.106(2)	−1.857(3)	−1.544(4)	−5.579(4)	−1.513(5)	−3.479(5)

of the lower level becomes dominant. The asymptotic ($n \gg 1$) dependence $q^{\pi} \propto -n^6$ has a sign opposite to that represented by Eqs. (19)–(22) for the higher-order components of the Lyman and Balmer series. This result shows, in particular, that the presence of a magnetic field leads to a decrease in the matrix element of a dipole transition between Rydberg states with circular orbits.

Nevertheless, the second factor in the expression (1) for the probability, the transition frequency $\omega_{nn'}$, always has a positive quadratic correction $\propto n^6$ for hydrogen-like lines when $n \gg 1$, as the formulas of Table II show. The relationship between the coefficient in this asymptotic behavior and the corresponding coefficient in the asymptotic behavior for $q^{\pi,\sigma}$ determines the sign of the correction to the probability. Such a correction is positive for all transitions with $\nu_{n'}=1$ and negative for transitions with $\nu_{n'}=2$. Thus, the transition probability between the nondegenerate states of adjacent levels of the hydrogen atom increases as the magnetic field strength increases, while for the transitions in which one shell is skipped this probability decreases.

Note that such a relationship between the corrections to the frequencies and matrix elements is specific only to the given group of transitions, since here an increase in the difference of the diamagnetic susceptibilities in the numerator of (7) is accompanied by a decrease in the transition frequency in the denominator. As a result, the asymptotic dependence on n of the factor $w^{(1)}$ for the correction to the frequency proves to be similar to the dependence of the correction to the matrix element. But if the principal quantum number of only the upper level is increased, then $w^{(1)} \propto n^4$, and the main contribution to the change in probability is provided by the correction matrix element, $q^{(1)} \propto n^6$, which is the same for transitions of all types.

The above results together with the formulas of Table II and the numerical data of Table III for $n \geqslant 6$ can be used not only for hydrogen but also for many-electron atoms, since essentially all Rydberg states of a valence electron with $m \geqslant 3$ are hydrogen-like.[2,3]

3. CORRECTIONS TO THE MATRIX ELEMENTS OF TRANSITIONS FROM DEGENERATE STATES

The above data do not allow for that fact that excited hydrogen-like levels are ordinarily degenerate, so that in pure form there are no nlm-states even in a weak magnetic field, with the exception of four states with $m \geqslant n-3$, for which the analytic relations and the numerical data are listed in Tables II and III. When $m < n-3$, only the magnetic quantum number m and the parity P are exact constants of the motion. Within the scope of perturbation theory techniques, the principal quantum number is an approximate constant of the motion. On the other hand, the orbital angular momentum l ceases to be a constant of the motion even in arbitrarily weak fields, since all states of a given n-shell with differing values of l have the same energy, and the diamagnetic interaction (3) is not spherically symmetric. A magnetic field mixes states with different angular momenta but the same parity [by virtue of the even parity of the diamagnetic operator (3)], and the atom goes into a state with parity $P=(-1)^m$, which is a superposition of states with $l=m$, $m+2,\cdots$, $n-1$, or into a state with parity $P=(-1)^{m+1}$, which is a superposition of states with $l=m+1$, $m+3,\cdots$, $n-1$. The general properties of such states had been studied by the early 1980s, and are described, for example, in Braun's review.[2]

Thus, highly excited levels ($|np\rangle$, in particular) are only a component of the Zeeman state $|nm\lambda^-\rangle$ with $m=0,1$ and negative parity, which also incorporates states with $l=3,5,\ldots$, while the levels $|ns\rangle$ and $|nd\rangle$ are components of the $|nm\lambda^+\rangle$ state with $m=0$ and positive parity, which also incorporates components with $l=4,6,\ldots$. The number of $|nm\lambda^{\pm}\rangle$ states (below we call them diamagnetic states and label them with the ''diamagnetic quantum number'' λ) is $n-m$ (an equal number of even- and odd-parity states if n and m have the same parity, or the number of states with parity $P=(-1)^m$ is greater by one than the number of states with parity $P=(-1)^n$). All these levels have different quadratic shifts, i.e., the diamagnetic interaction lifts the degeneracy in the hydrogen spectrum completely. Thus, the redistribution of the intensity of dipole radiation among the diamagnetic states $|nm\lambda\rangle$ in the Lyman series (corresponding to $|np\rangle \to |1s\rangle$) is determined by the contribution of the np-wave function of the spherical Coulomb basis. In the zeroth approximation, this contribution is independent of the field strength B, since the energy of degenerate states is the same, and the transition to diamagnetic states merely changes the symmetry of the wave function. This situation is

fully analogous to the redistribution of transition probabilities among the Stark levels in an electric field, which are described by parabolic quantum numbers.[16]

As B increases, both the matrix element of the $|np\rangle \rightarrow |1s\rangle$ transition (described in Sec. 2) and the contribution of the np-state to the diamagnetic wave function change. Moreover, the interaction (3) also induces transitions from states with $l=3,5,\ldots$, which enter into the wave function of the zeroth approximation as well. Hence, in addition to corrections to the matrix elements of type (9)–(11), for transitions from degenerate states two new types of transition related to the above causes emerge. To calculate these corrections, we use higher-order perturbation theory for degenerate states.[6]

The wave function of a diamagnetic state with given principal and magnetic quantum numbers satisfies the integral equation

$$\psi_{nm\lambda p}(\mathbf{r})=\sum_{l=m+p}^{l_{\max}} a_l(\lambda)[1+G'_E(\mathbf{r},\mathbf{r}') \times V_D(\mathbf{r}')]^{-1}|\varphi_{nlm}(\mathbf{r}')\rangle, \tag{23}$$

where $\varphi_{nlm}(\mathbf{r})$ is the wave function of the unperturbed state with given orbital angular momentum, $p=0(1)$ determines the parity $P=(-1)^{m+p}$ of the state and the number of l's in the sum, and the parameter λ labels the diamagnetic states; the number of states is equal to the multiplicity of the degeneracy, i.e., the number of terms in the sum (23). The upper summation limit $l_{\max}$ is equal to $n-2$ if the parities $m+p$ and n are the same, and to $n-1$ if the parities $m+p$ and n are opposite. The reduced Green's function $G'_E(\mathbf{r},\mathbf{r}')$ is orthogonal to the states of the unperturbed basis and allows for mixing by the magnetic field of the states from the complete sets not incorporated in the basis, including the continuum. In the zeroth approximation in V_D the term with the Green's function can be discarded, with the wave function remaining a superposition of wave functions with the same value of the principal quantum number n but with different orbital angular momenta.

Thus, in contrast to nondegenerate states, a radiative transition in a magnetic field occurs between atomic states with no definite orbital angular momentum. This leads to a situation in which a transition to the lower level, which in the absence of a magnetic field can originate only from a state φ_{nlm} with a given orbital angular momentum l and must meet the dipole selection rules, in a magnetic field originates from almost all mixed states (23) in which $a_l \neq 0$. Here the intensity of radiation from the $|nm\lambda\rangle$ state constitutes only a fraction of the intensity of a transition from the ''pure'' state φ_{nlm} proportional to $|a_l(\lambda)|^2$.

The coefficients a_l of the superposition (23) satisfy a system of homogeneous algebraic equations, which can be obtained by substituting the wave function (23) into the Schrödinger equation and projecting the result onto the spherical basis states $|\varphi_{nlm}\rangle$:

$$(E_n-E)a_l+\sum_{l'=m+p}^{l_{\max}} a_{l'}\langle\varphi_{nlm}(\mathbf{r})|\hat{W}(\mathbf{r},\mathbf{r}')|\varphi_{nl'm}(\mathbf{r}')\rangle=0,$$
$$l=m+p,m+p+2,\ldots,l_{\max}, \tag{24}$$

where

$$\hat{W}(\mathbf{r},\mathbf{r}')=V_D(\mathbf{r})[1+G'_E(\mathbf{r},\mathbf{r}')V_D(\mathbf{r}')]^{-1} =V_D(\mathbf{r})\sum_{s=0}^{\infty}[-G'_E(\mathbf{r},\mathbf{r}')V_F(\mathbf{r}')]^s \tag{25}$$

is the integral operator that exactly accounts for the diamagnetic interaction.

In setting up the perturbation theory series in powers of V_D, we must allow for the dependence of the Green's function G'_E on V_D (with E the energy of the atom in the field). We then set up a series in powers of B^2 for the matrix

$$W_{ll'}=\langle\varphi_{nlm}(\mathbf{r})|\hat{W}(\mathbf{r},\mathbf{r}')|\varphi_{nl'm}(\mathbf{r}')\rangle, \tag{26}$$

which determines the coefficients of the system of homogeneous algebraic equations (24). Thus, the perturbation theory for the energies and wave functions of the diamagnetic states of the hydrogen atom reduces to a perturbation for the eigenvectors and eigenvalues of the diamagnetic matrix (26), with the new theory of perturbations represented by a series in powers of B^2.

By analogy with the series expansion of diamagnetic energy,

$$\Delta E_\lambda=E_\lambda-E_n=-\sum_{s=1}^{\infty}\frac{\chi_\lambda^{(s)}}{2s!}B^{2s}, \tag{27}$$

where $\chi_\lambda^{(s)}$ is s-order susceptibility, we can write the series for $W_{ll'}$ in the form

$$W_{ll'}=-\sum_{s=1}^{\infty}\frac{v_{ll'}^{(s)}}{2s!}B^{2s}. \tag{28}$$

The coefficients a_l must also be expanded in power series, which in contrast to the quantities (27) and (28) (which vanish as $B\rightarrow 0$) begin with the zeroth term:

$$a_l=\sum_{s=0}^{\infty}a_l^{(s)}B^{2s}, \tag{29}$$

so that at $B=0$ the coefficients remain finite, i.e., $\lim_{B\rightarrow 0}a_l=a_l^{(0)}$. Substituting (27)–(29) into the left-hand side of Eq. (24) and setting the coefficient of each power of the parameter B^2 to zero, we obtain a sequence of equations for the $a_l^{(s)}$. In the lowest order we have

$$\sum_{l'=m+p}^{l_{\max}} a_{l'}^{(0)}(v_{ll'}^{(1)}-\chi^{(1)}\delta_{ll'})=0. \tag{30}$$

Thus, the set of the coefficients $a_l^{(0)}$ is an eigenvector and $\chi^{(1)}$ the eigenvalue of the matrix $v_{ll'}^{(1)}$, whose elements are given by the analytic expression in (8). Equation (30) clearly shows that the expansions (27)–(29) remove the magnetic field B from Eqs. (24), so solving these equations amounts to solving an eigenvalue problem for a matrix independent of the magnetic field. The eigenvalues of this matrix are the first-order diamagnetic susceptibilities $\chi_\lambda^{(1)}$, where $\lambda=1,2,\ldots,K$, with K being the multiplicity of the degeneracy, i.e., the number of terms in the sum (23). At $K=1$ we

have $\chi_\lambda^{(1)}=v_{ll}^{(11)}$. At $K=2$ we can also write analytic expressions for the susceptibilities as functions of the principal quantum number n (see Refs. 4, 6, and 9). The eigenvectors are the sets of the zeroth-order coefficients in the superposition, i.e., the field-independent part of the series (29).

The values of $\chi_\lambda^{(1)}$ and $a_l^{(0)}(\lambda)$ obtained in this manner are used to solve the system of equations (24) in the next order in B^2, with the system being

$$\sum_{l'=m+p}^{l_{\max}} a_{l'}^{(1)}(\lambda)(v_{ll'}^{(1)}-\chi_\lambda^{(1)}\delta_{ll'}) = \sum_{l'=m+p}^{l_{\max}} a_{l'}^{(0)}(\lambda)(\chi_\lambda^{(2)}\delta_{ll'}-v_{ll'}^{(2)}). \tag{31}$$

Multiplying both sides of (31) by $a_l^{(0)*}(\lambda)$, summing over l, and taking into account (30) and the completeness condition $\Sigma|a_l^{(0)}|^2=1$, we obtain an expression for the second-order susceptibility:

$$\chi_\lambda^{(2)}=\sum_{l=m+p}^{l_{\max}}\sum_{l'=m+p}^{l_{\max}} a_l^{(0)*}(\lambda)a_{l'}^{(0)}(\lambda)v_{ll'}^{(2)}. \tag{32}$$

Thus, the second-order diamagnetic susceptibility is determined by the eigenvector $a_{l'}^{(0)}(\lambda)$ of the first-order matrix (8) and the second-order matrix elements. After integrating with respect to the angular variables, we obtain explicit expressions for these matrix elements:

$$v_{ll}^{(2)}=\beta_{nl}^{(0)}+\frac{3m^2-l(l+1)}{l(2l-1)}\beta_{nl}^{(2)}+\frac{3(l^2+2l-5m^2)(l^2-5m^2-1)-10m^2(4m^2-1)}{l(2l-1)(2l-2)(2l-3)}\beta_{nl}^{(4)}, \tag{33}$$

$$v_{ll+2}^{(2)}=-\left[\frac{((l+1)^2-m^2)((l+2)^2-m^2)(n^2-(l+1)^2)(n^2-(l+2)^2)}{(2l-1)^2(2l+1)(2l+3)^2(2l+5)(2l+7)^2}\right]^{1/2}\frac{n^6}{32}[\gamma_{nl}^{(0)}+m^2\gamma_{nl}^{(2)}], \tag{34}$$

$$v_{ll+4}^{(2)}=\left[\frac{(l+1-m)_4(l+1+m)_4(n+l+1)_4(n-l-4)_4}{(2l+1)(2l+3)^2(2l+5)^2(2l+7)^2(2l+9)}\right]^{1/2}\frac{205n^6}{128}. \tag{35}$$

The m-independent parameters in the expressions are superpositions of second-order radial matrix elements of the operator r^2, which by virtue of orthogonality relations of type (17), can be written in closed form as functions of the principal and orbital quantum numbers:[6]

$$\beta_{nl}^{(0)}=\frac{n^6}{240}\{5n^2[97n^2-33l(l+1)+365] - 12(21l^4+42l^3+179l^2+158l+60)\}, \tag{36}$$

$$\beta_{nl}^{(2)}=\frac{ln^6}{336(2l+3)}\{n^2[802n^2-1005l(l+1)+2000] + 3(175l^4+350l^3+523l^2+348l-480)\}, \tag{37}$$

$$\beta_{nl}^{(4)}=\frac{l(l-1)n^6}{1120(2l+3)(2l+5)}\{15n^2[41n^2+22l(l+1) - 185]-1001l^3(l+2)+3421l^2+4422l-2160\}, \tag{38}$$

$$\gamma_{nl}^{(0)}=n^2[429l(l+3)-802]+237l^3(l+6) + 3145l^2+3036l-2484, \tag{39}$$

$$\gamma_{nl}^{(2)}=5(41n^2+37l^2+111l-75). \tag{40}$$

Having calculated the susceptibility $\chi_\lambda^{(2)}$, we can solve the system of equations (31) for the corrections $a_{l'}^{(1)}(\lambda)$ to the coefficients (29). However, the determinant of the system (31) exactly equals the determinant of the system of homogeneous equations (30) and consequent ly vanishes. Hence, in addition to (31), we must use the condition of conservation of the normalization of the wave function (23) in the first order in V_D:

$$\sum_{l=m+p}^{l_{\max}} a_l^{(0)}(\lambda)a_l^{(1)}(\lambda)=0. \tag{41}$$

The values of the susceptibilities (32) together with the coefficients $a_l^{(1)}(\lambda)$ must be calculated for each eigenvector of the diamagnetic matrix (8) corresponding to the specific eigenvalue $\chi_\lambda^{(1)}$. The $a_l^{(1)}(\lambda)$ determine the first-order corrections in B^2 to the wave function (23) of a degenerate state, which are due to the variation in a strong field of the contribution of the basis vectors φ_{nlm}. The variation of the contribution of the other states, which do not belong to the degenerate basis, is determined by the term with the Green's function. This part is fully analogous to the corrections for nondegenerate states.

Since in all transitions belonging to the Lyman and Balmer series the lower levels are nondegenerate, we limit ourselves to the study of the most interesting transition in practice, from the degenerate state $|nm\lambda\rangle$ to the nondegenerate state $|n'l'm'\rangle$. In the zeroth approximation the expression for the matrix element is given by a superposition of the form

$$d_{10}^{(0)}(\lambda)=\langle nm\lambda|d_z|n'l'm'\rangle = a_{l'-1}^{(0)}(\lambda)\langle\varphi_{nl'-1m}|d_z|\varphi_{n'l'm'}\rangle+a_{l'+1}^{(0)}(\lambda) \times\langle\varphi_{nl'+1m}|d_z|\varphi_{n'l'm'}\rangle. \tag{42}$$

When $m'=l'$, the first term on the right-hand side of (42) vanishes. Thus, the transformation of the excited states leads to a splitting of the line of the transition $|nlm\rangle\to|n'l'm\rangle$ into the diamagnetic components $|nm\lambda\rangle\to|n'l'm\rangle$, whose number is determined by the multiplicity K of the degeneracy ($\lambda=1,\ldots,K$), whose frequencies are determined by the eigenvalues $\chi_\lambda^{(1)}$, and whose intensities by the eigenvector components $a_{l'+1}^{(0)}$ of the diamagnetic matrix (8). In other words, the action of a magnetic field on the spectrum of the atom results in a redistribution of the intensity $I_{n,n'l'}$ of the $|nl'+1m\rangle\to|n'l'm\rangle$ transition among the diamagnetic sublevels $|nm\lambda\rangle$ with different quantum numbers $\lambda=1,\ldots,K$ in proportion to the contribution of the state $|nl'+1m\rangle$ to these sublevels, i.e., $I_{nm\lambda,n'}=|a_{l'+1}^{(0)}(\lambda)|^2 I_{n,n'l'}$. This relationship characterizes the B-independent distribution of the oscillator strengths of dipole transitions among the diamagnetic sublevels $|nm\lambda\rangle$ of the excited state,[2,3] similar to the distribution among the Stark states of a parabolic basis in an electric field.[16]

As B increases, so does the contribution of the field-dependent corrections to the matrix elements. The expression for these corrections in the first order in B^2 is

$$\begin{aligned}d_{10}^{(1)}(\lambda)=&\,a_{l'+1}^{(1)}B^2\langle\varphi_{nl'+1m}|d_z|\varphi_{n'l'm'}\rangle\\&+a_{l'+1}^{(0)}(\lambda)B^2[\langle\varphi_{nl'+1m}|V_D G'_{E_n}d_z|\varphi_{n'l'm'}\rangle\\&+\langle\varphi_{nl'+1m}|d_z G'_{E_n}d_z|\varphi_{n'l'm'}\rangle]\\&+a_{l'+3}^{(0)}(\lambda)B^2[\langle\varphi_{nl'+3m}|V_D G'_{E_n}d_z|\varphi_{n'l'm'}\rangle\\&+\langle\varphi_{nl'+3m}|d_z G'_{E_n}d_z|\varphi_{n'l'm'}\rangle].\end{aligned}\tag{43}$$

Dividing this by $d_{10}^{(0)}(\lambda)$, we can write the factor of the first B-dependent term in the expansion of the matrix element in (6) in the form

$$q^{\pi}_{n\lambda,n'l'}=q^{\pi}_{nl'+1,n'l'}+\frac{a_{l'+1}^{(1)}(\lambda)}{a_{l'+1}^{(0)}(\lambda)}+\frac{a_{l'+3}^{(0)}(\lambda)}{a_{l'+1}^{(0)}(\lambda)}q^{\pi}_{nl'+3,n'l'},\tag{44}$$

where $q^{\pi}_{nl'+1,n'l'}=q^{\pi}_{nl'+1,n'l'}(n)+q^{\pi}_{nl'+1,n'l'}(n')$ is the factor for the transition between nondegenerate states, whose components are given in (12). Similar expressions can be written for the components of the factor $q^{\pi}_{nl'+3,n'l'}=q^{\pi}_{nl'+3,n'l'}(n)+q^{\pi}_{nl'+3,n'l'}(n')$:

$$\begin{aligned}q^{\pi}_{nl'+3,n'l'}(n)=&\,\frac{1}{4(2l'+5)}\sqrt{\frac{3(l'+1)}{2l'+7}}\\&\times\frac{\langle nl'+3|r^2 g_{l'+1}^{(n)}(r,r')r'|n'l'\rangle}{\langle nl'+1|r|n'l'\rangle},\\q^{\pi}_{nl'+3,n'l'}(n')=&\,\frac{1}{4(2l'+5)}\sqrt{\frac{3(l'+1)}{2l'+7}}\\&\times\frac{\langle nl'+3|r g_{l'+2}^{(n')}(r,r')r'^2|n'l'\rangle}{\langle nl'+1|r|n'l'\rangle}.\end{aligned}\tag{45}$$

For each specific set of values of the quantum numbers of the lower level $n'l'$ there are simple formulas for these parameters. In particular, for transitions belonging to the Lyman series we have

$$\begin{aligned}q^{\pi}_{nf,1s}(n)=&\,\frac{n^4\sqrt{3(n^2-4)(n^2-9)/7}}{480Z^4(n^2-1)^3}\\&\times(55n^6+115n^4-411n^2-527),\\q^{\pi}_{nf,1s}(1s)=&\,\frac{n^2\sqrt{3(n^2-4)(n^2-9)/7}}{30Z^4(n^2-1)^3}(49n^2-1).\end{aligned}\tag{46}$$

For $|n0\lambda^-\rangle\to|2s\rangle$ transitions of the Balmer series we have

$$\begin{aligned}q^{\pi}_{nf,2s}(n)=&\,\frac{n^4\sqrt{3(n^2-4)(n^2-9)/7}}{480Z^4(n^2-4)^4}\\&\times(55n^8-80n^6-4832n^4+81\,664n^2+134\,912),\\q^{\pi}_{nf,2s}(2s)=&-\frac{256n^2\sqrt{3(n^2-4)(n^2-9)/7}}{15Z^4(n^2-4)^4}\\&\times(7n^2+4)(n^2+2).\end{aligned}\tag{47}$$

For $|n1\lambda^+\rangle\to|2p1\rangle$ transitions we have

$$\begin{aligned}q^{\pi}_{ng,2p}(n)=&\,\frac{n^4\sqrt{2(n^2-9)(n^2-16)/3}}{672Z^4(n^2-4)^3}(125n^6-220n^4\\&-4752n^2-34\,624),\\q^{\pi}_{ng,2p}(2p)=&-\frac{64n^2\sqrt{2(n^2-9)(n^2-16)/3}}{21Z^4(n^2-4)^3}(25n^2-4).\end{aligned}\tag{48}$$

Note that all these expressions have similar patterns, as pointed out earlier for the components of the factors (19) and (20): as n increases, the component $q^{\pi}(n)$ in (46)–(48) increases as n^6, while $q^{\pi}(n')$ remains essentially the same at large values of n. Moreover, the expression for the sum $q^{\pi}=q^{\pi}(n)+q^{\pi}(n')$ becomes much simpler in comparison to the expressions for the individual components:

$$\begin{aligned}q^{\pi}_{nf,1s}=&\,q^{\pi}_{nf,1s}(n)+q^{\pi}_{nf,1s}(1s)\\=&\,\frac{n^2\sqrt{3(n^2-4)(n^2-9)/7}}{480(n^2-1)}(55n^4+225n^2-16),\\q^{\pi}_{nf,2s}=&\,q^{\pi}_{nf,2s}(n)+q^{\pi}_{nf,2s}(2s)=\frac{n^2\sqrt{3(n^2-4)(n^2-9)/7}}{480(n^2-4)^2}\\&\times(55n^6+360n^4-2832n^2-4096),\\q^{\pi}_{ng,2p}=&\,q^{\pi}_{ng,2p}(ng)+q^{\pi}_{ng,2p}(2p)\\=&\,\frac{n^2\sqrt{2(n^2-9)(n^2-16)/3}}{672(n^2-4)}\\&\times(125n^4+780n^2-512).\end{aligned}\tag{49}$$

It must also be noted that the factor $q_{nl'+3,n'l'}$ in the case of nondegenerate states determines the matrix element of a magnetically induced dipole transition, which is forbidden in

TABLE IV. The factors $q^{\pi}_{nm\lambda^{\pm},n'l'}$ for magnetically induced corrections to the matrix elements of the radiative π-transitions of the Lyman series, $|n0\lambda\rangle\to|1s\rangle$, and of the Balmer series, $|n0\lambda\rangle\to|2s\rangle$ and $|nm\lambda\rangle\to|2pm\rangle$, with $m=0,1$. The index λ is determined by the absolute value represented in individual columns by the diamagnetic susceptibility of the upper level, $\chi^{(1)}_{nm\lambda^{\pm}}$; $(k)=10^k$.

n	λ	$-\chi^{(1)}_{n0\lambda^-}/2$	$q^{\pi}_{n0\lambda,1s}$	$q^{\pi}_{n0\lambda,2s}$	$-\chi^{(1)}_{n0\lambda^+}/2$	$q^{\pi}_{n0\lambda,2p0}$	$-\chi^{(1)}_{n1\lambda^+}/2$	$q^{\pi}_{n1\lambda,2p1}$
	1	9.0	1.368(2)	9.107(1)	5.171	−1.179(2)	9.0	6.609(1)
3	2				1.958(1)	1.920(3)		
	1	1.335(1)	1.525(3)	1.439(3)	1.868(1)	−1.014(3)	3.60(1)	6.293(2)
4	2	3.865(1)	5.396(2)	1.646(2)	6.532(1)	1.044(4)		
	1	3.425(1)	4.451(3)	4.158(3)	2.834(1)	1.539(2)	4.438(1)	7.788(3)
5	2	1.095(2)	2.776(3)	1.701(3)	6.875(1)	−2.033(3)	1.056(2)	2.666(3)
	3				1.654(2)	3.579(4)		
	1	5.215(1)	2.070(4)	2.025(4)	5.759(1)	−4.188(3)	1.056(2)	2.021(4)
6	2	1.150(2)	4.168(3)	3.098(3)	1.710(2)	−3.473(4)	2.454(2)	9.364(3)
	3	2.513(2)	8.941(3)	6.529(3)	3.519(2)	9.876(4)		
	1	9.064(1)	3.912(4)	3.831(4)	8.631(1)	5.290(3)	1.396(2)	8.490(4)
	2	2.545(2)	2.200(4)	1.964(4)	1.842(2)	−5.906(4)	2.507(2)	3.762(4)
7	3	5.001(2)	2.393(4)	1.923(4)	3.636(2)	−1.142(5)	4.918(2)	2.571(4)
	4				6.643(2)	2.349(5)		
	1	1.322(2)	1.070(5)	1.058(5)	1.352(2)	−6.683(3)	2.444(2)	1.468(5)
	2	2.842(2)	3.637(4)	3.379(4)	3.655(2)	−1.145(5)	4.992(2)	8.790(4)
8	3	5.081(2)	4.635(4)	4.168(4)	6.854(2)	−3.249(5)	8.885(2)	6.056(4)
	4	8.995(2)	5.549(4)	4.717(4)	1.150(3)	5.003(5)		

the absence of a magnetic field, and the above expression can serve as a means of numerically estimating the values of such matrix elements.

We also write the expression for the factor q corresponding to the transition of the Balmer series from the even component of Zeeman states with $m=0$, i.e., to the $|n0\lambda^+\rangle\to|2p0\rangle$ transition:

$$q^{\pi}_{n0\lambda^+,2p0}=\left[q^{\pi}_{nd0,2p0}+\frac{\sqrt{5}a_0^{(0)}}{2a_2^{(0)}}q^{\pi}_{ns,2p}+\frac{a_2^{(1)}(\lambda)}{a_2^{(0)}}+\frac{a_0^{(1)}(\lambda)}{8a_2^{(0)}}\sqrt{\frac{5(n^2-4)}{n^2-1}}+2\sqrt{\frac{3}{10}}\frac{a_4^{(0)}}{a_2^{(0)}}q^{\pi}_{ng,2p}\right]\times\left[1+\sqrt{\frac{5(n^2-4)}{n^2-1}}\frac{a_0^{(0)}}{8a_2^{(0)}}\right]^{-1},\qquad(50)$$

where $q^{\pi}_{ns,2p}$, $q^{\pi}_{nd0,2p0}$, and $q^{\pi}_{ng,2p}$ are defined in (21), (22), and (49).

In the general case of $m\leqslant l'-3$, the right-hand side of Eq. (43) acquires three terms, which can be obtained from those that have already been written if we replace $l'+1$ with $l'-1$ and $l'+3$ with $l'-3$; these correspond to the contribution to the diamagnetic state $|nm\lambda\rangle$ of the states $|nlm\rangle$ with orbital angular momenta $l=l'-1$ and $l=l'-3$.

4. RESULTS OF NUMERICAL CALCULATIONS AND DISCUSSION

The expressions (44)–(50) are sufficient to calculate the numerical values of the factor q^{π} of any transitions belonging to the first two series of a hydrogen-like atom. We carried out calculations of the q^{π} for the $|n0\lambda^-\rangle\to|1s\rangle$ transitions of the Lyman series and of three types of transition, $|n0\lambda^-\rangle\to|2s\rangle$, $|n0\lambda^+\rangle\to|2p0\rangle$, and $|n1\lambda^+\rangle\to|2p1\rangle$, of the Balmer series.

The transitions to the $1s$ and $2s$ states originate from the same excited states of negative parity, $|n0\lambda^-\rangle$, are determined by the same coefficients $a_1^{(0)}$, $a_3^{(0)}$, and $a_1^{(1)}$ of the wave function (23) of the upper level, and differ only in the parameters $q^{\pi}_{np,n's}$ and $q^{\pi}_{nf,n's}$. Formulas (19)–(20), and (49) show that for large values of n the difference between these parameters disappears. Hence the correction factors (44) for the matrix elements of radiative transitions from highly excited magnetic states in the Lyman and Balmer series are essentially identical.

The difference in the π-transitions into the states $2p_0$ and $2p_1$ is not only quantitative but also qualitative even in the zeroth approximation, since the former, in contrast to the latter, contains a contribution corresponding to transitions from ns states. This fact manifests itself in the dependence of the factor (50) on the coefficients $a_0^{(0)}$ and $a_0^{(1)}$ and also on the parameter $q^{\pi}_{ns,2p}$, which has negative values that exceed in absolute value the positive values of the parameter $q^{\pi}_{nd0,2p0}$ by about a factor of 3 (see Eqs. (21) and (22) and the data given in Table I).

Table IV lists the numerical values of the factors $q^{\pi}_{nm\lambda,n'l'm}$ for the transitions of the Lyman and Balmer series from states with $n\leqslant 8$. To label the transitions (so as to determine the numerical value of the parameter λ), we use the energy distribution, which in the first order is determined by the susceptibility $\chi^{(1)}_{nm\lambda}$ of the upper level with the principal quantum number n. The numerical values of this susceptibility are listed in Table IV together with the factors q for each group of transitions that originate from the same excited

states. Note the existence of a correlation between the magnetically induced corrections to the energy, $\chi^{(1)}_{nm\lambda}$, and to the matrix elements of the radiative transitions, $q^{\pi}_{nm\lambda,n'l'm}$. As for the energy, the problem is a trivial one, since at large values of n the diamagnetic susceptibility $\chi^{(1)}_0$ of the lower level is small compared to the susceptibility $\chi^{(1)}_1$ of the upper, so that in the expression for the correction [Eq. (7)] we can neglect χ_0. Thus, the correction coefficient $w^{(1)}_{nm\lambda,n'}$ for the transition frequencies of the diamagnetic lines if the Lyman series is approximately four times smaller than the corresponding coefficient of the Balmer series (in proportion to the ratio of the unperturbed binding energies of the states with $n=1$ and $n=2$). The fact that the coefficients (44) for the Lyman and Balmer series at large values of n are identical is not obvious, since even the unperturbed matrix elements in the series differ in the asymptotic region by a factor of larger than 20 (see Ref. 16). Nevertheless, the proximity of the coefficients $q^{\pi}_{nm\lambda,1s}$ and $q^{\pi}_{nm\lambda,2s}$ is obvious from the data of Table IV by $n=5$. Here the difference between the coefficients is at its minimum for the lower diamagnetic components of the emission line ($\lambda=1$) and at its maximum for the higher diamagnetic components of the emission line ($\lambda=K$). Calculations show that for $n\geqslant 30$ the maximum difference does not exceed 1%: $(q^{\pi}_{nmK,1s}-q^{\pi}_{nmK,2s})/q^{\pi}_{nmK,1s}\leqslant 0.01$.

The decay of excited states of positive parity with $m=0$, which contributes nothing to the Lyman series, can proceed by a π-transition of the Balmer series to the $|2p0\rangle$ state. Despite the fact that the susceptibilities of $|n0\lambda^{+}\rangle$ states differ from the susceptibilities of the diamagnetic levels $|n0\lambda^{-}\rangle$ by a factor no greater than two, as one can see from Table IV (for $n\gg 1$ the susceptibilities of the lowest diamagnetic levels of opposite parities, the so-called diamagnetic doublets, are essentially identical[2]), the correction factors q for the $|n0\lambda^{+}\rangle\rightarrow|2p0\rangle$ transitions differ substantially from the corresponding factors for the $|n0\lambda^{-}\rangle\rightarrow|2s\rangle$ transitions in both absolute value and sign.

The ground state in the set with $m=1$ is $|2p1\rangle$, with the result that π-transitions within this set are encountered only in the Balmer series. The correction factor $q^{\pi}_{n1\lambda,2p1}$ is determined by (44), as for transitions to the $2s$ states. Hence both quantitatively and qualitatively the distribution of the given factor among the even-parity diamagnetic sublevels $|n0\lambda^{+}\rangle$ is identical to the distribution of the factor $q^{\pi}_{n0\lambda,2s}$ among the odd-parity sublevels. These distributions are compared in Fig. 1, which depicts the numerical values of the factors $q^{\pi}_{n1\lambda,2p1}$ (dark circles) and $q^{\pi}_{n0\lambda,2s}$ (open circles) for the matrix elements of the diamagnetic components of the Balmer series, corresponding to the radiative decay of states with a principal quantum number $n=40$. The resonance-like jump in the region of the transition from the doublet diamagnetic spectrum to the split states of different parities[2] is observed in both curves, symbolizing the dependence on the diamagnetic energy, which is represented on the horizontal axis by the first-order susceptibility $\chi^{(1)}_{nm\lambda}$.

Figure 1 also depicts the numerical values of the factors $q^{\pi}_{n0\lambda,2p0}$ for the transitions from even-parity states in the doublet region of the spectrum (the $\times$'s). These transitions also

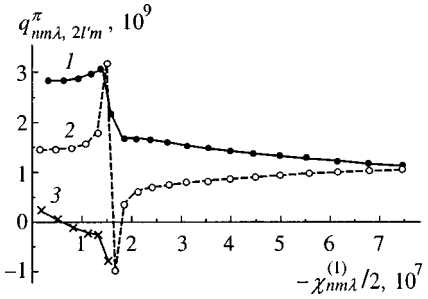


FIG. 1. Numerical values of the correction factors $q^{\pi}_{n1\lambda,2p1}$ (dark circles connected by the solid curve *1*) and $q^{\pi}_{n0\lambda,2s}$ (open circles connected by the dashed curve *2*) of the diamagnetic components of the Balmer line corresponding to the radiative decay of the level with $n=40$. The $\times$'s connected by the curve *3* depict the first six values for $q^{\pi}_{n0\lambda,2p0}$ corresponding to those depicted in Fig. 2.

exhibit a sudden change in the nature of the energy dependence in the vicinity of the transition from doublet to split states, accompanying the crossover from positive to negative values of the factors. The entire set of values of $q^{\pi}_{n0\lambda,2p0}$ is depicted in Fig. 2 with a 20-fold change in scale along the vertical axis in comparison to Fig. 1. As Fig. 2 shows, the more sudden jump in the values of $q^{\pi}_{n0\lambda,2p0}$ occurs in the high-frequency part of the diamagnetic spectrum, i.e., approximately at a distance of one-fifth of the entire width of the diamagnetic set from the upper limit (on the energy scale). Here for almost all states below this jump the factor $q^{\pi}_{n0\lambda,2p0}$ is negative, and since its absolute value exceeds the corresponding diamagnetic susceptibilities by a factor of 100 or more (see the scaling factors for the quantities along both axes), the intensity of the corresponding lines decreases as the magnetic field becomes stronger. For the last three lines in the high-frequency part of the spectrum, as for essentially all lines whose diamagnetic corrections are depicted in Fig. 1, the intensity increases with field strength, exceeding the intensity of the corresponding radiation emitted by a free atom in a field $B\approx 1$ T by 30–40%.

5. CONCLUSION

In a sufficiently strong magnetic field, there can be not only complete splitting of the emission lines of atoms but the probability of the atoms emitting and absorbing radiation may be redistributed dramatically. The intensity of essen-

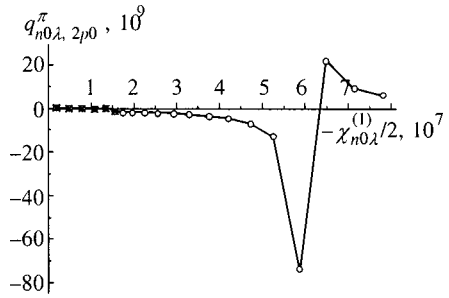


FIG. 2. Numerical values of the correction factors $q^{\pi}_{n0\lambda,2p0}$ of the diamagnetic components of the Balmer line corresponding to the radiative transition from the even-parity states $|n0\lambda^{+}\rangle$ of the level with $n=40$. The first six values (the $\times$'s) are plotted in Fig. 1 on a larger scale.

tially all π-components of the Lyman series increases with field strength. In the Balmer series the same increase in intensity is observed for transitions in the Zeeman set with $m=1$. In the set with $m=0$ there is a redistribution of intensities in favor of the decay lines of odd-parity states and a small number of states from the low- and high-frequency parts of the diamagnetic set of positive parity. On the other hand, the intensity of the lines of decay of even-parity states from the middle of the diamagnetic spectrum decreases with increasing field strength.

Our calculations show that the changes in the radiative properties of a hydrogen-like atom in a magnetic field can be described analytically by using the spherical Coulomb basis. We did so and arrived at a complete solution of the problem of the hydrogen atom in a magnetic field with allowance for diamagnetic corrections not only for the energy but also for the wave functions. The method of determining the corrections to optical characteristics of degenerate states of the atom in a field by solving the system of algebraic equations (24) combined with the expressions for higher-order matrix elements of the type (19)–(22), (33)–(40), and (46)–(49) could be useful in solving other problems of the interaction between excited atoms and electromagnetic fields.

The analytic expressions and quantitative data listed in the Table I–IV and Figs. 1 and 2 provide information about changes in the intensity of atomic lines. They can be applied not only to hydrogen but also to many-electron atoms in Zeeman states with large magnetic quantum numbers ($m \geqslant 3$) and can be used to describe the spectral characteristics of Rydberg atoms in moderate and strong magnetic fields, generated in laboratories on the Earth and produced in stellar plasma.

This work was sponsored by the Russian Fund for Fundamental Research (Grant No. 97-02-16407) and the Russian Ministry of Education (Grant No. 97-0-5.1-63).

*)E-mail: vit@ovd.vsu.ru

[1)]In this paper we use the atomic system of units.

[1] P. F. A. Klinkenberg, in *Atomic Physics 15 (invited papers of the 15th ICAP)*, World Scientific, Singapore (1997), p. 221.
[2] P. A. Braun, Rev. Mod. Phys. **65**, 115 (1993).
[3] D. Kleppner, N. G. Littman, and M. I. Zimmerman, in *Rydberg States of Atoms and Molecules*, R. F. Stebbings and F. B. Dunning (eds.), Cambridge Univ. Press, London (1983), p. 73.
[4] T. P. Grozdanov and H. S. Taylor, J. Phys. B **19**, 4075 (1986).
[5] E. A. Solov'ev, Zh. Eksp. Teor. Fiz. **82**, 1762 (1982) [Sov. Phys. JETP **55**, 1017 (1982)].
[6] V. D. Ovsiannikov and S. V. Goossev, Phys. Scr. **57**, 506 (1998).
[7] I. I. Sobelman, *Atomic Spectra and Radiative Transitions*, Springer-Verlag, Berlin (1979), Sec. 9.2.
[8] N. L. Manakov, V. D. Ovsyannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
[9] S. V. Goossev and V. D. Ovsiannikov, J. Phys. B **28**, 5251 (1995).
[10] V. M. Vaĭnberg, V. A. Gani, and A. E. Kudryavtsev, Zh. Éksp. Teor. Fiz. **113**, 550 (1998) [JETP **86**, 305 (1998)].
[11] V. D. Ovsiannikov, Phys. Rev. A **57**, 3719 (1998).
[12] P. Falsaperla and G. Fonte, Phys. Rev. A **50**, 3051 (1994).
[13] J.-H. Wang and C.-S. Hsue, Phys. Rev. A **52**, 4508 (1995).
[14] Yu. P. Kravchenko, M. F. Liberman, and B. Johansson, Phys. Rev. A **54**, 287 (1996).
[15] T. P. Grozdanov, L. Andric, C. Manescu, and R. McCarroll, Phys. Rev. A **56**, 1865 (1997).
[16] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, Plenum Press, New York (1977).
[17] A. Erdélyi, *Higher Transcendental Functions* (Bateman Project) Vol. 1, McGraw-Hill, New York (1953), Chap 5; Vol. 2, Chap. 10.

Translated by Eugene Yankovsky

Escape into vacuum of fast electrons generated by oblique incidence of an ultrashort, high-power laser pulse on a solid target

A. A. Andreev,*) I. A. Litvinenko, and K. Yu. Platonov

All-Russian Scientific Center ''S. I. Vavilov State Optics Institute,'' 199034 St. Petersburg, Russia; Russian Federal Nuclear Center, All-Russian Scientific Research Institute of Technical Physics, 199034 St. Petersburg, Russia

(Submitted 29 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 1184–1197 (October 1999)

The possibility that fast electrons can escape in a direction close to the trajectory of a reflected ultrashort laser pulse at extremely high laser radiation fluxes is examined analytically and numerically. Analytic estimates are made of the feasibility of forming electron bursts in the plasma and of their subsequent motion. The self-consistent, collisionless motion of a plasma acted on by specified incident and reflected ultrashort laser pulses is modeled in two dimensions by the particle-in-cell method. It is shown that a substantial number of electrons located in the subcritical region are gathered into bunches by the resultant forces and escape to the vacuum in a direction different from the normal to the target surface within a narrow range of solid angles. This demonstrates the feasibility of laser acceleration of an electron burst during reflection of an ultrashort laser pulse from a solid target. © *1999 American Institute of Physics.* [S1063-7761(99)00510-7]

1. INTRODUCTION

The development of laser technology for generating high power ($q \geqslant 10^{18}$ W/cm^2), ultrashort ($\tau_i \leqslant 0.1$ ps) pulses[1] has meant that earlier ideas regarding the mechanisms by which this laser radiation interacts with matter are inadequate for explaining the experimentally observed physical effects. In particular, data on the angular distribution of the fast electrons escaping from a target under obliquely incident laser light have been obtained in numerical and real experiments.[2,3] Analysis of these data shows that the bulk of the fast electrons escaping from the target moved in the direction of the reflected laser pulse. One possible reason for the formation of these beams is the acceleration of electrons escaping from the laser plasma in the electromagnetic fields of the incident and reflected laser pulses. This mechanism differs from the acceleration of electrons in vacuum by a plasma wave, which has been studied experimentally and theoretically[4] for a laser pulse of duration $\leqslant 1$ ns and intensity 10^{15} W/cm^2. The interaction of this kind of pulse with a target leads to significant hydrodynamic outflow of the resulting plasma. In this weakly inhomogeneous (compared to the scale of the laser wavelength) plasma, intense plasma oscillations develop and accelerate electrons in the direction of the plasma wave, owing both to resonant absorption and stimulated Raman scattering.

This paper is devoted to a study of the dynamics of electrons in the incidence and reflection regions of an ultrashort laser pulse. For pulses of this type lasting less than 1 ps, plasma ions are unable to move and the plasma boundary remains sharp. In this sort of weakly inhomogeneous plasma, the generation of longitudinal plasma waves is suppressed and there is no resonant acceleration mechanism. In this paper, we assume that the laser plasma is collisionless and that the motion of electrons is determined by the external electromagnetic field and by the ambipolar potential owing to charge separation. We have studied the electron dynamics with the aid of analytic estimates and numerical calculations.

In Sec. 2 the trajectory of an electron is constructed analytically and its emergence angle and energy are estimated. This makes it possible to understand the physics of the phenomenon, and subsequently to interpret the results of the numerical calculations. In Sec. 3 the same problem is solved numerically using a particle-in-cell (PIC) code for particular parameters of the laser pulse and plasma. In Sec. 4 we discuss the results of the numerical calculations, compare the analytic and numerical results, and explain the resulting dependences.

2. ANALYTIC TREATMENT OF THE ESCAPE OF ELECTRONS FROM THE TARGET SURFACE

We consider the oblique incidence (at an angle θ) of a laser pulse of finite width D on a target surface. Here the following obvious distribution of fields develops (Fig. 1): region *i* contains the field of the incident wave, *o* that of the reflected wave, *b* a superposition of the incident and reflected waves, and *pl* the field within the plasma. We assume that near the target surface there is a small group of electrons and that those electrons have no effect on the distribution of the fields in the vacuum or in the plasma. This group results from the electron density profile being steepened as a result of the ponderomotive pressure of the laser pulse and a small number of electrons with a density below critical remains in the vacuum. This sort of distribution of electron density was obtained via one-dimensional numerical simulation,[5] in which processes within the plasma were the main concern. These electrons will also be examined subsequently. Since

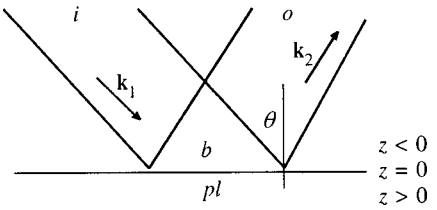


FIG. 1. Incidence of a laser beam on a plasma surface ($z=0$).

the longitudinal and transverse dimensions of the laser pulse are of the order of tens of wavelengths, the laser field consists of an electromagnetic wave with slowly varying amplitude $E_0(\mathbf{r},t)$. In dimensionless variables $t\leftrightarrow\omega t$, $y\leftrightarrow(\omega/c)y$, $z\leftrightarrow(\omega/c)z$, $E\leftrightarrow eE/m\omega c$, $B\leftrightarrow eB/m\omega c$, $A\leftrightarrow eA/mc^2$, $\phi\leftrightarrow e\phi/mc^2$, and $v\leftrightarrow v/c$, the laser fields corresponding to the various regions of Fig. 1 are

for region i

$$B_x^{(i)}=-E_0^{(i)}(\mathbf{r},t)\cos[z\cos\theta+y\sin\theta-t],$$

$$E_z^{(i)}=B_x^{(i)}\sin\theta,\quad E_y^{(i)}=-B_x^{(i)}\cos\theta; \tag{1}$$

for region o

$$B_x^{(o)}=-\eta E_0^{(o)}(\mathbf{r},t)\cos[-z\cos\theta+y\sin\theta-t+\alpha],$$

$$E_z^{(o)}=B_x^{(o)}\sin\theta,\quad E_y^{(o)}=-B_x^{(o)}\cos\theta; \tag{2}$$

and, for region b

$$B_x^{(b)}=B_x^{(i)}+B_x^{(o)},\quad E_z^{(b)}=E_z^{(i)}+E_z^{(o)},\quad E_y^{(b)}=E_y^{(i)}+E_y^{(o)}. \tag{3}$$

The reflection coefficient η and phase α of the reflected laser wave derive from matching of the electromagnetic fields (3) with the fields within the plasma in region *pl*. In general, the fields in region *pl* are self-consistent solutions of Maxwell's equations and the kinetic equations for the plasma particles. There are obviously no analytic expressions for η and α in some arbitrary case. In the case of resonant absorption in the normal skin effect regime,[6]

$$\eta=\frac{|\cos\theta-\zeta|}{|\cos\theta+\zeta|},\quad \alpha=\arg\left(\frac{\cos\theta-\zeta}{\cos\theta+\zeta}\right), \tag{4}$$

where

$$\zeta=-1.38i\xi^{1/3}\{1+\xi^{2/3}[0.72-\sin^2\theta(0.21+0.48\ln\xi)]\}+\pi\xi\sin^2\theta. \tag{5}$$

The parameter ξ is the dimensionless scale length for density inhomogeneities of the plasma: $\xi=\omega L/c$. To order of magnitude, L is the duration τ_i of the laser pulse multiplied by the ion sound speed c_s. Equation (5) holds for $\xi\leqslant 1$. Numerical calculations of η and α for reflection of a laser pulse in the anomalous skin effect regime can be found elsewhere.[5,7]

Having determined the parameters of the laser fields above the plasma surface, we now proceed to describe electron dynamics. In the fields (1)–(3), the trajectory of electrons is two-dimensional, and the equations of motion in region b have the form

$$\frac{dp_z}{dt}=(\sin\theta-v_y)(B_x^{(i)}+B_x^{(o)})-\frac{\partial U_{am}}{\partial z}, \tag{6}$$

and

$$\frac{dp_y}{dt}=v_z(B_x^{(i)}+B_x^{(o)})-\cos\theta(B_x^{(i)}-B_x^{(o)}), \tag{7}$$

where $B_x^{(i,o)}$ is the magnetic field in the corresponding region and $U_{am}(z,t)$ is the ambipolar potential of the plasma ions, which we determine below. In regions i and o, $B_x^{(o)}$ and $B_x^{(i)}$, respectively, are absent from Eqs. (6) and (7).

We first examine the solution of Eqs. (6) and (7) for nonrelativistic laser fields, $|B_x^{(i,o)}|<1$. Then $\mathbf{p}=\mathbf{v}$ in the chosen system of dimensionless variables. We distinguish oscillating (frequency ω) and mean components of the coordinates and velocity of the electron: $\mathbf{v}=\langle\mathbf{v}\rangle+\delta\mathbf{v}$. Then, averaging Eqs. (6) and (7) over the period of the oscillations, we obtain for the mean components,

$$\frac{d\langle\mathbf{v}\rangle}{dt}=-\frac{\partial(U_{\rm eff}+U_{am})}{\partial\mathbf{r}}, \tag{8}$$

where

$$U_{\rm eff}(\mathbf{r},t)=(E_0^{(i)})^2+\eta^2(E_0^{(o)})^2-\eta\cos(2\theta)E_0^{(i)}E_0^{(o)}\cos(2z\cos\theta-\alpha) \tag{9}$$

is the effective potential, which is the time-averaged square of the laser electric field:[8] $U_{\rm eff}=\langle\mathbf{E}^2\rangle/2$.

The potential (9) contains three substantially different time scales: 1) the laser wavelength; 2) the transverse extent of the laser beam, which is tens of wavelengths; and 3) the transverse scale length of the laser pulse, which is hundreds of wavelengths. The greatest contribution to the force on an electron in Eq. (8) is obtained by differentiating the potential (9) with respect to the very shortest scale length. In this approximation, Eq. (8) takes the form

$$\frac{d\langle v_z\rangle}{dt}=-\eta\cos(2\theta)\cos\theta E_0^{(i)}E_0^{(o)}\times\sin(2z\cos\theta-\alpha)-\frac{\partial U_{am}}{\partial z}, \tag{10}$$

$$\frac{d\langle v_y\rangle}{dt}=0,$$

$$\langle v_y\rangle=\sin\theta[(E_0^{(i)})^2+\eta^2(E_0^{(o)})^2-2\eta\cos(2\theta)E_0^{(i)}E_0^{(o)}\cos(2z\cos\theta-\alpha)]/4. \tag{11}$$

The quantity $\langle v_y\rangle$ in Eq. (11) follows from conservation of the y-component of the canonical momentum of the particle, which happens when the dependence of the amplitude $E_0^{(i,o)}$ on the coordinate is neglected. The motion of electrons in this approximation can be described qualitatively as follows. Initially electrons are near the plasma surface ($z=0$). They then approach the plane of the nearest minimum in the potential (9), i.e., $2z\cos\theta-\alpha=0$. As they oscillate in the potential well near the plane of the minimum, the particles move at velocity $\langle v_y\rangle$ toward region o, where they are entrained by the reflected wave and acquire a mean velocity

$\eta^2(E_0^{(o)})^2/4$ in the direction of the reflected pulse.[9] Taking the transverse structure of the laser pulse into account leads to a modification of the physics of electron motion: large scale potential variation shows up, and the moving electrons tend to occupy positions that minimize the potential energy.

This kind of motion can be approximately represented by a mass point sliding over a curvilinear surface. Here the kinetic energy of the electron after it escapes from regions occupied by the field will equal its initial potential energy,

$$U(z=0)=(E_0^{(i)})^2+\eta^2(E_0^{(o)})^2-\eta\cos(2\theta)E_0^{(i)}E_0^{(o)}\times\cos\alpha+U_{am}(z=0),$$

and the angle of escape will depend on the specific form of the function $U(y,z)$.

We now estimate the angle of escape for a laser beam with a triangular transverse distribution of field amplitudes,

$$E_0^{(i)}(y,z)=E_0(1-|z\sin\theta-y\cos\theta|/R),\quad |z\sin\theta-y\cos\theta|\leqslant R, \tag{12}$$

$$E_0^{(o)}(y,z)=E_0(1-|z\sin\theta+y\cos\theta|/R),\quad |z\sin\theta+y\cos\theta|\leqslant R, \tag{13}$$

where R is the radius of the laser beam in units of c/ω, i.e., $R=D\omega/2c$. Note that estimates of the time the electron moves in regions *b*, *o*, and *i* for $E_0\geqslant 0.1$, along with transverse beam sizes of the order of twenty wavelengths, yield a time much shorter than the pulse duration. Thus, Eqs. (12) and (13) do not contain the laser pulse length. Electrons initially concentrated in the neighborhood of $z\sim 0$, $y\sim 0$ escape to the vacuum, since the potential falls off toward the vacuum near that point. The equation of motion for electrons escaping to the vacuum in region *b* has the form

$$\frac{d^2z}{dt^2}=\frac{2E_0^2\sin\theta}{R}\left(1+\frac{z\sin\theta-y\cos\theta}{R}\right)+\frac{2E_0^2\eta^2\sin\theta}{R}\left(1+\frac{z\sin\theta+y\cos\theta}{R}\right), \tag{14}$$

$$\frac{d^2y}{dt^2}=-\frac{2E_0^2\cos\theta}{R}\left(1+\frac{z\sin\theta-y\cos\theta}{R}\right)+\frac{2E_0^2\eta^2\cos\theta}{R}\left(1+\frac{z\sin\theta+y\cos\theta}{R}\right). \tag{15}$$

In regions *i* and *o* we must remove terms respectively containing or not containing η in Eqs. (14) and (15). In writing Eqs. (14) and (15) we have averaged over the small scale (of the order of the wavelength). Equations (14) and (15) do not contain a force corresponding to the ambipolar field. As will be shown below, this is true for small values of the phase α of the reflected wave. Equations (14) and (15) comprise a set of linear equations with constant coefficients. Their solution in regions *i*, *b*, and *o* and the matching of solutions at the boundary present no difficulties. As a result, for the rightward escape angle $\theta_e^{(+)}$ (between the velocity of the electron and the *z* axis) of an electron from a laser spot, we obtain

$$\theta_e^{(+)}=\arctan\left(\tan\theta\times\frac{\chi(1-\eta)-\cos(2\theta)(1-\eta)/(\mu^2-\eta^2)}{\chi(1+\eta)+\cos(2\theta)(1-\eta)/(\mu^2-\eta^2)}\right), \tag{16}$$

where $\chi=\sqrt{2}E_0/R$ and $\mu^2=\{1+\eta^2+[(1+\eta^2)^2-4\eta^2\sin^2(2\theta)]^{1/2}\}/2$.

Obviously, it is impossible to determine the angle of escape for a laser beam with an arbitrary transverse profile by the method considered above. We now suggest a means of estimating the angle of escape. Since electrons move in the plane of the minimum in the effective potential $U_{\rm eff}(y,z)$, we can assume that they move along a field line of $U_{\rm eff}$ that passes through the neighborhood of $y\sim 0$, $z\sim 0$. (For a convex transverse laser profile, electrons initially near the maximum of the effective potential will escape to the vacuum.) The differential equation for the field line is

$$\frac{dy}{dz}=\frac{\partial U_{\rm eff}/\partial z}{\partial U_{\rm eff}/\partial y},\quad y(0)=0.$$

Hence, it is easy to find the slope of the field line at the boundary of region *b* and to estimate the speed by invoking energy conservation. These quantities will be the initial data for the electron's motion in regions *i* and *o*. In these regions the field lines are perpendicular to the wave vector of the incident and reflected waves, respectively, so the electron acquires an additional component of velocity in these directions, whose magnitude is also determined by energy conservation, i.e., the potential $U_{\rm eff}$ at the point of entry into region *i* or *o*. Knowing the final values of the velocity components, it is easy to find the angle of escape of an electron for an arbitrary field configuration.

We now proceed to analyze the ambipolar potential $U_{am}(z)$. The ambipolar potential, like $U_{\rm eff}$, is proportional to E_0^2, since the ambipolar field balances the ponderomotive pressure of the laser light. To find the ambipolar field and the corresponding potential, we assume that the ponderomotive force of the radiation acts on the plasma from the vacuum side, pressing the bulk of electrons into the depth of the plasma and not affecting the small part of the plasma lying in the transparent region. Neglecting the effect of this small part on the formation of the ambipolar field, in the region *pl* we have equilibrium between the ponderomotive force and the ambipolar field created by the bulk of the electrons:

$$E_{am}=-(\sin\theta-v_y)(B_x^{(i)}+B_x^{(o)}),\quad z>0. \tag{17}$$

Thus, the bulk charge density ρ corresponding to the ambipolar field is

$$\rho=\frac{1}{4\pi}\frac{\partial E_{am}}{\partial z}. \tag{18}$$

The charge density is nonzero only within the skin layer, i.e., in a very narrow region near the surface. For example, under actual conditions the scale length of the skin layer is of the order of a tenth of the wavelength of the incident radiation. Thus, at distances from the plasma boundary comparable to

the wavelength, but less than the width of the laser beam, the charge configuration (18) is equivalent to a plane with surface charge density σ given by

$$\sigma=\int_0^\infty \rho dz=\frac{1}{4\pi}E_{am}\Big|_{z=0}=-\frac{1}{4\pi}(\sin\theta-v_y)B_x\Big|_{z=0}. \tag{19}$$

The ambipolar field in vacuum, which affects the escape of particles, can be assumed to be the field due to the surface charge σ, i.e., $E_{am}=4\pi\sigma$. As a result, the ambipolar potential takes the form

$$U_{am}(y,z)\approx\frac{\partial U_{\text{eff}}}{\partial z}\Big|_{z=0}\cdot|z|=-\eta\cos(2\theta) \times\cos\theta E_0^{(i)}E_0^{(o)}\sin\alpha|_{z=0}\cdot|z|. \tag{20}$$

At distances comparable to the physical dimensions of region *b*, the ambipolar field falls off in accordance with Coulomb's law. Thus, Eq. (20) holds at small *z*, and is actually the Taylor series expansion of the ambipolar potential. At small *z* the ambipolar field balances the force owing to ponderomotive pressure up to the linear term of the Taylor series. This leads to a reduction in the dependence of the energy of an escaping electron on the laser pulse intensity for $E_0<1$.

To conclude this section we consider relativistic laser intensities, for which $E_0>1$. The averaged equation of motion then takes the form[8]

$$\frac{d\mathbf{p}}{dt}=-\frac{\nabla\langle\mathbf{E}^2\rangle}{2(1+\mathbf{p}^2+\langle\mathbf{E}^2\rangle)}, \tag{21}$$

where **p** is the electron momentum. This equation can be solved analytically in regions *i* and *o*. Solving Eq. (21) in region *o* yields the angle of escape of an electron from the region occupied by the field,

$$\theta_e^{(+)}=\theta+\arctan\frac{[2(\gamma/\gamma_b-1)(1+v_b)]^{1/2}}{\gamma-\gamma_b(1-v_b)}, \tag{22}$$

where γ_b, v_b, γ, and v are the Lorentz factor and velocity upon entering and leaving region *b*. One important consequence of Eq. (22) is that acceleration of an electron by the reflected laser pulse ($\gamma>\gamma_b$) causes it to escape in the specular direction, i.e., $\theta_e^{(+)}\to\theta$ for $\gamma\to\infty$.

Summarizing the analytic treatment of electron motion in incident and reflected laser fields, we find that for nonrelativistic fields the parameters of the trajectory are determined by the form of the effective potential (i.e., the spatial distribution of the laser fields). At relativistic intensities, entrainment of the electron by the reflected pulse induces it to move in the direction of the latter.

3. NUMERICAL SIMULATION OF ELECTRON MOTION

The analytic model for the motion of electrons above the plasma constructed in the preceding section contains many approximations, but it provides a qualitative description of the escape of electrons from the surface. This model is accurate to within a factor of order unity. A more rigorous examination of this problem requires a comprehensive solution of the electron equations of motion and Maxwell's equations in two-dimensional geometry. This complicated computational problem will be solved by the method described below. To simplify the problem we take η and α from Eq. (4), i.e., we assume that the transverse field above the plasma surface ($z<0$) is given, and we calculate the ambipolar field and electron trajectories self-consistently.

The superposition principle is used to calculate **E** and **B** in the region where the incident and reflected laser radiation intersect. For $z>0$ and outside the beam, all components of the electric and magnetic fields are assumed equal to zero. The ambipolar field is calculated at each time using the Poisson equation and is added to the fields from the laser pulse.

We take the parameters of the laser pulse and target to be close to the possible experimental values. Consider a plane laser beam of width $R=10$ μm incident on a target at an angle $\theta=30°$ (Fig. 1). We take the reflection coefficient to be 0.8 and the phase to be $\alpha=0$. This is consistent with estimates using Eqs. (4) and (5) for a laser pulse of duration $\tau_i=100$ fs and intensity 10^{18} W/cm^2. We take the laser frequency to be $\omega=1.8\times10^{15}$ s^{-1}. In the calculations we examined several values of the maximum laser intensity and various transverse and temporal profiles for the laser pulse.

The target plasma with which the laser radiation interacts has initial electron density $N_e(z)$ and mean electron temperature 0.1 keV. The $N_e(z)$ dependence is taken from estimates of the expansion dynamics of a plasma exposed to a picosecond preheating laser pulse. At $z>0$ we have $N_e(z)=aN_c$, and at $z<0$,

$$N_e(z)=aN_c\exp(-z/z_c),$$

where $N_c=10^{21}$ cm^{-3} is the critical density, $a=3$ is the ratio of the maximum density to the critical density, and $z_c=0.3$ μm is the scale length of the inhomogeneity at the critical point.

To describe the motion of plasma electrons we have used the particle-in-cell (PIC) method. Three-dimensional particle motion is described relativistically, including the Lorentz force. The difference scheme employed here is a second-order approximation in space and time. To improve the accuracy of the fields at the particles, they were calculated separately for each particle at each time step. The number of steps was set to 18000. The time for the calculations was $3\tau_i$. The number of particles in each calculation was 100000. They were all injected into the laser spot region at the initial time in accordance with the density profile and initial energy of the target plasma, and they continue to move until the pulse terminates. To represent particle motion as clearly as possible, the spatial distribution of electron density is recorded several times during the laser pulse. After the calculation is complete, the energy and the angle of escape φ (the angle between the momentum and the normal to the surface at the time the calculation ends) of each particle is retained. The distribution of electrons $n_e(\varphi)$ with respect to the angle of emission φ is determined by summing the accumulated information over all particles, and the mean electron energy is calculated.

In this model the ions are also assumed to move. The ion and electron density profiles initially coincide, and the ions subsequently move under the influence of the same fields as the electrons.

4. NUMERICAL CALCULATIONS

The calculations were carried out for peak laser powers 10^{17} and 10^{18} W/cm^2 with triangular spatial and temporal distributions of the laser fields, as well as at an intensity 10^{18} W/cm^2 with a rectangular pulse shape. The overall duration of the triangular pulses was 200 fs, and that of the rectangular pulses was 100 fs.

We begin our examination of the results by analyzing a triangular pulse with intensity of 10^{18} W/cm^2. Figure 2 shows the spatial distribution of electron density at two times, $t=40$ and 308.9 fs. The abscissa in both plots of Fig. 2 is the linear position z of an electron above the plasma surface, and ranges from -4 μm to 10 μm; the ordinate is the position y, which ranges from -12 μm to 10 μm. The density scales are in units of electrons/cm^3. The calculations show that the laser radiation is already ''pressuring'' electrons into the plasma by 40 fs. The resulting ponderomotive potential repels some electrons from the point $z\sim z_{cr}$. Those that are ''turned back'' by the laser light and ambipolar potential subsequently form two broad beams and emerge to the right.

Comparison of the distributions at times $t=40$ fs and 308.9 fs indicates that a beam of electrons that does not move along the surface normal is actually formed from the subcritical plasma. This is illustrated most clearly by Fig. 3, which shows the angular distribution $n_e(\varphi)$ of electrons at $z>0$ at time $t=308.9$ fs.

Analysis of Fig. 3 indicates the presence of a beam of electrons with an energy of roughly 15 keV emitted at an angle of $-58°$ to the normal. The results of this calculation for intensity 10^{17} W/cm^2 are shown in Fig. 4. At lower intensities the angular distribution of the emitted electrons becomes narrower (they emerge closer to the normal), and the angular asymmetry of the distribution is reduced. The angular distribution in Fig. 4 has three characteristic peaks: one central and two side peaks. The central peak in the plot of $n_e(\varphi)$ (Fig. 4) was obtained using a numerical model in which the unconstrained ends of the laser plasma move under the influence of the pulse. Note that the energy of the emerging electrons increases slowly as the laser intensity rises, and is 7.9 keV at 10^{17} W/cm^2 and 13 keV at 10^{18} W/cm^2. This effect is related to compensation of the accelerating force by the ambipolar field, as noted in Sec. 2 [Eq. (21)].

Ions also leave the target, but unlike electrons, they leave normal to the surface. The mean ion energy is of the same order of magnitude as the electron energy (for $Z=1$). The ions move because of the ambipolar potential.

We now analyze the calculations for rectangular spatial and temporal distributions of the laser intensity at 10^{18} W/cm^2. Figure 5 shows two calculated two-dimensional electron density distributions at times 50 fs and 240 fs, as done previously for a triangular distribution. The abscissa in both plots of Fig. 5 is the linear position z, which ranges between the same limits as in Fig. 2. The calculations show that the laser radiation also ''pressures'' electrons into the plasma at 40–50 fs. The resulting ponderomotive potential repels some electrons from the point $z\sim z_{cr}$. In contrast to the case of a triangular pulse shape, the part that is ''turned back'' falls into a minimum of the ponderomotive potential (9) and is retained there for the duration of the laser pulse, undergoing oscillations at the minimum of the potential

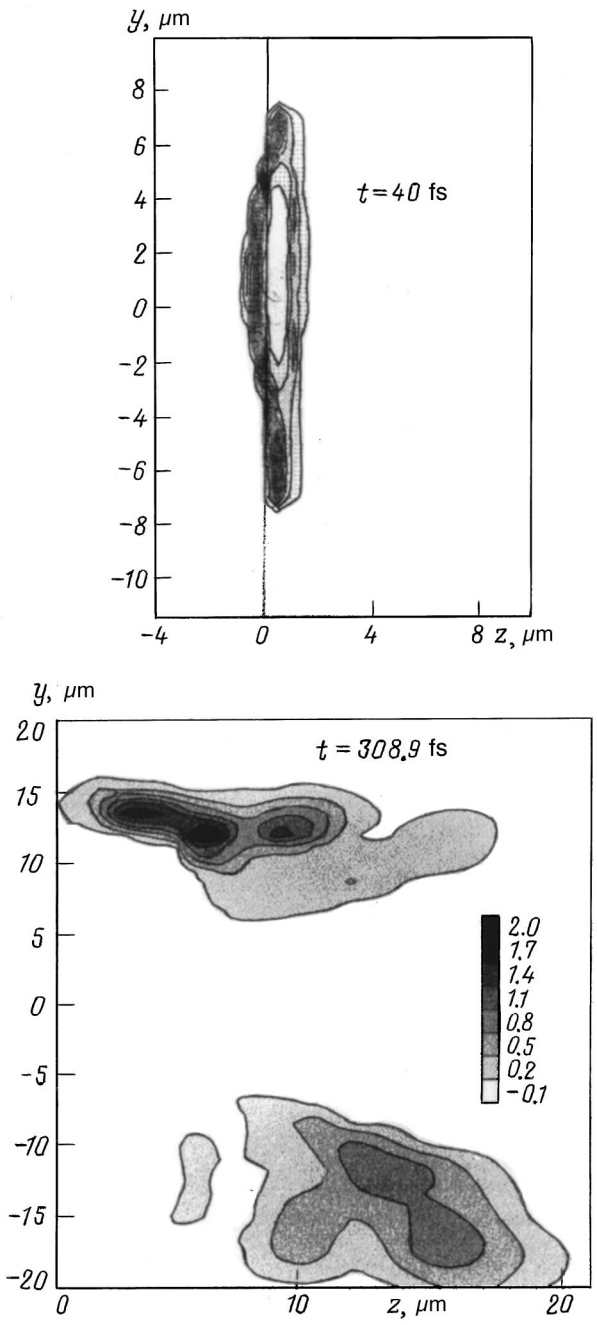


FIG. 2. Spatial distribution of electron density for a laser pulse with a triangular transverse profile at various times. The laser intensity is 10^{18} W/cm^2. Electron density is given in units of 10^{19} cm^{-3}. The abscissa is the position normal to the target surface and the ordinate is the position along the surface. The origin of the abscissa corresponds to the position of the target surface at the initial time, and that of the ordinate to the center of the laser spot.

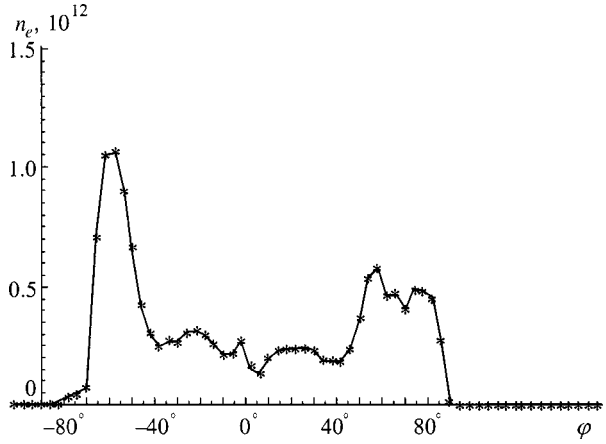


FIG. 3. Angular distribution of escaping electrons for a laser pulse with a triangular transverse profile and intensity 10^{18} W/cm^2. The abscissa is the emission angle relative to the surface normal and the ordinate is the number of electrons.

$U_{\text{eff}} \propto \cos(2z\cos\theta - \alpha)$. Figure 5 shows that the electron burst has a transverse size of 0.8–1 μm at 50 fs and is stretched out along the entire surface of the laser spot.

After the laser pulse ends (>100 fs), the resulting electron bunch escapes the target at a characteristic angle of 34° with mean energy 19 keV. Comparison of the results at equal intensities 10^{18} W/cm^2 but different transverse profiles shows that the mean electron energy is insensitive to pulse shape. The spatial structure of electron density and the angle at which electrons are ejected into the vacuum, however, differ substantially for rectangular and triangular transverse laser beam profiles. Assuming that the contribution of the ambipolar field decreases as the laser intensity increases, we have done some additional calculations of the range of emission angles and energies for triangular distributions at intensities 10^{19} W/cm^2 and 10^{20} W/cm^2, neglecting the ambipolar field. The corresponding emission angles are 48°–57° and 42°–51°, while the energies are 450–570 keV and 4600–5200 keV. These data imply that electrons emerge at angles between the specular direction and the direction of the wave

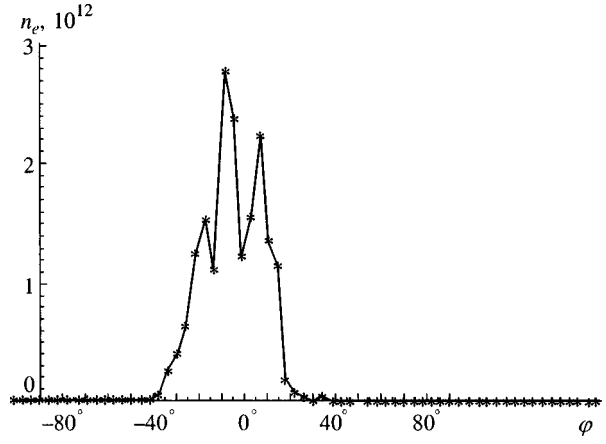
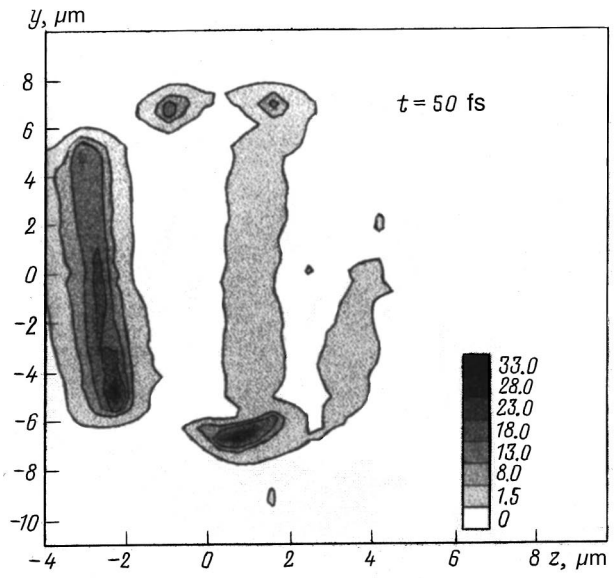


FIG. 4. Angular distribution of escaping electrons for a laser pulse with a triangular transverse profile and intensity 10^{17} W/cm^2. The abscissa is the emission angle relative to the surface normal, and the ordinate is the number of electrons.



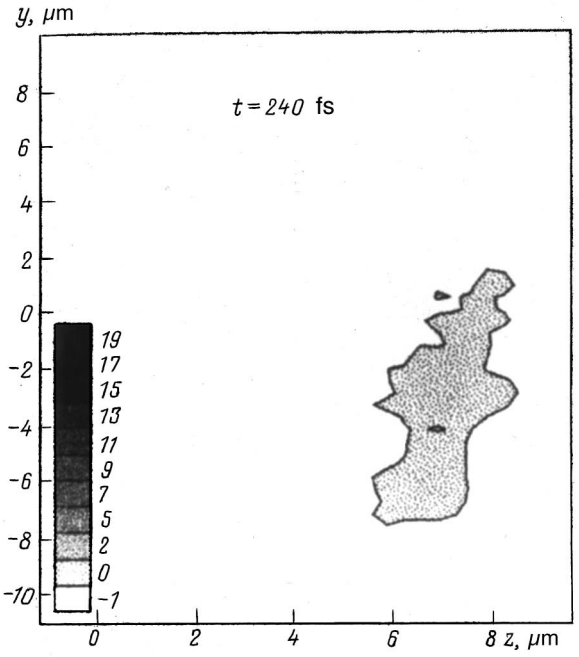


FIG. 5. Spatial distribution of electron density for a laser pulse with a rectangular transverse profile at various times. The laser intensity is 10^{18} W/cm^2. Electron density is given in units of 10^{19} cm^{-3}. The abscissa is the position normal to the target surface, and the ordinate is the position along the surface. The origin of the abscissa corresponds to the position of the target surface at the initial time, and that of the ordinate to the center of the laser spot.

electric field, while the electron energy (neglecting the ambipolar field) increases in proportion to the intensity of the laser radiation in accordance with the relativistic dynamics of an electron in the field of a monochromatic plane wave (see expression (4) on p. 152 of Ref. 9). To evaluate the influence of the initial conditions on the angle at which electrons are ejected from the laser field region, we have done some calculations with special initial conditions. It was assumed that electrons initially leave the plasma ($z=0$) either at a fixed angle of 30° to the normal or isotropically. The

initial electron energy was 2 keV and the laser intensity was 10^{18} W/cm^2. In either case, the angular distribution and energy of the emitted electrons were the same as in Fig. 3. Thus, the parameters of the emitted electrons are determined by the laser field configuration above the plasma surface.

5. COMPARISON OF NUMERICAL AND ANALYTIC RESULTS

We now compare the results of the numerical and analytic models. Above all, we note the qualitative agreement of the calculations with the theory of electron motion in the effective potential. Comparing Figs. 2 and 5, where the intensities and energies of the laser pulses are the same, while the transverse profiles are different, we see that in Fig. 2 for a triangular pulse shape, electrons are pushed out of the high field region to form two lateral bunches. In Fig. 5, where the pulse profile is rectangular, there is a single electron bunch that is uniform along the surface of the plasma at times 0–100 fs. Figures 2 and 5 imply that during initial interaction of the laser pulse with electrons (i.e., at times of 0–40 fs), the particles are divided into two groups, moving respectively into the vacuum and into the plasma. This is also explained by the structure of the effective potential (9): near the surface there is a maximum of the potential (9) at $2z\cos\theta = -\pi$. At $\theta = 30°$ and wavelength 1.04 μm, z is 0.3 μm. The initial concentration of electrons falls off exponentially with a characteristic scale length $z_c = 0.3$ μm. Electrons are repelled from the region $z \sim z_c$ as they approach the minimum in the effective potential. The subsequent minima of U_{eff} along the z axis are weakly populated by electrons, owing to the exponential drop in initial electron density, and (cf. Figs. 2 and 5) are absent up to 10^{18} cm^{-3} on the chosen density scale. If the density resolution is increased, it is possible to observe a ''striped'' structure in electron density along the z axis. After formation of the initial electron bunches over times of 0–70 fs under the influence of the strong gradients of U_{eff} over scale lengths of the order of the wavelength, these bunches begin to move under the influence of the weaker gradients of U_{eff} on scale lengths of the order of the transverse size of the laser beam. The time over which this motion takes place is of order $R/(U_{\text{eff}}/m_e)^{1/2}$, which for the numerical parameters chosen here is less than the pulse duration. An electron burst can therefore cover the entire region occupied by the field. For nonrelativistic intensities, the function $U_{\text{eff}}(y,z)$ ends up symmetric with respect to the normal to an accuracy of the order of the difference between the reflection coefficient η and unity. Thus, the angular distribution of the emitted electrons in Fig. 4 ends up almost symmetric.

When the ambipolar field is taken into account at intensities 10^{17}–10^{18} W/cm^2, the mean energy of the emitted electrons is not proportional to U_{eff} (i.e., to the intensity of the radiation), but has a weaker dependence on U_{eff} (7.9 keV at 10^{17} W/cm^2 and 13 keV at 10^{18} W/cm^2). Finally, when the intensity rises to relativistic levels 10^{18} W/cm^2, the number of electrons emerging in the direction of the reflected pulse increases (cf. Fig. 3, corresponding to 10^{18} W/cm^2, and Fig. 4, corresponding to 10^{17} W/cm^2). For ultrarelativistic intensities 10^{19} and 10^{20} W/cm^2, the emission angle approaches the specular value.

We now compare the emission angle and electron energy for a triangular laser beam with intensity 10^{17} W/cm^2 [Eqs. (16) and (17)] with the numerical results of Fig. 4. A calculation using Eqs. (16) and (17) yields $\theta_e^{(i)} \approx 17°$ and $\theta_e^{(o)} \approx 28°$; the numerical simulation yields $\theta_e^{(i)} \approx 10°$ and $\theta_e^{(o)} \approx 20°$. Finally, estimates of the emission angle from the slope of the field lines of the effective potential (as described in Sec. 2) yield $\theta_e^{(i)} \approx 6°$ and $\theta_e^{(o)} \approx 16°$. The electron energy, according to estimates of U_{eff} at $z \approx 0$, $y \approx 0$, is 12 keV and the numerical calculation yields 7.9 keV. Given the approximations entailed in the analytic estimates (the lack of an ambipolar field), we can say that the numerical and analytic results are in agreement.

This analysis implies that by varying the angle of incidence, the intensity, and the transverse profile of the laser beam, it is possible to control the motion of the electron bunch produced by the interaction of a laser pulse with a solid target.

6. CONCLUSION

An analytic model and two-dimensional PIC calculations of the collisionless expansion of a plane plasma from a solid target under the influence of incident and reflected subpicosecond laser pulses with energy fluxes exceeding 10^{16} W/cm^2 have shown that electrons in the subcritical density region are driven into the vacuum by ponderomotive and ambipolar forces in directions different from the target surface normal.

The angular distribution of the emitted electrons has the following features.

a) At intensities below 10^{17} W/cm^2, the angular distribution is essentially symmetric with respect to the normal, and has a characteristic emission angle that is determined by the transverse shape of the laser beam and by the angle of incidence.

b) When the intensity rises there is an increase in angular separation between the peaks in the angular distribution, as well as an increase in the number of electrons emerging closer to the direction of the reflected laser pulse. This results from entrainment of electrons by the reflected laser light.

At intensities above 10^{18} W/cm^2, the energy of the emitted electrons is proportional to the laser intensity, since the increase in accelerating force is not balanced by the ambipolar field. Ions leave the target normal to the surface under the influence of the ambipolar field.

In the region where the incident and reflected laser beams intersect, electron bunches develop in the regions where the ponderomotive potential has a minimum.

These results imply that a brief, intense laser pulse can be used to produce and control bunches of accelerated electrons.

This work was supported by the MNTTs (Grant No. N-107) and INTAS (Grant No. 94-934).

*)E-mail: andreev@ilph.spb.su

[1]M. Perry and G. Mourou, Science **264**, 917 (1994).
[2]W. L. Kruer and K. Estabrook, Phys. Fluids **28**, 430 (1985).
[3]G. C. Gauthier, S. Bastiani *et al.*, Proc. SPIE Vol. 23, p.124 (1997).
[4]F. Amiranoff, C. Labaune, G. Matthieussent, and C. Rousseaux, Phys. Fluids B **4**, 2589 (1992).
[5]H. Ruhl and P. Mulser, Phys. Lett. A **205**, 388 (1995).
[6]A. A. Andreev and A. N. Semahin, in *Proc. SPIE, Vol. 2097 Laser Applications* (1993), p. 327.
[7]S. C. Wilks and W. L. Kruer, Phys. Rev. Lett. **69**, 1383 (1992).
[8]B. Quesnel and P. Mora, Phys. Rev. E **58**, 3719 (1998).
[9]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon, New York (1975).

Translated by D. H. McNeill

Relativistic ponderomotive forces

D. R. Bituk

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

M. V. Fedorov*)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

(Submitted 5 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1198–1209 (October 1999)

The interaction of a relativistic classical electron with an inhomogeneous electromagnetic field is investigated. In second-order perturbation theory the motion is separated into fast and slow motions, and the relativistic Newtonian equation is averaged over the fast oscillations. The rate of change obtained for the slow component of the electron momentum is interpreted as a relativistic ponderomotive force. The result is generalized to the relativistic case of the well-known expression for the Gaponov–Miller force acting on an electron at rest. The expressions obtained for the relativistic ponderomotive forces are very complicated in the general case. They simplify in the limit of a stationary field (pulses of long duration) and a small gradient. The most typical and simplest special case of an inhomogeneous field—a stationary plane-focused beam—is investigated. The main difference between relativistic ponderomotive forces and their nonrelativistic limit is they have multiple components. In addition to the usual force directed along the gradient of the field, the relativistic case is also characterized by force components that do not have the form of the gradient of a potential and are parallel to the wave vector and the direction of the field polarization. It is shown that when a relativistic electron travels in a direction close to the direction of the wave vector of a focused laser beam, these components can greatly exceed the gradient force. A force directed along the field polarization vector arises even when the gradient of the field in this direction is zero. © *1999 American Institute of Physics.* [S1063-7761(99)00610-1]

1. INTRODUCTION

The concept of ponderomotive forces is well known for its widespread applications in plasma physics, in laser physics, and in many other branches of physics. In continuum physics (classical field theory) ponderomotive forces are interpreted as time-averaged forces acting on an element of a medium because either the field or the medium *per se* is inhomogeneous.[1] In low-density media ponderomotive forces are often viewed as time-averaged forces acting on an individual electron in a spatially inhomogeneous electromagnetic field. This definition is widely used in plasma physics and in the physics of the interaction of free electrons or beams of ionized atoms with the inhomogeneous field of a focused laser beam. In the nonrelativistic approximation a ponderomotive force acting on an individual electron is directed along the gradient of the average intensity of the field; for this reason it is often called a *gradient force*. An explicit expression for the gradient force acting on a nonrelativistic electron appears to have first been derived by Gaponov and Miller,[2] and for this reason the concept is often known by the alternative name ''Gaponov–Miller force.'' The expression for the ponderomotive force in Ref. 2 was derived by applying a field iteration procedure to the nonrelativistic Newtonian equation and then averaging the second-order (in the field) equations over time. It has been shown[3,4] that the same result can be obtained by the elementary procedure of averaging over time the nonrelativistic electron Hamiltonian in an inhomogeneous electromagnetic field $\mathbf{H}=(\mathbf{p}-e\mathbf{A}(\mathbf{r},t)/c)^2/2m$, which gives

$$\mathbf{F}_{\mathrm{GM}}=-\nabla U_{\mathrm{pond}}(\mathbf{r},t), \tag{1}$$

where

$$U_{\mathrm{pond}}(\mathbf{r},t)=\frac{e^2}{2mc^2}\,\overline{\mathbf{A}^2(\mathbf{r},t)}=\frac{e^2\varepsilon_0^2(\mathbf{r})}{4m\omega^2} \tag{2}$$

is the ponderomotive potential, $\mathbf{A}(\mathbf{r},t)=-(c/\omega)\boldsymbol{\varepsilon}_0(\mathbf{r})\times\cos(\omega t)$ is the vector potential of the field, $\boldsymbol{\varepsilon}_0(\mathbf{r})$ and ω are the amplitude and frequency of the field, and the overbar signifies averaging over the period of the field.

The validity of Eqs. (1) and (2) rests on the condition that the electron translational velocity v and the velocity of the electron oscillations in the radiation field $v_\varepsilon\equiv e\varepsilon_0/m\omega$ are small in comparison with the light velocity c: v, $v_\varepsilon\ll c$. The assumption that the electron oscillation velocity be small imposes conditions that limit the strength ε and intensity I of the field: $\varepsilon\ll\varepsilon_{\mathrm{rel}}=mc\omega/e$ and $I\ll I_{\mathrm{rel}}=m^2c^3\omega^2/8\pi e^2$. At frequencies in the optical range we have $\varepsilon_{\mathrm{rel}}\sim 10^{10}$ V/cm and $I_{\mathrm{rel}}\sim 10^{18}$ W/cm^2. The restriction on the field $v_\varepsilon\ll c$ is retained in the present study, but no such limitations are assumed in the ratio of the velocity and energy of translation motion of the electron. In other words, the Gaponov–Miller iteration and averaging procedure[2] is gen-

eralized to the case in which the translational motion of the electron is relativistic. As in Ref. 2, the calculations are carried out in second-order perturbation theory with respect to the inhomogeneous field. The main result obtained below is that in the relativistic case the gradient force acquires increments, which under certain conditions are not small, do not take the form of the gradient of some potential, and are not directed along the gradient of the field.

It is important to note that in a similarly posed problem back in 1966 Kibble[5] averaged the classical equations of motion of an electron to second order with respect to the inhomogeneous field. In Ref. 5, however, the averaged force acting on a relativistic electron was not calculated, and so the characteristics of relativistic ponderomotive forces were not revealed. These characteristics were also missed in two papers[6,7] dealing with the motion of relativistic electrons in strong fields $I \geqslant I_{\mathrm{rel}}$. As mentioned, the case of strong fields is not considered in this paper. Consequently, our objective does not include a detailed comparison with the approximations and conditions for the validity of the results in Refs. 6 and 7.

In Sec. 2 we derive general expressions for the relativistic ponderomotive force. In Sec. 3 we consider the most typical special case of electromagnetic field configuration: a plane-focused light beam.

2. AVERAGING OF THE RELATIVISTIC NEWTONIAN EQUATION IN SECOND-ORDER PERTURBATION THEORY

We consider the relativistic Newtonian equation with the Lorentz force:

$$\frac{d\mathbf{p}}{dt} = e\boldsymbol{\varepsilon}(\mathbf{r},t) + \frac{e}{c}\mathbf{v}\times\mathbf{H}(\mathbf{r},t), \tag{3}$$

where $\boldsymbol{\varepsilon}(\mathbf{r},t)$ and $\mathbf{H}(\mathbf{r},t)$ are the electric and magnetic fields,

$$\boldsymbol{\varepsilon}(\mathbf{r},t) = -\frac{1}{c}\frac{\partial \mathbf{A}(\mathbf{r},t)}{\partial t},$$

$\mathbf{H}(\mathbf{r},t) = \operatorname{curl}\mathbf{A}(\mathbf{r},t)$, $\mathbf{p}$ is the momentum of the electron, $\mathbf{v}$ is its velocity, $\mathbf{v} = c^2\mathbf{p}/E$, and E is the energy, whose rate of change is defined as the work done by the field in unit time:

$$\frac{dE}{dt} = e\mathbf{v}\cdot\boldsymbol{\varepsilon}(\mathbf{r},t). \tag{4}$$

Equation (4) can be used to write the relativistic Newtonian equation (3) in an equivalent form as an equation for the electron velocity $\mathbf{v}$:

$$\frac{d\mathbf{v}}{dt} = \frac{c^2}{E}\left\{ -\frac{e}{c}\frac{\partial \mathbf{A}(\mathbf{r},t)}{\partial t} + \frac{e}{c}\mathbf{v}\times[\nabla\times\mathbf{A}(\mathbf{r},t)] + \frac{e}{c^3}\mathbf{v}\left(\mathbf{v}\cdot\frac{\partial \mathbf{A}(\mathbf{r},t)}{\partial t}\right)\right\}. \tag{5}$$

In its most general form the vector potential of a field that satisfies Maxwell's equations is represented by the Fourier expansion

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{2}\int d\mathbf{k}\mathbf{A}_{\mathbf{k}}\exp[i(\mathbf{k}\cdot\mathbf{r} - \omega_k t)] + \text{c.c.}, \tag{6}$$

where $\mathbf{A}_{\mathbf{k}}$ are the coefficients of the expansion, which are arbitrary functions of the three-dimensional wave vector $\mathbf{k}$, and $\omega_k = c|\mathbf{k}|$ is the frequency corresponding to the wave vector $\mathbf{k}$. We shall assume below that the spectral width of the field $\mathbf{A}(\mathbf{r},t)$ is not too large, $\Delta\omega/\omega_0 \ll 1$, where $\Delta\omega$ is the width of the spectrum, and ω_0 is the center frequency of the radiation.

We note that the specification of the field in the form (6) provides a fairly general description of both the spatial and the temporal inhomogeneities, i.e., the distribution of the field intensity at the focus together with the activation and deactivation of the field during transmission of the light pulse.

The initial conditions for Eq. (6) are given in the form

$$\mathbf{r}(t)|_{t\to t_0} = \mathbf{r}_0, \quad \mathbf{v}(t)|_{t\to t_0} = \mathbf{v}_0, \tag{7}$$

where $\mathbf{r}_0$ and $\mathbf{v}_0$ are the initial values of the radius vector and velocity of the electron, and it is assumed here that the field is zero at the initial time t_0 in the region occupied by the electron.

Clearly, the relativistic equations of motion of an electron in the general form (3)–(5) cannot be solved exactly by analytical means. We therefore confine the present investigation to relatively weak fields, $v_\varepsilon \ll c$ or $I \ll I_{\mathrm{rel}}$. We also assume that as the electron moves through the inhomogeneous field, different time scales can be distinguished, and the process can be separated into fast and slow motions. Fast motions comprise oscillations of the field on the unperturbed path of the electron, which take place at the Doppler-shifted frequency of the field. Depending on the direction of the electron velocity, the light frequency along its path varies from $2\omega_0$ to $\omega_0/2\gamma^2$ (Refs. 8 and 9), where γ is the relativistic factor of the translational motion of the electron, $\gamma = [1-(v_0/c)^2]^{-1/2}$, and motion at these frequencies is regarded as fast. Superposed on these fast oscillations is a slow motion in the form of a smooth variation of the field amplitude along the path of the electron. We assume that the amplitudes of the electron coordinates and velocity induced by the field of the fast oscillations are small in comparison with their smooth, slow, but large-scale variations. It is now meaningful to pose the problem of averaging the motion of an electron over fast oscillations and finding averaged equations for the slowly varying components of the electron velocity and momentum. We note that this statement of the problem is meaningful in the given situation only when restrictions are imposed on the size of the region of inhomogeneity of the field and on the initial electron energy. Indeed, the Doppler frequency shift of the field oscillations along the electron path is also manifested as an increase in the spatial period of the field-induced rapid oscillations of the electron. We know[9] that when a relativistic electron moves along the axis of a focused laser beam, the spatial period of the oscillations of the electron coordinates and velocity in the field becomes equal to $2\gamma^2\lambda$, where $\lambda = 2\pi c/\omega_0$ is the wavelength of the radiation. Averaging over the rapid oscillations makes sense if the number of oscillations over the focal length L is large, i.e., if

$$L \gg 2\gamma^2\lambda. \tag{8}$$

Consequently, the main ideas of the foregoing discussion are the method of iterations with respect to the interaction of an electron with a field and averaging over the rapid oscillations. In the iterative procedure the time-dependent radius vector $\mathbf{r}(t)$ and the electron velocity $\mathbf{v}(t)$ are sought in the form of perturbation series:

$$\mathbf{r}(t)=\mathbf{r}^{(0)}(t)+\mathbf{r}^{(1)}(t)+\mathbf{r}^{(2)}(t)+\ldots,$$
$$\mathbf{v}(t)=\mathbf{v}^{(0)}(t)+\mathbf{v}^{(1)}(t)+\mathbf{v}^{(2)}(t)+\ldots, \tag{9}$$

where $\mathbf{r}^{(n)}(t)\sim\mathbf{v}^{(n)}(t)\sim\varepsilon_0^n$, and the zeroth-order solutions are determined by the initial conditions (7):

$$\mathbf{r}^{(0)}(t)=\mathbf{r}_0+\mathbf{v}_0 t, \quad \mathbf{v}^{(0)}(t)=\mathbf{v}_0. \tag{10}$$

It is obvious that when $\mathbf{r}=\mathbf{r}^{(0)}(t)$ and $\mathbf{v}=\mathbf{v}^{(0)}(t)$ (10) are substituted into the right-hand side of Eq. (5), the equation for $\mathbf{v}^{(1)}(t)$ acquires only rapidly oscillating terms. Representing the vector potential of the field in the form (6), we can integrate the resulting equation for $\mathbf{v}^{(1)}(t)$ and find explicit first-order solutions:

$$\mathbf{v}^{(1)}(t)=-\frac{ec}{2E_0}\int d\mathbf{k}\frac{\omega_{\mathbf{k}}\mathbf{B}_{\mathbf{k}}}{\omega_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{v}_0}\times\exp[i(\mathbf{k}\cdot\mathbf{r}^{(0)}(t)-\omega_{\mathbf{k}}t)]+\text{c.c.}, \tag{11}$$

$$\mathbf{r}^{(1)}(t)=-\frac{ec}{2E_0}\int d\mathbf{k}\frac{\omega_{\mathbf{k}}\mathbf{B}_{\mathbf{k}}}{(\omega_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{v}_0)^2}\times\exp[i(\mathbf{k}\cdot\mathbf{r}^{(0)}(t)-\omega_{\mathbf{k}}t)]+\text{c.c.}, \tag{12}$$

where the following notation has been introduced for economy of space:

$$\mathbf{B}_{\mathbf{k}}=\mathbf{A}_{\mathbf{k}}+\frac{\mathbf{v}_0\times(\mathbf{k}\times\mathbf{A}_{\mathbf{k}})}{\omega_k}-\frac{\mathbf{v}_0}{c^2}(\mathbf{v}_0\cdot\mathbf{A}_{\mathbf{k}}), \tag{13}$$

and $E_0=mc^2[1-(v_0/c)^2]^{-1/2}$ is the initial electron energy.

Clearly, the direct averaging of Eqs. (11) and (12) yields zero and has no significance. Nonzero average values of the electron velocity and momentum arise only in second-order perturbation theory. A second-order equation is conveniently deduced from Eq. (3) after expressions for the fields $\boldsymbol{\varepsilon}$ and $\mathbf{H}$ in terms of the vector potential $\mathbf{A}$ are substituted into its right-hand side, together with the quantities $\mathbf{r}=\mathbf{r}^{(1)}(t)$ and $\mathbf{v}=\mathbf{v}^{(1)}(t)$, whereupon we obtain

$$\frac{d\mathbf{p}^{(2)}}{dt}=-\frac{e}{c}\left\{(\mathbf{r}^{(1)}(t)\times\nabla)\frac{\partial\mathbf{A}(\mathbf{r},t)}{dt}+\mathbf{v}^{(1)}(t)\times\nabla\mathbf{A}(\mathbf{r},t)+\mathbf{v}_0\times[(\mathbf{r}^{(1)}(t)\times\nabla)\nabla\mathbf{A}(\mathbf{r},t)]\right\}\Bigg|_{\mathbf{r}=\mathbf{r}^{(0)}(t)}, \tag{14}$$

where the functions $\mathbf{v}^{(1)}(t)$ and $\mathbf{r}^{(1)}(t)$ are given by Eqs. (11) and (12). Substituting these expressions into Eq. (14), we average the result over rapid oscillations. Essentially this means that we are omitting terms proportional to $\exp\{\pm i(\omega_{\mathbf{k}}+\omega_{\mathbf{k}'})t\}$ in the resulting sums and integrals and are leaving only difference-frequency terms proportional to $\exp\{\pm i(\omega_{\mathbf{k}}-\omega_{\mathbf{k}'})t\}$. The rate of change of the slow component of the electron momentum induced by averaging is interpreted as the ponderomotive force $\mathbf{F}_{\text{pond}}(t)$ acting on the electron at any instant t during its motion through the region occupied by the inhomogeneous field. The final general result of our calculations has the form

$$\mathbf{F}_{\text{pond}}(t)=\overline{\frac{d\mathbf{p}^{(2)}}{dt}}=-\frac{e^2}{2E_0}\text{Im}\left(\int d\mathbf{k}d\mathbf{k}'\times\exp\{i[(\mathbf{k}-\mathbf{k}')\mathbf{r}^{(0)}(t)-(\omega_{\mathbf{k}}-\omega_{\mathbf{k}'})t]\}\times\omega_{\mathbf{k}}\left\{\frac{\mathbf{B}_{\mathbf{k}}\times(\mathbf{k}'\times\mathbf{A}^*_{\mathbf{k}'})}{\omega_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{v}_0}+\frac{\mathbf{k}'\cdot\mathbf{B}_{\mathbf{k}}}{(\omega_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{v}_0)^2}\times(\omega_{\mathbf{k}'}\mathbf{A}^*_{\mathbf{k}'}+\mathbf{v}_0\times(\mathbf{k}'\times\mathbf{A}^*_{\mathbf{k}'}))\right\}\right). \tag{15}$$

We note that, generally speaking, the definition of the average (ponderomotive) force as the time average of the rate of change of the momentum $\mathbf{p}$ is not the only one possible in the relativistic case. For example, the average 4-force can be defined as the derivative of the electron 4-momentum averaged over φ with respect to the phase $\varphi=\omega t-\mathbf{k}\cdot\mathbf{r}$ (Ref. 6). The definition (15) is not relativistically covariant, but this does not prevent its use in the laboratory coordinate frame. The laboratory frame is distinguished from all others and therefore contains the light source. As a rule, the field assumes the simplest form in this frame. For example, in the case of long-duration pulses discussed below, the laboratory frame is the only frame in which the field is stationary. The physical significance of the definition (15) is that the time integral of the force $\mathbf{F}_{\text{pond}}(t)$ is equal to the change in the electron momentum $\Delta\mathbf{p}$ during the entire interaction time. The momentum variation, in turn, governs (e.g.) the variation of the direction of electron motion after traversing the region of localization of the inhomogeneous field (relative to its initial direction of motion). As for the path of the electron as a whole averaged over the rapid oscillations, $\overline{\mathbf{r}(t)}$, in the general relativistic case it does not depend on the average force (15). Its determination requires time-averaging Eq. (5) for the velocity $\mathbf{v}$ rather than Eq. (3) for the momentum $\mathbf{p}$. The averaged Eq. (5) can then be integrated to obtain $\overline{\mathbf{r}(t)}$. In this paper, however, we are content to find the relativistic ponderomotive force given by Eq. (15).

The general equation for $\mathbf{F}_{\text{pond}}(t)$ (15) is valid for an inhomogeneous field of arbitrary spatial and temporal configuration. An important special case discussed below is the case of a stationary field. This approximation is valid if the laser pulse lasts much longer than the electron time of flight through the region of the inhomogeneous field. In this case the spectral width of the radiation $\Delta\omega$ is so small that we can set $\omega_{\mathbf{k}}=\omega_{\mathbf{k}'}=\omega_0$ in the integrand of Eq. (15), whereupon we obtain

$$\mathbf{F}^{\text{st}}_{\text{pond}}(t)=-\frac{e^2\omega_0}{2E_0}\text{Im}\left(\int d\mathbf{k}d\mathbf{k}'\exp\{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}^{(0)}(t)\}\times\left\{\frac{\mathbf{B}_{\mathbf{k}}\times(\mathbf{k}'\times\mathbf{A}^*_{\mathbf{k}'})}{\omega_0-\mathbf{k}\cdot\mathbf{v}_0}+\frac{\mathbf{k}'\cdot\mathbf{B}_{\mathbf{k}}}{(\omega_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{v}_0)^2}\times(\omega_0\mathbf{A}^*_{\mathbf{k}'}+\mathbf{v}_0\times(\mathbf{k}'\times\mathbf{A}^*_{\mathbf{k}'}))\right\}\right). \tag{16}$$

Strictly speaking, the integration over $\mathbf{k}$ and $\mathbf{k}'$ in Eq. (16) can be carried out in general form without specifying the functions $\mathbf{A}_{\mathbf{k}}$ and $\mathbf{A}_{\mathbf{k}'}$. It should be noted in this regard that the spectral expansion (6) assumes the form

$$\mathbf{A}(\mathbf{r},t)=\mathrm{Re}[\mathbf{A}_0(\mathbf{r})\exp(-i\omega_0 t)], \tag{17}$$

where $\mathbf{A}_0(\mathbf{r})$ is the complex amplitude of the field:

$$\mathbf{A}_0(\mathbf{r})=\int d\mathbf{k}\mathbf{A}_{\mathbf{k}}\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{18}$$

Finally, the power-law dependence on $\mathbf{k}$ and $\mathbf{k}'$ in the integrals of Eq. (16) can be replaced by differentiation operators according to the rules $\mathbf{k}\to i\nabla=i\partial/\partial\mathbf{r}$ and $\mathbf{k}'\to-i\nabla'=-i\partial/\partial\mathbf{r}'$. As a result, we obtain the following operator representation for the relativistic ponderomotive force:

$$\mathbf{F}^{\mathrm{st}}_{\mathrm{pond}}(t)=\frac{e^2\omega_0}{2E_0}\mathrm{Im}\left\{ i\frac{\mathbf{B}_0(\mathbf{r})\times[\nabla'\times\mathbf{A}_0^*(\mathbf{r}')]}{\omega_0-i\mathbf{v}_0\cdot\nabla}+\frac{\nabla'\cdot\mathbf{B}_0(\mathbf{r})}{(\omega_0-i\mathbf{v}_0\cdot\nabla)^2}(i\omega_0\mathbf{A}_0^*(\mathbf{r}')+\mathbf{v}_0[\nabla'\times\mathbf{A}_0^*(\mathbf{r}')])\right\}\Bigg|_{\mathbf{r}=\mathbf{r}'=\mathbf{r}^{(0)}(t)}, \tag{19}$$

where the operator fractions are defined by power series in ∇ and ∇', and the function $\mathbf{B}_0(\mathbf{r})$ is equal to

$$\mathbf{B}_0(\mathbf{r})=\mathbf{A}_0(\mathbf{r})+i\frac{\mathbf{v}_0\times[\nabla\times\mathbf{A}_0(\mathbf{r})]}{\omega_k}-\frac{\mathbf{v}_0}{c^2}(\mathbf{v}_0\cdot\mathbf{A}_0(\mathbf{r})). \tag{20}$$

Bearing in mind this definition of the operator fractions in terms of power series in ∇, we can state that in general the relativistic ponderomotive forces in a stationary inhomogeneous field of arbitrary configurations are determined by an infinite number of higher derivatives of the average intensity of the field $\overline{I(\mathbf{r})}$ at an arbitrary point $\mathbf{r}$ on the path of the electron, $\mathbf{r}=\mathbf{r}^{(0)}(t)$. This expression for the force $\mathbf{F}^{\mathrm{st}}_{\mathrm{pond}}(t)$ given by (19) can be simplified in the small-gradient approximation, when the expansion in powers of ∇ can be restricted to the first order only. This approximation and the corresponding results will be described in the next section for the case of a plane-focused stationary light beam.

3. PONDEROMOTIVE FORCES IN A WEAKLY INHOMOGENEOUS LIGHT FIELD

We now consider the frequently encountered case of a weakly inhomogeneous field. One example of this situation under certain conditions is the field of radiation focused by a lens. Criteria of weak inhomogeneity will be formulated below. Qualitatively, a field is weakly inhomogeneous if it differs only slightly from the field of a plane wave, which, in turn, is characterized by an average wave vector $\mathbf{k}_0$. For a stationary field this means that in Eqs. (17) and (18) the complex amplitude $\mathbf{A}_0(\mathbf{r})$ can be assigned a specific form by writing it as the product of a slowly varying function of the coordinates $\mathbf{A}_{00}(\mathbf{r})$ and the quantity $\exp(i\mathbf{k}_0\cdot\mathbf{r})$:

$$\mathbf{A}(\mathbf{r},t)=\mathrm{Re}[\mathbf{A}_{00}(\mathbf{r})\exp\{i(\mathbf{k}_0\cdot\mathbf{r}-\omega_0 t)\}], \tag{21}$$

where once again, by analogy with (18), the slow function $\mathbf{A}_{00}(\mathbf{r})$ can be expanded into a Fourier integral:

$$\mathbf{A}_{00}(\mathbf{r})=\int d\mathbf{k}\mathbf{A}_{0\mathbf{k}}\exp(i\mathbf{k}\mathbf{r}). \tag{22}$$

In this notation the ponderomotive force is described by the previous equation (16), but now with $\mathbf{k}$ and $\mathbf{k}'$ replaced by $\mathbf{k}_0+\mathbf{k}$ and $\mathbf{k}_0+\mathbf{k}'$. The slowness of the function $\mathbf{A}_{00}(\mathbf{r})$ in comparison with $\exp(i\mathbf{k}_0\cdot\mathbf{r})$ implies that on the average $|\mathbf{k}|\ll k_0$. By a procedure analogous to the transition from Eq. (16) to (19) we can once again integrate over the variables $\mathbf{k}$ and $\mathbf{k}'$ and write the equation for the ponderomotive force in the form

$$\mathbf{F}^{\mathrm{st}}_{\mathrm{pond}}(t)=\frac{e^2\omega_0}{2E_0}\mathrm{Im}\left\{ i\frac{\mathbf{B}_0(\mathbf{r})\times[(\nabla'+i\mathbf{k}_0)\times\mathbf{A}_{00}^*(\mathbf{r}')]}{\omega_0-i\mathbf{k}_0\cdot\mathbf{v}_0-i\mathbf{v}_0\cdot\nabla}+\frac{(\nabla'+i\mathbf{k}_0)\cdot\mathbf{B}_0(\mathbf{r})}{(\omega_0-i\mathbf{k}_0\cdot\mathbf{v}_0-i\mathbf{v}_0\cdot\nabla)^2}(i\omega_0\mathbf{A}_{00}^*(\mathbf{r}')+\mathbf{v}_0\times[(\nabla'+i\mathbf{k}_0)\times\mathbf{A}_{00}^*(\mathbf{r}')])\right\}\Bigg|_{\mathbf{r}=\mathbf{r}'=\mathbf{r}^{(0)}(t)}, \tag{23}$$

where

$$\mathbf{B}_0(\mathbf{r})=\mathbf{A}_{00}(\mathbf{r})+i\frac{\mathbf{v}_0\times[(\nabla-i\mathbf{k}_0)\times\mathbf{A}_{00}(\mathbf{r})]}{\omega_k}-\frac{\mathbf{v}_0}{c^2}(\mathbf{v}_0\cdot\mathbf{A}_{00}(\mathbf{r})). \tag{24}$$

We can interpret the slowness of the function $\mathbf{A}_{00}(\mathbf{r})$ in Eq. (22) as slowness of the gradients ∇ and ∇' in comparison with $(\omega_0-\mathbf{k}_0\cdot\mathbf{v}_0)/v_0$ in the denominators of Eq. (23):

$$|\mathbf{v}_0\cdot\nabla|\ll\omega_0-\mathbf{k}_0\cdot\mathbf{v}_0. \tag{25}$$

Of course, inequality (25) is a conditional statement and needs to be interpreted. An appropriate explicit estimate will be given below. For the time being we merely note that by virtue of condition (25) the expansion of the operator fractions in powers of ∇ and ∇' in Eq. (23) can be restricted to the lowest (linear) approximation in the gradients, which gives

$$\mathbf{F}_{\mathrm{pond}}=\mathbf{F}_{\mathrm{pol}}+\mathbf{F}_{\mathbf{k}_0}+\mathbf{F}_{\mathrm{grad}}, \tag{26}$$

where $\mathbf{F}_{\mathrm{pol}}$ and $\mathbf{F}_{\mathbf{k}_0}$ are directed along the polarization vector $\mathbf{e}\equiv\mathbf{A}_{00}/|\mathbf{A}_{00}|$, and the force component $\mathbf{F}_{\mathrm{grad}}$ is directed along the gradient of the average intensity of the field. Calculations by this scheme yield the following equations for the indicated components of the relativistic ponderomotive force:

$$\mathbf{F}_{\mathrm{pol}}=\frac{e^2}{2E_0(\omega_0-\mathbf{k}_0\cdot\mathbf{v}_0)}(\mathbf{v}_0\nabla)\times\mathrm{Re}\left\{\mathbf{A}_{00}^*(\mathbf{r})\left(\mathbf{k}_0\cdot\mathbf{A}_{00}(\mathbf{r})+\frac{\omega_0}{c^2}(\mathbf{v}_0\cdot\mathbf{A}_{00}(\mathbf{r}))\right)\right\}, \tag{27}$$

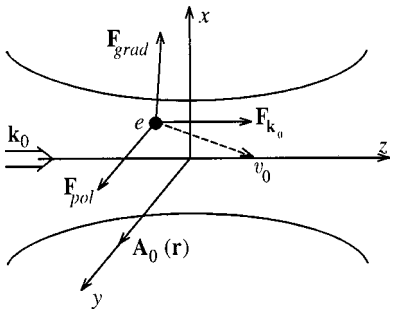


FIG. 1. Ponderomotive forces acting on a relativistic electron in an inhomogeneous laser beam.

$$\mathbf{F}_{\mathbf{k}_0}=-\frac{e^2}{2E_0(\omega_0-\mathbf{k}_0\cdot\mathbf{v}_0)^2}\mathbf{k}_0(\mathbf{v}_0\cdot\nabla)\times\mathrm{Re}[(\mathbf{v}_0\cdot\mathbf{A}_{00}(\mathbf{r}))(\mathbf{k}_0\cdot\mathbf{A}_{00}^*(\mathbf{r}))], \tag{28}$$

$$\mathbf{F}_{\mathrm{grad}}=-\frac{e^2}{2E_0}\nabla\left\{\left|\mathbf{A}_{00}(\mathbf{r})\right|^2+\frac{\mathrm{Re}[(\mathbf{v}_0\cdot\mathbf{A}_{00}(\mathbf{r}))(\mathbf{k}_0\cdot\mathbf{A}_{00}^*(\mathbf{r}))]}{\omega_0-\mathbf{k}_0\cdot\mathbf{v}_0}\right\}. \tag{29}$$

We now estimate the conditions under which the weakly inhomogeneous field approximation (25) is valid. We consider the simple scheme of a plane-focused laser beam (Fig. 1), for which the gradient of its intensity is directed along the x axis, the wave vector $\mathbf{k}_0$ is directed along the z axis, and the polarization vector $\mathbf{e}=\mathbf{A}_{00}/|\mathbf{A}_{00}|$ is directed along the y axis. The main characteristic parameter of this scheme is the angle of diffraction spreading of the beam θ_{las}:

$$\theta_{\mathrm{las}}=\frac{1}{k_0 d}=\frac{2\pi\lambda_0}{d}\sim\frac{d}{L}\sim\left(\frac{\lambda_0}{L}\right)^{1/2}\ll 1, \tag{30}$$

where d is the diameter of the focal spot, L is the focal length, and $\lambda_0=2\pi/k_0$. The parameters θ_{las} and λ_0 can be used to estimate characteristic values of the transverse and longitudinal (relative to $\mathbf{k}_0$) gradients of the field intensity:

$$\nabla_\perp=\frac{\partial}{\partial x}\sim\frac{1}{d}\sim\frac{\theta_{\mathrm{las}}}{\lambda_0},\quad \nabla_\parallel=\frac{\partial}{\partial z}\sim\frac{1}{L}\sim\frac{\theta_{\mathrm{las}}^2}{\lambda_0}. \tag{31}$$

We note that the estimates (31) characterize the maximum possible gradients. In particular, the estimates of $\nabla_\perp$ and $\nabla_\parallel$ refer to regions sufficiently far from the focal axis and from the focal plane, respectively ($\nabla_\perp=0$ on the axis, and $\nabla_\parallel=0$ in the focal plane). Let θ_e be a small angle between the electron velocity vector $\mathbf{v}_0$ and the direction of propagation of the laser beam (i.e., the vector $\mathbf{k}_0$). Also, let the electron velocity v_0 be close to the light velocity c, and let $1-v_0\approx 1/2\gamma^2$, where $\gamma\gg 1$. Under these conditions we can use the estimates (31) to write the criterion (25) in the form

$$1+\frac{\theta_e}{\theta_{\mathrm{las}}}\ll\frac{1}{2\gamma^2\theta_{\mathrm{las}}^2}+\left(\frac{\theta_e}{\theta_{\mathrm{las}}}\right)^2. \tag{32}$$

Figure 2 shows graphs of the left-hand and right-hand sides of these inequalities as functions of the ratio of the angles $\theta_e/\theta_{\mathrm{las}}$ for two different values of the parameter $\gamma^2\theta_{\mathrm{las}}^2$. If the relativistic character of the electron is not too pronounced, i.e., for $\gamma^2\theta_{\mathrm{las}}^2\ll 1$, inequality (32) is satisfied for any values of $\theta_e/\theta_{\mathrm{las}}$. But if the electron energy is so high that $\gamma^2\theta_{\mathrm{las}}^2\gg 1$, then inequality (32) is satisfied only for $\theta_e>\theta_{\mathrm{las}}$, i.e., only if the angle between the direction of the electron velocity $\mathbf{v}_0$ and the vector $\mathbf{k}_0$ is not too small. This means that for $\gamma^2\theta_{\mathrm{las}}^2\gg 1$ and for motion of the electron along or almost along the focal axis ($\theta_e<\theta_{\mathrm{las}}$) the weakly inhomogeneous field approximation is invalid, and the relativistic ponderomotive force is given by the general equations (16), (19), (23), and (24), but not by the simplified equations (27)–(29), which are linear in the gradient of the field. In the important special case $\gamma=1/\theta_e$ condition (32) is satisfied only for $\theta_e>\theta_{\mathrm{las}}$.

We now analyze the dependences and estimates of the components of the ponderomotive force (27)–(29). We consider the same plane focusing scheme as above (Fig. 1). Because the polarization vector of the field is assumed to be orthogonal both to the gradient of the field and to the wave vector $\mathbf{k}_0$, we have $\mathbf{k}_0\cdot\mathbf{A}_{00}(\mathbf{r})\equiv 0$ and, accordingly, $\mathbf{F}_{\mathbf{k}_0}\equiv 0$. If, in addition, the velocity vector $\mathbf{v}_0$ lies in the plane formed by the vectors $\mathbf{k}_0$ and $\mathbf{e}$, Eqs. (27) and (29) for $\mathbf{F}_{\mathrm{pol}}$ and $\mathbf{F}_{\mathrm{grad}}$ then assume the form

$$\mathbf{F}_{\mathrm{pol}}=\mathbf{e}_y\frac{e^2\theta_e\gamma^2}{2E_0(1+\gamma^2\theta_e^2)}\frac{\partial}{\partial z}|\mathbf{A}_{00}(\mathbf{r})|^2,$$

$$|\mathbf{F}_{\mathrm{pol}}|\sim\frac{e^2\theta_e\theta_{\mathrm{las}}^2\gamma^2|\mathbf{A}_{00}|^2}{2E_0\lambda_0(1+\gamma^2\theta_e^2)}, \tag{33}$$

$$\mathbf{F}_{\mathrm{grad}}=-\frac{e^2}{2E_0}\left(\mathbf{e}_x\frac{\partial}{\partial x}+\mathbf{e}_z\frac{\partial}{\partial z}\right)|\mathbf{A}_{00}(\mathbf{r})|^2\sim-\frac{e^2|\mathbf{A}_{00}|^2}{2E_0\lambda_0}(\mathbf{e}_x\theta_{\mathrm{las}}+\mathbf{e}_z\theta_{\mathrm{las}}^2), \tag{34}$$

where $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are unit vectors along the x, y, and z axes (see Fig. 1).

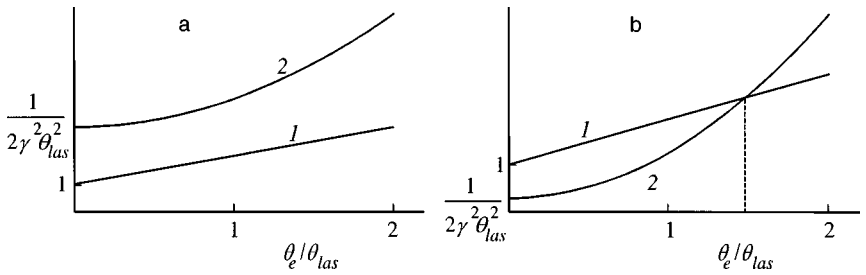


FIG. 2. Graphs of the left-hand side (*1*) and right-hand side (*2*) of inequality (32) as functions of $\theta_e/\theta_{\mathrm{las}}$ for $\gamma\sqrt{2}<2\theta_{\mathrm{las}}$ (a) and $\gamma\sqrt{2}>2\theta_{\mathrm{las}}$ (b).

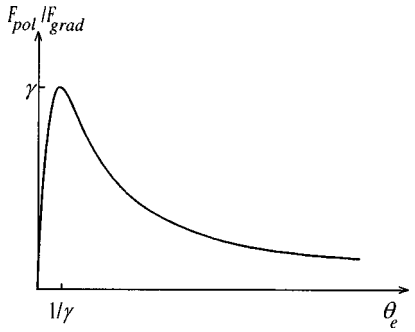


FIG. 3. Ratio of the component of the relativistic ponderomotive force directed along the field polarization vector $F_{\rm pol}$ given by (33) to the gradient component $F_{\rm grad\|}$ given by (34) as a function of the angle θ_e.

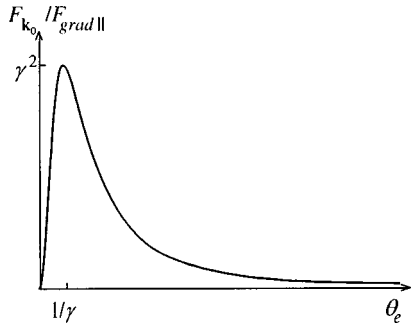


FIG. 4. Dependence of the ratio of the component of the relativistic ponderomotive force directed along the wave vector $F_{\mathbf{k}_0}$ (37) to the gradient component $F_{\rm grad\,\|}$ (34) on the angle θ_e.

We see at once that Eq. (34) is a direct, elementary generalization of the Gaponov–Miller force: $F_{\rm grad}$ (34) is obtained from Eqs. (1) and (2) by replacing the electron mass m with $E_0/c^2=\gamma m$.

It follows from Eq. (33) that the dependence of the force $F_{\rm pol}$ on θ_e has a maximum (Fig. 3) at $\theta_e=1/\gamma$, and at the maximum

$$F_{\rm pol}^{\max}=\frac{e^2\gamma}{2E_0}\frac{\partial}{\partial z}|\mathbf{A}_{00}(\mathbf{r})|^2\sim\frac{e^2\theta_{\rm las}^2\gamma|\mathbf{A}_{00}|^2}{2E_0\lambda_0}$$
$$\sim\gamma F_{\rm grad\,\|}\gg F_{\rm grad\,\|}, \tag{35}$$

i.e., the force directed along the polarization vector $\mathbf{e}_y$ is γ times larger than the longitudinal (along $\mathbf{k}_0$) component of the gradient force $F_{\rm grad\,\|}\equiv F_{\rm grad\,z}$. Owing to the weak inhomogeneity condition (32) for $\gamma=1/\theta_e$, which holds only if $\theta_e>\theta_{\rm las}$, the force $F_{\rm pol}^{\max}$ in Eq. (35) also exceeds the transverse gradient force $F_{\rm grad\,\perp}\equiv F_{\rm grad\,x}$:

$$\frac{F_{\rm pol}^{\max}}{F_{\rm grad\,x}}\sim\frac{\gamma}{\theta_{\rm las}}=\frac{\theta_e}{\theta_{\rm las}}>1. \tag{36}$$

It is important to note that, generally speaking, one should compare the gradient force and its increments in the same direction. In the present case of plane focusing, however, the gradient force in the y-direction is zero, whereas $F_{\rm pol}^{\max}\neq 0$. In this sense we can state that the nongradient ponderomotive force directed along the field polarization vector is infinitely many times greater than the gradient force in the same direction.

In view of the somewhat unexpected nature of this result, we explain its origin qualitatively. The main reason for the onset of a ponderomotive force directed along the field polarization vector in the relativistic case is that the relativistic equation for the electron velocity $\mathbf{v}$ (5) differs from its nonrelativistic limit. This difference is governed mainly by the last term on the right-hand side of Eq. (5), which is proportional to $\mathbf{v}(\mathbf{v}\cdot\boldsymbol{\varepsilon})$. When we solve this equation iteratively, the velocity acquires in the first order a correction directed along the initial electron velocity vector $\mathbf{v}^{(1)}\propto\mathbf{v}_0$. The substitution of the resulting component $\mathbf{v}^{(1)}$ into the last term of Eq. (3) causes the second-order force to acquire a term proportional to the cross product $\mathbf{v}_0\times\mathbf{H}$. When the projection of the initial electron velocity onto the direction of the wave vector $\mathbf{k}_0$ has a nonzero value, the cross product $\mathbf{v}_0\times\mathbf{H}$ has a component directed along the field polarization vector $\boldsymbol{\varepsilon}$. The calculation of the second-order force component by this elementary procedure and averaging of the result over fast oscillations yields a result that coincides with Eq. (33).

For a different field polarization with the vector $\mathbf{A}_{00}(\mathbf{r})$ situated in the common plane of the gradient and the vector $\mathbf{k}_0$ (the xz plane in Fig. 1), the product $\mathbf{k}_0\cdot\mathbf{A}_{00}(\mathbf{r})$ is not constant and in general has a nonzero value. If the electron velocity $\mathbf{v}_0$ also lies in the same xz plane as $\mathbf{A}_{00}(\mathbf{r})$, both $\mathbf{F}_{\rm pol}$ (27) and $\mathbf{F}_{\mathbf{k}_0}$ (28) are nonzero quantities. Now the force component directed along the wave vector $\mathbf{k}_0$ assumes the form

$$\mathbf{F}_{\mathbf{k}_0}=-\mathbf{e}_z\frac{2e^2\gamma^4}{E_0(1+\gamma^2\theta_e^2)^2}\left(\theta_e\frac{\partial}{\partial x}+\frac{\partial}{\partial z}\right)$$
$$\times\mathrm{Re}[\theta_e A_{00x}+(A_{00z})A_{00z}^*]. \tag{37}$$

Taking into account the estimates of the maximum possible projections of the vector potential of the focused field $A_{00x}\sim A_{00}$, $A_{00z}\sim\theta_{\rm las}A_{00}$ and the estimates of the derivatives of the potential (31), we can also estimate the magnitude of the force $\mathbf{F}_{\mathbf{k}_0}$:

$$F_{\mathbf{k}_0}\sim\frac{e^2\gamma^4A_{00}^2\theta_{\rm las}^2(\theta_{\rm las}+\theta_e)^2}{E_0\lambda_0(1+\gamma^2\theta_e^2)^2}. \tag{38}$$

A qualitative graph of $|F_{\mathbf{k}_0}(\theta_e)|$ is shown in Fig. 4. The maximum of the curve is pronounced for $\gamma\theta_{\rm las}<1$ and $\theta_e>\theta_{\rm las}$. It is attained at $\theta_e\sim 1/\gamma>\theta_{\rm las}$, and at the maximum we have

$$F_{\mathbf{k}_0}^{\max}=-\frac{e^2\gamma^2A_{00}^2\theta_{\rm las}^2}{E_0\lambda_0}\sim\gamma^2F_{\rm grad\,\|}\gg F_{\rm grad\,\|}. \tag{39}$$

This result shows that in the present case of a weakly inhomogeneous, plane-focused laser field under optimal conditions the force directed along the focal axis can be γ^2 times greater than the longitudinal gradient force. In other words, the estimate of the force acting on the electron in the direction of $\mathbf{k}_0$, according to Eq. (34), for the trivial relativistic generalization of the Gaponov–Miller force ($m\to\gamma m$) can produce a large—by a factor of γ^2—error, where $\gamma\gg 1$. This effect occurs for radiation polarized in the plane of the gradient of the field (xz plane) if, in addition, the initial electron velocity vector $\mathbf{v}_0$ lies in the same xz plane and forms an

angle $\theta_e = 1/\gamma$ with the focal axis (z axis) and if the relativity factor γ and the angle of diffraction spreading of the focused beam satisfy the conditions $\gamma \gg 1$, $\gamma\theta_{las} < 1$, and $\theta_{las} < \theta_e$.

CONCLUSION

To summarize, we have found a general expression (15) for the ponderomotive forces exerted on a relativistic electron by an inhomogeneous electromagnetic field of arbitrary configuration with an arbitrary time dependence. We have analyzed the results specifically for the approximation of a stationary, weakly inhomogeneous, plane-focused field. We have estimated the conditions underlying the validity of the small inhomogeneous field approximation (32), from which it follows, in particular, that for very high electron energies ($\gamma \gg 1/\theta_{las}$) and an electron velocity directed very close to the direction of the focal axis ($\theta_e \leqslant \theta_{las}$) the weakly inhomogeneous field approximation fails, and the general equation (15) obtained for relativistic ponderomotive forces must be used. When the conditions supporting the approximation of weak inhomogeneity of the field hold, we have determined that the relativistic ponderomotive forces have several components, i.e., besides the usual gradient forces they also contain terms directed along the field polarization vector and the focal axis. Our estimates show that the resulting ''increments'' can greatly exceed the gradient component of the relativistic ponderomotive force. We have found conditions under which the relativistic ponderomotive force differs substantially from its gradient component. These conditions essentially entail the proper choice of polarization of the field and direction of the initial electron velocity vector $\mathbf{v}_0$, as well as a relation between the directions of $\mathbf{v}_0$ and the focal axis ($\mathbf{k}_0$) so as to satisfy the condition $\gamma = 1/\theta_e$.

In addition to the expressions obtained for the relativistic ponderomotive forces *per se*, it is also important to determine the averaged electron paths in an inhomogeneous field as a whole. As mentioned above, this problem can be solved by time-averaging the relativistic equation (5) for the electron velocity $\mathbf{v}(t)$.

The results of the present study have been obtained in second order with respect to the light field, which corresponds to a relatively low intensity $I \ll I_{rel}$. Of unquestionable interest is the generalization to stronger fields. Moreover, together with the classical treatment of the problem in this paper, it would also be interesting to solve the quantum-mechanical problem of relativistic ponderomotive forces. Under certain conditions the results of the present study can most likely be reproduced in terms of the quantum solution if in the latter the initial state of the electron is specified in the form of localized wave packets.[10,11]

The authors are grateful to N. B. Narozhnyĭ for helpful critical comments.

*)E-mail: fedorov@gon.ran.gpi.ru

[1] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed. (rev. and enl., with L. P. Pitaevskii), Pergamon Press, Oxford, New York (1984).
[2] É. V. Gaponov and M. É. Miller, Zh. Éksp. Teor. Fiz. **34**, 242 (1958) [Sov. Phys. JETP **7**, 168 (1958)].
[3] M. V. Fedorov, Zh. Éksp. Teor. Fiz. **52**, 1434 (1967) [Sov. Phys. JETP **25**, 952 (1967)].
[4] L. S. Bartell, J. Appl. Phys. **38**, 1561 (1967).
[5] T. W. B. Kibble, Phys. Rev. **150**, 1060 (1966).
[6] S. P. Goreslavsky, N. B. Narozhny, O. V. Shcherbachev, and V. P. Yakovlev, Laser Phys. **3**, 418 (1993).
[7] D. Bauer, P. Mulser, and W.-H. Steeb, Phys. Rev. Lett. **75**, 4622 (1995).
[8] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford, New York (1975).
[9] M. Fedorov, *Atomic and Free Electrons in a Strong Light Field*, World Science, Singapore (1997).
[10] M. V. Fedorov, S. P. Goreslavsky, and V. S. Letokhov, Phys. Rev. E **55**, 1015 (1997).
[11] D. R. Bitouk and M. V. Fedorov, Phys. Rev. A **58**, 1195 (1998)

Translated by James S. Wood

Spontaneous interference bremsstrahlung effect in the scattering of a relativistic electron by a nucleus in the field of two light waves

S. P. Roshchupkin*) and O. B. Lysenko

Sumy State University, 244007 Sumy, Ukraine; Institute of Applied Physics, Ukrainian National Academy of Sciences, 244024 Sumy, Ukraine

(Submitted 10 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1210–1240 (October 1999)

This paper presents, for the general relativistic case, a theoretical study of nonresonance spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two elliptically polarized light waves propagating in the same direction. We show that there are two significantly different kinematic regions: the noninterference region where the main multiphoton parameters are the Bunkin–Fedorov quantum parameters $\gamma_{1,2}$, and the interference region where interference effects play an important role and where the quantum interference parameters $\alpha(\pm)$ act as multiphoton parameters. We encounter the spontaneous interference bremsstrahlung effect in two cases: in the special case of the same linear polarization of both waves, and in the general case of elliptical polarization of the waves. The effect manifests itself in the interference region and is due to stimulated, correlated emission and absorption of photons of both waves. For moderately strong fields, we find the cross sections of spontaneous bremsstrahlung by an electron scattered by a nucleus in the given kinematic regions. Finally, we show that the differential cross section in the interference region with correlated emission (absorption) of equal numbers of photons of both waves can be much greater than the corresponding cross section in any other geometry.
[S1063-7761(99)00710-6]

1. INTRODUCTION

Spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of a plane electromagnetic wave has been investigated for a long time (see, e.g. Refs. 1–6). Attention was focused primarily on the study of resonances associated with mass-shell effects in the Green's function of the intermediate electron in the plane-wave field. For the general relativistic case, the resonances were studied in Ref. 7 (see also the review in Ref. 8).

Lately there has been an upsurge of interest in elementary quantum processes that occur in a superposition of several laser fields. For instance, Karapetyan and Fedorov[9] studied, in the nonrelativistic limit (in the dipole approximation), stimulated bremsstrahlung and inverse bremsstrahlung by electrons scattered by a nucleus in the presence of two plane electromagnetic waves. This process for the general relativistic case in the field of two waves was examined in Refs. 10–12 and in the field of an arbitrary number of plane electromagnetic waves, in Ref. 13. These studies revealed a new physical effect, the interference effect[11,12] associated with correlated emission and absorption of equal numbers of photons of the two waves by an electron scattered by a nucleus in a specific plane into given angles (the interference effect was generalized to an arbitrary number of waves in Ref. 13). Note that stimulated bremsstrahlung and inverse bremsstrahlung by an electron scattered by another electron in the field of two waves were studied in Ref. 14. The spontaneous emission of a photon by an electron in the field of two waves was studied earlier (see, e.g., Refs. 15–19). The same process in the interference region in the field of two waves was studied in Ref. 20 and in the field of an arbitrary number of plane waves, in Ref. 21. Here a spontaneous combination effect was discovered: the emission by the electron of a spontaneous photon at combination frequencies of the external waves.

In the present paper we investigate nonresonant spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two plane electromagnetic waves propagating in the same direction. Note that we do not study the resonance associated with mass-shell effects in the Green's function of the intermediate electron in the plane-wave field. We do, however, investigate in detail spontaneous bremsstrahlung in the noninterference and interference regions, and for the latter we predict the emergence of a spontaneous interference bremsstrahlung effect in two cases: for equal linear polarizations of the waves and elliptical polarization of the waves. In the latter case the effect amounts to strong correlation of the exit angles of the electron and spontaneous photon, to a dependence of the emission spectrum on the energy and polar incidence angle of the initial electron, and to stimulated, correlated emission and absorption (due to wave interference) of photons of both waves. We also show that the differential cross section of spontaneous bremsstrahlung in the interference region can be much greater than the corresponding cross section in any other geometry. In this paper we use natural units, with $\hbar = c = 1$.

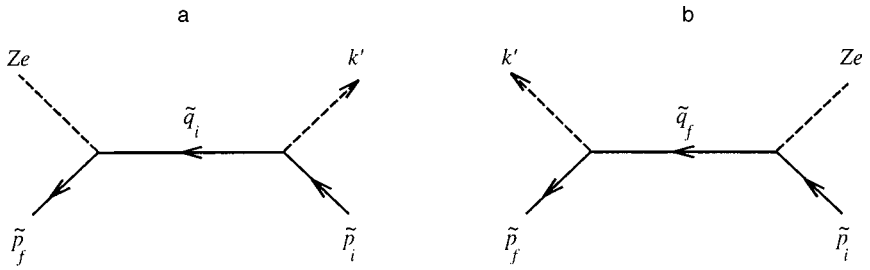


FIG. 1. Spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two light waves. The solid incoming and outgoing lines correspond to the wave functions of the initial and final electrons in the field of two waves (Volkov functions), the dashed lines correspond to the spontaneous photon k' and the ''pseudophoton'' of the nucleus, and the internal lines correspond to the Green's function of an electron in the field of two waves.

2. AMPLITUDE OF SPONTANEOUS BREMSSTRAHLUNG

We select the 4-vector potential of the external field in the form of the sum of two plane electromagnetic waves propagating parallel to the z axis:

$$A=A_1(\varphi_1)+A_2(\varphi_2), \tag{1}$$

where

$$A_j(\varphi_j)=\frac{F_j}{\omega_j}(e_{jx}\cos\varphi_j+\delta_j e_{jy}\sin\varphi_j). \tag{2}$$

Here δ_j is the ellipticity parameter ($\delta_j=0$ for linear polarization and $\delta_j^2=1$ for circular polarization of waves), $e_{jx}=(0,\mathbf{e}_{jx})$ and $e_{jy}=(0,\mathbf{e}_{jy})$ are the 4-vectors of wave polarization, F_j and ω_j are the field strength and frequency of the first ($j=1$) and second ($j=2$) waves, and the argument φ_j has the form

$$\varphi_j=\omega_j(t-z), \quad j=1,2. \tag{3}$$

The amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in the light wave (1) has the form (see Fig. 1)

$$S_{fi}=-ie^2\int d^4x_1 d^4x_2\,\overline{\psi_f}(x_2|A)\times[\tilde{\gamma}_0 A_0(x_2)G(x_1x_2|A)\hat{A}'(x_1,k')+\hat{A}'(x_2,k')G(x_1x_2|A)\tilde{\gamma}_0 A_0(x_1)]\psi_i(x_1|A), \tag{4}$$

where $\psi_i(x_1|A)$ and $\overline{\psi_f}(x_2|A)$ are the wave function of the electron in the initial and final states in the light wave (1) (the Volkov function),[22] $G(x_1,x_2|A)$ is the Green's function of the electron in the field of the plane wave (1) (see Refs. 7 and 23–26),

$$A_0(x)=\frac{Ze}{|\mathbf{x}|} \tag{5}$$

is the Coulomb potential Ze of the nucleus, the ''hatted'' expression $\hat{A}'(x,k')=\tilde{\gamma}^\mu A'_\mu$ denotes the scalar product of the Dirac matrices $\tilde{\gamma}^\mu$ ($\mu=1,2,3,4$) and the 4-vector potential of the spontaneous photon,

$$A'_\mu(x,k')=\sqrt{\frac{2\pi}{\omega'}}\varepsilon^*_\mu\exp\{ik'x\}, \tag{6}$$

with ε^*_μ and $k'=\omega' n'=\omega'(1,\mathbf{n}')$ the 4-vector of polarization and 4-momentum of the spontaneous photon. After performing simple manipulations (see Refs. 7 and 22) with the amplitude (4), we can write

$$S_{fi}=\sum_{l=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}S_{ls}, \tag{7}$$

where the partial amplitude with emission ($l>0$ and $s>0$) or absorption ($l<0$ and $s<0$) of $|l|$ photons of the first wave and $|s|$ photons of the second is

$$S_{ls}=-i\frac{8\pi^{5/2}Ze^3}{\sqrt{2\omega'\tilde{E}_i\tilde{E}_f}}\exp\{i\phi_{if}\}[\overline{u_f}H_{ls}u_i]\frac{\delta(q_0)}{\mathbf{q}^2}, \tag{8}$$

with

$$H_{ls}=\sum_{l'=-\infty}^{\infty}\sum_{s'=-\infty}^{\infty}\left[M_{l-l',s-s'}(\tilde{p}_f,\tilde{q}_i)\frac{\hat{\tilde{q}}_i+m_*}{\tilde{q}_i^2-m_*^2}K_{l',s'}\times(\tilde{q}_i,\tilde{p}_i)+K_{l',s'}(\tilde{p}_f,\tilde{q}_f)\frac{\hat{\tilde{q}}_f+m_*}{\tilde{q}_f^2-m_*^2}M_{l-l',s-s'}(\tilde{q}_f,\tilde{p}_i)\right]. \tag{9}$$

Here ϕ_{if} is the phase, which is independent of the summation indices, u_i and $\overline{u_f}$ are Dirac bispinors, $q=(q_0,\mathbf{q})$ is the 4-momentum transfer, and $\tilde{q}_i$ and $\tilde{q}_f$ are the 4-momenta of the intermediate electrons for the amplitudes of diagrams a and b (see Fig. 1), respectively:

$$q=\tilde{p}_f-\tilde{p}_i+k'+lk_1+sk_2, \tag{10}$$

$$\tilde{q}_i=\tilde{p}_i-k'-l'k_1-s'k_2, \tag{11}$$

$$\tilde{q}_f=\tilde{p}_f+k'+l'k_1+s'k_2. \tag{12}$$

In Eqs. (9)–(12), $k_1=\omega_1 n=\omega_1(1,\mathbf{n})$ and $k_2=\omega_2 n=\omega_2(1,\mathbf{n})$ are the 4-momenta of the photons of the first and second waves, $\tilde{p}_i$ and $\tilde{p}_f$ are the 4-quasimomenta of the electron before and after scattering, and m_* is the effective mass of the electron in the light wave (1):

$$\tilde{p}_j=p_j+\frac{m^2}{4(k_1p_j)}[(1+\delta_1^2)\eta_1^2+(1+\delta_2^2)\eta_2^2]k_1, \quad j=i,f, \tag{13}$$

$$m_*=m\sqrt{1+\frac{1}{2}(1+\delta_1^2)\eta_1^2+\frac{1}{2}(1+\delta_2^2)\eta_2^2}. \tag{14}$$

Here $p_j=(E_j,\mathbf{p}_j)$ is the 4-momentum of the electron before ($j=i$) and after ($j=f$) scattering, and

$$\eta_{1,2}=\frac{eF_{1,2}}{m\omega_{1,2}} \tag{15}$$

is the classical Lorentz-invariant parameter characterizing the intensity of the first and second waves, respectively.

In Eq. (9), the operators $M_{rr'}$, which determine the amplitude of the scattering of an electron (as if the intermediate electron were real; see Ref. 11) by a nucleus in the field of two waves, have the form

$$M_{rr'}(\tilde{p}_2,\tilde{p}_1)=\tilde{\gamma}_0 I_{rr'}+\frac{\omega_1 m^2}{8(k_1\tilde{p}_1)(k_1\tilde{p}_2)}B_{rr'}\hat{k}_1 + \frac{m}{4(k_1\tilde{p}_1)}\tilde{\gamma}_0\hat{k}_1\hat{D}_{rr'} + \frac{m}{4(k_1\tilde{p}_2)}\hat{D}_{rr'}\hat{k}_1\tilde{\gamma}_0, \tag{16}$$

where $r=l-l'$, $r'=s-s'$, $\tilde{p}_1=\tilde{q}_i$ and $\tilde{p}_2=\tilde{p}_f$ for the amplitude a, and $\tilde{p}_1=\tilde{p}_i$ and $\tilde{p}_2=\tilde{q}_f$ for the amplitude b (see Fig. 1). The operators $K_{l's'}$ in (9), which determine the amplitude of spontaneous emission of a photon by an electron (as if the intermediate electron were real; see Ref. 20) in the field of two waves, have the form

$$K_{l's'}(\tilde{p}_2,\tilde{p}_1)=\hat{\varepsilon}^* I_{l's'}+\frac{\omega_1 m^2}{8(k_1\tilde{p}_1)(k_1\tilde{p}_2)}B_{l's'}\hat{k}_1 + \frac{m}{4(k_1\tilde{p}_1)}\hat{\varepsilon}^*\hat{k}_1\hat{D}_{l's'} + \frac{m}{4(k_1\tilde{p}_2)}\hat{D}_{l's'}\hat{k}_1\hat{\varepsilon}^*. \tag{17}$$

where $\tilde{p}_1=\tilde{p}_i$ and $\tilde{p}_2=\tilde{q}_i$ for the amplitude a, and $\tilde{p}_1=\tilde{p}_f$ and $\tilde{p}_2=\tilde{p}_f$ for the amplitude b. In Eqs. (16) and (17), the functions $I_{rr'}$ and $B_{rr'}$ and the 4-vector $D_{rr'}$, first introduced in Ref. 11 (see also Refs. 12 and 13), have the form

$$D_{rr'}=\eta_1(e_1^* I_{r+1,r'}+e_1 I_{r-1,r'}) + \eta_2(e_2^* I_{r,r'+1}+e_2 I_{r,r'-1}), \tag{18}$$

$$e_j=e_{jx}+i\delta_j e_{jy}, \quad j=1,2,$$

$$B_{rr'}=\eta_1^2[2(1+\delta_1^2)I_{rr'}+(1-\delta_1^2)(I_{r+2,r'}+I_{r-2,r'})] + \eta_2^2[2(1+\delta_2^2)I_{rr'}+(1-\delta_2^2)(I_{r,r'+2}+I_{r,r'-2})] + 2\eta_1\eta_2[d_- I_{r-1,r'-1}+d_-^* I_{r+1,r'+1} + d_+ I_{r-1,r'+1}+d_+^* I_{r+1,r'-1}], \tag{19}$$

$$d_\pm=(1\pm\delta_1\delta_2)\cos\Delta+i(\delta_1\pm\delta_2)\sin\Delta, \tag{20}$$

$$\Delta=\angle(\mathbf{e}_{1x},\mathbf{e}_{2x}),$$

$$I_{rr'}\equiv I_{rr'}(\chi_1,\gamma_1,\beta_1;\chi_2,\gamma_2,\beta_2;\tau_-,\tau_+,\alpha_+,\alpha_-) = \frac{1}{(2\pi)^2}\int_0^{2\pi}d\phi_1\int_0^{2\pi}d\phi_2\exp\{i(\Phi-r\phi_1-r'\phi_2)\}, \tag{21}$$

$$\Phi=\gamma_1\sin(\phi_1-\chi_1)+\beta_1\sin 2\phi_1+\gamma_2\sin(\phi_2-\chi_2) + \beta_2\sin 2\phi_2+\alpha_+\sin(\phi_1+\phi_2-\tau_-) + \alpha_-\sin(\phi_1-\phi_2-\tau_+). \tag{22}$$

The arguments of the functions $I_{rr'}$ are

$$\tan\tau_\pm=\frac{\mathrm{Im}\,d_\pm}{\mathrm{Re}\,d_\pm}=\frac{\delta_1\pm\delta_2}{1\pm\delta_1\delta_2}\tan\Delta, \quad \tan\chi_j=\delta_j\frac{e_{jy}g_j}{e_{jx}g_j}, \tag{23}$$

$$g_j\equiv g_j(\tilde{p}_2,\tilde{p}_1)=\frac{\tilde{p}_2}{(k_j\tilde{p}_2)}-\frac{\tilde{p}_1}{(k_j\tilde{p}_1)},$$

$$\gamma_j\equiv\gamma_j(\tilde{p}_2,\tilde{p}_1)=\eta_j m\sqrt{(e_{jx}g_j)^2+\delta_j^2(e_{jy}g_j)^2}, \tag{24}$$

$$\beta_j\equiv\beta_j(\tilde{p}_2,\tilde{p}_1)=\frac{(1-\delta_j^2)\eta_j^2 m^2}{8}\left[\frac{1}{k_j\tilde{p}_2}-\frac{1}{k_j\tilde{p}_1}\right], \quad j=1,2, \tag{25}$$

$$\alpha_\pm\equiv\alpha_\pm(\tilde{p}_2,\tilde{p}_1)=\eta_1\eta_2\frac{|d_\mp|m^2}{2(k_1\pm k_2)}\left(\frac{1}{\tilde{p}_1}-\frac{1}{\tilde{p}_2}\right). \tag{26}$$

We note that the integer values of r and r' in Eqs. (18), (19), and (21), and of the 4-momenta $\tilde{p}_1$ and $\tilde{p}_2$ in Eqs. (18), (19), and (21)–(26), are equal to the respective values in formulas (16) and (17) (see text following these formulas). We also note that the functions (21) and (22) can be expanded in Bessel functions J_r of integer order:[11]

$$I_{rr'}=\sum_{j=-\infty}^{\infty}\sum_{j'=-\infty}^{\infty}\exp\{-i(j\tau_-+j'\tau_+)\}J_j(\alpha_+)J_{j'}(\alpha_-) \times L_{r-j-j'}(\chi_1,\gamma_1,\beta_1)L_{r'+j'-j}(\chi_2,\gamma_2,\beta_2). \tag{27}$$

Here the functions $L_{r-j-j'}$ and $L_{r'+j'-j}$ are determined by the parameters of the first and second waves and describe multiphoton processes in the field of one wave:[11,27–29]

$$L_r(\chi,\gamma,\beta)=\exp\{-ir\chi\} \times\sum_{r'=-\infty}^{\infty}\exp\{2ir'\chi\}J_{r-2r'}(\gamma)J_{r'}(\beta). \tag{28}$$

Note that the γ_j in (24) are the well-known Bunkin–Fedorov multiphoton quantum parameters.[27–29] The quantum parameters β_j in (25) play an important role for linear polarization of the waves at high electron energies ($\beta_j=0$ for circular polarization of the waves and for elliptical polarization, but in the dipole approximation for the interaction of the electron and the electrostatic fields of both waves), and the $\alpha_\pm$ in (26) are quantum interference parameters, which determine the interference effects in the scattering of an electron by a nucleus and in the spontaneous emission of a photon by the electron in the field of two waves. We also note that in Eqs. (16) and (17) the functions $I_{rr'}$ in (27) determine the multiphoton processes in the field of two waves; a detailed analysis of these functions can be found in Ref. 11. Here we merely note that if $\alpha_\pm\gtrsim 1$, correlated emission and absorption of photons of the two waves become important. But if $\alpha_\pm\ll 1$ (this condition holds in the dipole approximation, in the optical frequency range at arbitrary electron energies and moderate wave intensities, $\eta_1\eta_2\ll\omega_{1,2}E_i/m^2\nu_i$, and for linearly polarized waves of arbitrary intensity at $\Delta=\pi/2$; see (26) and (20) and Refs. 11 and 12), interference processes can be ignored ($j=j'=0$) and the functions $I_{rr'}$ in (27) become the product of functions that determine the independent emission and absorption of photons of the first and second waves:

$$I_{rr'}(\chi_1,\gamma_1,\beta_1;\chi_2,\gamma_2,\beta_2;0,0,0,0) = L_r(\chi_1,\gamma_1,\beta_1)L_{r'}(\chi_2,\gamma_2,\beta_2). \quad (29)$$

The expressions (7)–(9) for the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus are valid for arbitrary intensities and frequencies of both waves and for electron velocities $v_{i,f} \gg Z/137$. It can easily be shown that when one wave is switched off (say, when $F_2=0$), the expressions (7)–(9) determine the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of one wave,[7] and when both waves are switched off ($F_1=F_2=0$), the expressions (7)–(9) determine the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field.[22]

We also note, then, that for equal frequencies of the waves ($\omega_1=\omega_2$) and equal polarizations ($\delta_1=\delta_2$), the expressions (7)–(9) become the expression for the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of one wave, whose strength F and polarization vectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are linked to the initial parameters of the waves through the relations

$$F=\sqrt{F_1^2+F_2^2+2F_1F_2\cos\Delta}, \quad (30)$$

$$\mathbf{e}_x=\frac{F_1\mathbf{e}_{1x}+F_2\mathbf{e}_{2x}}{F}, \quad \mathbf{e}_y=\frac{F_1\mathbf{e}_{1y}+F_2\mathbf{e}_{2y}}{F}. \quad (31)$$

We therefore assume that the frequencies of the waves differ by $\Delta\omega=\omega_2-\omega_1$. Then the expressions (10)–(12) for the 4-momenta become

$$q=\tilde{p}_f-\tilde{p}_i+k'+[l+s(1+\Delta\omega/\omega_1)]k_1, \quad (32)$$

$$\tilde{q}_i=\tilde{p}_i-k'-[l'+s'(1+\Delta\omega/\omega_1)]k_1, \quad (33)$$

$$\tilde{q}_f=\tilde{p}_f+k'+[l'+s'(1+\Delta\omega/\omega_1)]k_1. \quad (34)$$

These relations (32)–(34) show that if

$$\frac{|\Delta\omega|}{\omega_1}\ll 1, \quad (35)$$

this term can be ignored.

We can then introduce new photon numbers $r=l+s$ and $r'=l'+s'$, and sum the amplitude in (7)–(9) first over all values of s and then over all values of s'. In doing so we use the easily verifiable relations

$$\sum_{s=-\infty}^{\infty} I_{r-r'+s'-s,s-s'}=\exp\{-i\alpha_-\sin\tau_+\}L_{r-r'}(\chi,\gamma,\beta),$$

$$\sum_{s=-\infty}^{\infty} I_{r-s',s'}=\exp\{-i\alpha'_-\sin\tau_+\}L_{r'}(\chi',\gamma',\beta'), \quad (36)$$

where

$$\tan\chi=\frac{\gamma_1\sin\chi_1+\gamma_2\sin\chi_2}{\gamma_1\cos\chi_1+\gamma_2\cos\chi_2},$$

$$\gamma=\sqrt{\gamma_1^2+\gamma_2^2+2\gamma_1\gamma_2\cos(\chi_1-\chi_2)}, \quad (37)$$

$$\beta=\beta_1+\beta_2+\alpha_+=\frac{(1-\delta_1^2)m^2}{8}\left[\frac{1}{k_1\tilde{p}_2}-\frac{1}{k_1\tilde{p}_1}\right](\eta_1^2+\eta_2^2+2\eta_1\eta_2\cos\Delta). \quad (38)$$

Note that in (36) the primed parameters differ from the corresponding unprimed parameters by the values of the 4-momenta $\tilde{p}_1$ and $\tilde{p}_2$ [see text following formulas (16) and (17)]. Allowing for all this, we find

$$\sum_{s=-\infty}^{\infty} M_{r'-r'+s'-s,s-s'}\to\exp\{-i\alpha_-\sin\tau_+\}M_{r-r'},$$

$$\sum_{s'=-\infty}^{\infty} K_{r'-s',s'}\to\exp\{-i\alpha'_-\sin\tau_+\}K_{r'}, \quad (39)$$

and

$$\sum_{s=-\infty}^{\infty} H_{r-s,s}\to\exp\{-i\alpha_-(\tilde{p}_f,\tilde{p}_i)\sin\tau_+\}H_r, \quad (40)$$

where $M_{r-r'}$ and $K_{r'}$ determine the amplitude of electron scattering by the nucleus and the amplitude of spontaneous emission of a photon by an electron in the field of one wave, and H_r determines, with allowance for (8), the amplitude of spontaneous bremsstrahlung by the electron scattered by the nucleus in the field of one wave.[7,8] Note that the phase in (40) is large ($|\alpha_-|\gg 1$), but in calculating the probability it is balanced by being multiplied by the complex-conjugate value. Thus, if condition (35) holds, we have the thoroughly studied process of spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of a plane wave.[7] In what follows, we therefore assume that the opposite of (35) holds:

$$\frac{|\Delta\omega|}{\omega_1}\gtrsim 1. \quad (41)$$

Moreover, we assume that the wave frequencies obey the conditions

$$\omega_1>\omega_2, \quad \omega_{1,2}\ll\begin{cases} m & \text{if } E_i\gtrsim m, \\ \frac{1}{2}mv_i^2 & \text{if } v_i\ll 1. \end{cases} \quad (42)$$

Since the Bunkin–Fedorov parameters γ_j [Eq. (24)] strongly depend on the kinetics of electron scattering and of emission of a spontaneous photon, following Refs. 11–13, 20, and 21, we can specify two kinematic regions: the noninterference region where the main multiphoton parameters are the Bunkin–Fedorov quantum parameters, and the interference region where these parameters do not manifest themselves (by virtue of special kinematics, all eight parameters γ_j vanish), and where multiphoton processes are determined primarily by the quantum interference parameters $\alpha_\pm$ [Eq. (26)], i.e., stimulated emission and absorption are governed by the interaction between the electron and the interference field of both waves.

3. AMPLITUDE OF SPONTANEOUS BREMSSTRAHLUNG IN THE INTERFERENCE REGION

In the interference region, the following relationships must hold in the general case of arbitrary wave intensities:

$$e_{jx}g_j(\tilde{p}_f,\tilde{q}_i)=e_{jy}g_j(\tilde{p}_f,\tilde{q}_i)=0,$$
$$e_{jx}g_j(\tilde{q}_i,\tilde{p}_i)=e_{jy}g_j(\tilde{q}_i,\tilde{p}_i)=0,\quad j=1,2, \tag{43a}$$

for amplitude a (Fig. 1), and

$$e_{jx}g_j(\tilde{q}_f,\tilde{p}_i)=e_{jy}g_j(\tilde{q}_f,\tilde{p}_i)=0,$$
$$e_{jx}g_j(\tilde{p}_f,\tilde{q}_f)=e_{jy}g_j(\tilde{p}_f,\tilde{q}_f)=0,\quad j=1,2, \tag{43b}$$

for amplitude b. Here the 4-vectors g_j are given by (23). By virtue of (43a) and (43b), the Bunkin–Fedorov quantum parameters (24) and the phases χ_j [Eq. (23)] must vanish:

$$\gamma_j(\tilde{p}_f,\tilde{q}_i)=\gamma_j(\tilde{q}_i,\tilde{p}_i)=0,$$
$$\gamma_j(\tilde{q}_f,\tilde{p}_i)=\gamma_j(\tilde{p}_f,\tilde{q}_f)=0,\quad j=1,2, \tag{44}$$
$$\chi_j(\tilde{p}_f,\tilde{q}_i)=\chi_j(\tilde{q}_i,\tilde{p}_i)=0,$$
$$\chi_j(\tilde{q}_f,\tilde{p}_i)=\chi_j(\tilde{p}_f,\tilde{q}_f)=0,\ j=1,2. \tag{45}$$

For arbitrary wave intensities, the conditions (43) hold in two cases:

(1) in the special case of equal linear polarizations of both waves ($\Delta=0$ and $\mathbf{e}_{1x}=\mathbf{e}_{2x}=\mathbf{e}_x$) in the scattering of the electron and the ejection of a photon in the plane perpendicular to the polarization vectors of both waves: (the yz plane: $\mathbf{e}_x\cdot\mathbf{k}'=\mathbf{e}_x\cdot\mathbf{p}_i=\mathbf{e}_x\cdot\mathbf{p}_f=0$; see Sec. 5);

(2) in the general case of elliptical polarization of both waves, for which the corresponding vectors $\mathbf{g}_j$ [see Eqs. (43)] are either parallel or antiparallel to the direction of propagation of the two waves (vector $\mathbf{n}$), i.e., perpendicular to the polarization plane $(\mathbf{e}_{jx},\mathbf{e}_{jy})$ (see Sec. 6).

Thus, in the interference region given by (43), the parameters γ_j vanish and the functions $I_{rr'}$ in (27), which determine the amplitude of spontaneous bremsstrahlung [Eqs. (7)–(9)], become the functions $J_{r_+r_-}$ (see Ref. 11):

$$\begin{aligned}&I_{rr'}(0,0,\beta_1;0,0,\beta_2;\tau_+,\tau_-,\alpha_+,\alpha_-)\\&\quad\to J_{r_+r_-}(\beta_1,\beta_2;\chi_+,\chi_-,\alpha_+,\alpha_-)\\&\quad=\exp\{-i(r_+\chi_-+r_-\chi_+)\}\sum_{j=-\infty}^{\infty}\sum_{j'=-\infty}^{\infty}\exp\{i(j\chi_+\\&\qquad+j'\chi_-)\}J_{r_+-j-j'}(\alpha_+)J_{r_--j+j'}(\alpha_-)J_j(\beta_1)J_{j'}(\beta_2),\end{aligned} \tag{46}$$

where we have introduced the notation

$$\chi_+=\tau_-+\tau_+,\quad \chi_-=\tau_--\tau_+,$$
$$r_+=\frac{1}{2}(r+r'),\quad r_-=\frac{1}{2}(r-r'). \tag{47}$$

We see that in the interference region the numbers of photons of both waves emitted and absorbed by the electron are correlated in such a way that half the sum and half the difference of these numbers (r and r') are integers (r_+ and r_-).

We now rewrite the expressions (7)–(12) and (16)–(19) for the amplitude of spontaneous bremsstrahlung in the new notation (47). Here the sum and difference of the integer indices l',s' and l,s can be either even,

$$l'+s'=2s_+,\quad l+s=2l_+,$$
$$l'-s'=2s_-,\quad l-s=2l_-, \tag{48}$$

or odd,

$$l'+s'=2s_+-1,\quad l+s=2l_+-1,$$
$$l'-s'=2s_--1,\quad l-s=2l_--1. \tag{49}$$

Assuming that they are even [Eqs. (48)], we obtain correlated emission (absorption) of equal numbers of photons of both waves, and the 4-momenta (10)–(12) transform as

$$\begin{aligned}&q\to q_{(0)}=\tilde{p}_f-\tilde{p}_i+k'+l_+(k_1+k_2)+l_-(k_1-k_2),\\&\tilde{q}_i\to\tilde{q}_{i(0)}=\tilde{p}_i-k'-s_+(k_1+k_2)-s_-(k_1-k_2),\\&\tilde{q}_f\to\tilde{q}_{f(0)}=\tilde{p}_f+k'+s_+(k_1+k_2)+s_-(k_1-k_2).\end{aligned} \tag{50}$$

Here $I_{l's'}\to J_{s_+s_-}$, $I_{l'\pm2,s'}\to J_{s_+\pm1,s_-\pm1}$, $I_{l',s'\pm2}\to J_{s_+\pm1,s_-\mp1}$, $I_{l'\pm1,s'\pm1}\to J_{s_+\pm1,s_-}$, $I_{l'\pm1,s'\mp1}\to J_{s_+,s_-\pm1}$, and $D_{l's'}\to0$ (the 4-vector $D_{l's'}$ is finite only if (49) holds). In the ''odd'' case (conditions (49) hold), instead of (50) we have

$$\begin{aligned}&q\to q_{(1)}=\tilde{p}_f-\tilde{p}_i+k'+\left(l_+-\frac{1}{2}\right)(k_1+k_2)\\&\qquad+\left(l_--\frac{1}{2}\right)(k_1-k_2),\\&\tilde{q}_i\to\tilde{q}_{i(1)}=\tilde{p}_i-k'-\left(s_+-\frac{1}{2}\right)(k_1+k_2)\\&\qquad-\left(s_--\frac{1}{2}\right)(k_1-k_2),\\&\tilde{q}_f\to\tilde{q}_{f(1)}=\tilde{p}_f+k'+\left(s_+-\frac{1}{2}\right)(k_1+k_2)\\&\qquad+\left(s_--\frac{1}{2}\right)(k_1-k_2).\end{aligned} \tag{51}$$

Here $I_{l'+1,s'}\to J_{s_+s_-}$, $I_{l',s'+1}\to J_{s_+,s_--1}$, $I_{l'-1,s'}\to J_{s_+-1,s_--1}$, $I_{l',s'-1}\to J_{s_+-1,s_-}$, $I_{l's'}\to0$, and $B_{l's'}\to0$. Equations (50) and (51) suggest that in the interference region the electron, during electron deceleration at the nucleus and spontaneous emission of a photon, emits (absorbs), in a correlated manner, equal numbers of photons of the first and second waves ($s=\pm l$ and $s'=\pm l'$) or numbers that differ by 1 [$s=\pm(l+1)$ and $s'=\pm(l'+1)$]; processes in which the numbers of emitted (absorbed) photons of both differ by more than one are suppressed. Formally, this looks like the electron emits (absorbs) in a stimulated manner an integer or half-integer number of photons at the combination frequencies $\omega_1\pm\omega_2$ (although actually there are no such photons). The final formula for the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in the interference region is

$$S_{fi}=\sum_{l_+=-\infty}^{\infty}\sum_{l_-=-\infty}^{\infty}[S^{(0)}_{l_+l_-}+S^{(1)}_{l_+l_-}], \tag{52}$$

where $S^{(0)}_{l_+l_-}$ and $S^{(1)}_{l_+l_-}$ are the partial amplitudes with correlated emission (absorption) of equal numbers of photons of both waves and of numbers of such photons that differ by 1, and are specified by Eq. (8), in which q must be replaced by $q_{(0)}$ and H_{ls} by $H^{(0)}_{l_+l_-}$, and q by $q_{(1)}$ and H_{ls} by $H^{(1)}_{l_+l_-}$, respectively. Here

$$H^{(0)}_{l_+l_-}=\sum_{s_+=-\infty}^{\infty}\sum_{s_-=-\infty}^{\infty}[T^{(0,0)}_{l_+-s_+,l_--s_-}+T^{(1,1)}_{l_+-s_++1,l_--s_-+1}],$$

$$H^{(1)}_{l_+l_-}=\sum_{s_+=-\infty}^{\infty}\sum_{s_-=-\infty}^{\infty}[T^{(1,0)}_{l_+-s_+,l_--s_-}+T^{(0,1)}_{l_+-s_+,l_--s_-}], \tag{53}$$

where

$$T^{(a,b)}_{l_+-s_+,l_--s_-}=M^{(a)}_{l_+-s_+,l_--s_-}(\tilde{p}_f,\tilde{q}_{i(b)})\frac{\hat{\tilde{q}}_{i(b)}+m_*}{\tilde{q}^2_{i(b)}-m^2_*}\times K^{(b)}_{s_+s_-}(\tilde{q}_{i(b)},\tilde{p}_i)+K^{(b)}_{s_+s_-}(\tilde{p}_f,\tilde{q}_{f(b)})\frac{\hat{\tilde{q}}_{f(b)}+m_*}{\tilde{q}^2_{f(b)}-m^2_*}\times M^{(a)}_{l_+-s_+,l_--s_-}(\tilde{q}_{f(b)},\tilde{p}_i), \tag{54}$$

$a,b=0,1$, and

$$K^{(0)}_{s_+s_-}(\tilde{p}_2,\tilde{p}_1)=\hat{\varepsilon}^*J_{s_+s_-}+\frac{\omega_1 m^2}{8(k_1\tilde{p}_1)(k_1\tilde{p}_2)}B^{(0)}_{s_+s_-}\hat{k}_1, \tag{55}$$

$$K^{(1)}_{s_+s_-}(\tilde{p}_2,\tilde{p}_1)=\frac{m}{4(k_1\tilde{p}_1)}\hat{\varepsilon}^*\hat{k}_1\hat{D}^{(1)}_{s_+s_-}+\frac{m}{4(k_1\tilde{p}_2)}\hat{D}^{(1)}_{s_+s_-}\hat{k}_1\hat{\varepsilon}^*, \tag{56}$$

$$M^{(0)}_{l_+-s_+,l_--s_-}(\tilde{p}_2,\tilde{p}_1)=\tilde{\gamma}_0J_{l_+-s_+,l_--s_-}+\frac{\omega_1 m^2}{8(k_1\tilde{p}_1)(k_1\tilde{p}_2)}\times B^{(0)}_{l_+-s_+,l_--s_-}\hat{k}_1, \tag{57}$$

$$M^{(1)}_{l_+-s_+,l_--s_-}(\tilde{p}_2,\tilde{p}_1)=\frac{m}{4(k_1\tilde{p}_1)}\tilde{\gamma}_0\hat{k}_1\hat{D}^{(1)}_{l_+-s_+,l_--s_-}+\frac{m}{4(k_1\tilde{p}_2)}\hat{D}^{(1)}_{l_+-s_+,l_--s_-}\hat{k}_1\tilde{\gamma}_0. \tag{58}$$

In Eqs. (55)–(58), the functions $B^{(0)}_{rr'}$ and the 4-vector $D^{(1)}_{rr'}$ are of the form

$$B^{(0)}_{rr'}=\eta_1^2[2(1+\delta_1^2)J_{rr'}+(1-\delta_1^2)(J_{r+1,r'+1}+J_{r-1,r'-1})]+\eta_2^2[2(1+\delta_2^2)J_{rr'}+(1-\delta_2^2)\times(J_{r+1,r'-1}+J_{r-1,r'+1})]+2\eta_1\eta_2[d_-J_{r-1,r'}+d^*_-J_{r+1,r'}+d_+J_{r,r'-1}+d^*_+J_{r,r'+1}], \tag{59}$$

$$D^{(1)}_{rr'}=\eta_1(e_1^*J_{rr'}+e_1J_{r-1,r'-1})+\eta_2(e_2^*J_{r,r'-1}+e_2J_{r-1,r'}). \tag{60}$$

Note that the operator determining the correlated emission and absorption of photons of both waves whose numbers differ by 1 [second line of (53)] is proportional to the intensities of the external fields, $H^{(1)}_{l_+l_-}\propto\eta_{1,2}$, and disappears as $\eta_{1,2}\to 0$, which sets it apart from the operator $H^{(0)}_{l_+l_-}$, which in the limit of weak fields, $\eta_{1,2}\to 0$, is given by an expression that determines the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in vanishing external fields.

4. SPONTANEOUS BREMSSTRAHLUNG IN THE NONINTERFERENCE REGION

We now turn to spontaneous bremsstrahlung of an electron scattered by a nucleus in the noninterference region, where the conditions (43) fail, i.e., in the kinematic region where the Bunkin–Fedorov quantum parameters γ_j in (24) are not small and are the main multiphoton parameters. Note that this region is broad. Only fixed-angle scattering of an electron and emission of a spontaneous electron in the plane perpendicular to the polarization vector of both waves (for equal linear polarizations of the waves see Sec. 5) and in the plane of the initial electron momentum and the wave vector (see Sec. 6) are ruled out in this region.

We begin with relativistic electron energies. Then the quantum parameters (24)–(26) have the following orders of magnitude in the noninterference region:

$$\gamma_{1,2}\sim\frac{\eta_{1,2}m}{\omega_{1,2}},\quad \beta_{1,2}\sim\gamma_{1,2}\xi_{1,2},\quad \alpha_\pm\sim\gamma_1\xi_2\sim\gamma_2\xi_1, \tag{61}$$

where

$$\xi_{1,2}=\frac{\eta_{1,2}m}{|\mathbf{p}_i|} \tag{62}$$

is the classical parameter that determines the constants of the motion in the noninterference region.[8,11–13] In moderately strong fields we have $\xi_{1,2}\ll 1$, which is equivalent to the following conditions on the intensities of the fields as functions of the electron energy:

$$\eta_{1,2}\ll\begin{cases}1 & \text{if}\quad E_i\sim m,\\ E_i/m & \text{if}\quad E_i\gg m.\end{cases} \tag{63}$$

Under these conditions, the values of the quantum parameters are such that $\beta_{1,2}\ll\gamma_{1,2}$ and $\alpha_\pm\ll\gamma_{1,2}$, and multiphoton processes are primarily determined by the corresponding Bunkin–Fedorov parameters ($l'\lesssim\gamma_1$ and $s'\lesssim\gamma_2$). Noting that $l'\omega_1/E_{i,f}\lesssim\xi_1\ll 1$ and $s'\omega_2/E_{i,f}\lesssim\xi_2\ll 1$, we

find that in moderately strong fields [the conditions (63)] the expression for the amplitude of spontaneous bremsstrahlung [Eqs. (8)–(17)] becomes much simpler. For instance, the expressions for the 4-momenta (10)–(12) and amplitudes (16) and (17) assume the form

$$q=p_f-p_i+k', \quad q_i=p_i-k', \quad q_f=p_f+k', \tag{64}$$

$$M_{l-l',s-s'}=\tilde{\gamma}_0 I_{l-l',s-s'}, \quad K_{l's'}=\hat{\varepsilon}^* I_{l's'}. \tag{65}$$

Equation (64) shows that in moderately strong fields, the resonances associated with mass-shell effects in the Green's function of the intermediate electron in the field of the waves do not emerge ($q_i^2\neq m^2$ and $q_f^2\neq m^2$), i.e., the range of field strengths given by (63) also determines the nonresonance region.

Since the arguments of the functions $I_{l's'}$ and $I_{l-l',s-s'}$ in (65) are independent of the sum of the summation indices, if we allow for (65), the amplitude (9) can easily be summed over all integers l' and s'. The final expression for the amplitude given by (8) and (9) is

$$S_{ls}=I_{ls}\exp\{i\phi_{if}\}S_*, \tag{66}$$

where S_* is the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field,[22] and the functions I_{ls} are determined by the expressions (27) and (28) in whose arguments [Eqs. (23)–(26)] we must put $\tilde{p}_1=p_i$ and $\tilde{p}_2=p_f$, i.e.,

$$\gamma_{1,2}=\gamma_{1,2}(p_f,p_i), \quad \beta_{1,2}=\beta_{1,2}(p_f,p_i), \quad \alpha_\pm=\alpha_\pm(p_f,p_i). \tag{67}$$

Allowing for the expression (66) for the amplitude, we can easily derive a formula for the differential cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two waves:

$$d\sigma_{ls}=|I_{ls}|^2 d\sigma_*, \tag{68}$$

where $d\sigma_*$ is the differential cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus with the emission of a photon of given frequency and direction and with ejection of the final electron in a given direction (in a vanishing external field).[22] Note that the derived expression is valid for ultrarelativistic electrons moving in the initial or final states in a narrow cone with the external-field photons at angles that are not too small ($\theta_{i,f}^2\gg\eta_{1,2}m/E_{i,f}$), but is not valid for ultrarelativistic electrons moving in a narrow cone with the spontaneous photon (the appropriate expressions for a single wave were derived by Borisov and Zhukovskiĭ[2]).

Equation (68) makes it clear that in moderately strong fields [the conditions (63)], the spontaneous bremsstrahlung cross section becomes the product of probabilities of emission (absorption) of l photons of the first wave and s photons of the second wave by the cross section of spontaneous bremsstrahlung for an electron scattered by a nucleus in a vanishing external field. Thus, the emission of a spontaneous photon and the emission (absorption) of photons of both waves in the electron deceleration at the nucleus proceed independently. Here the emission and absorption of photons of the first and second wave are generally correlated via the quantum interference parameters $\alpha_\pm$.

Note that for $\gamma_{1,2}\ll\xi_{1,2}^{-1}\gg 1$ and $\gamma_{1,2}\ll\xi_{2,1}^{-1}\gg 1$, which for the intensities of the fields considered as functions of the electron energy is equivalent to

$$\eta_{1,2}^2\ll\begin{cases}\omega_{1,2}/m & \text{if} \quad E_i\sim m,\\ \omega_{1,2}E_i/m^2 & \text{if} \quad E_i\gg m,\end{cases}$$
$$\eta_1\eta_2\ll\begin{cases}\omega_{1,2}/m & \text{if} \quad E_i\sim m,\\ \omega_{1,2}E_i/m^2 & \text{if} \quad E_i\gg m,\end{cases} \tag{69}$$

the quantum parameters are small: $\beta_{1,2}\ll 1$ and $\alpha_\pm\ll 1$ [see (61)]. The functions I_{ls} in (27), (28), and (68) therefore become products of independent Bessel functions [see (29)], and the differential cross section assumes the form

$$d\sigma_{ls}=J_l^2(\gamma_1)J_s^2(\gamma_2)d\sigma_*. \tag{70}$$

Thus, at the field intensities given by (69), the emission (absorption) of photons of the first and second waves and the emission of a spontaneous photon as a result of deceleration of an electron by a nucleus proceed independently. Note that the conditions (69) are much more stringent than the conditions (63). For example, in moderately strong fields [see (63)], the energy of the final electron is independent of the number of absorbed or emitted photons of both waves, and the differential cross sections (68) and (70) can easily be summed over all possible emission and absorption processes. The result is predictable: the total cross section coincides with the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field, i.e., as a result of summation, all essentially quantum contributions completely cancel (as for a single wave[7]):

$$d\sigma=\sum_{l=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}d\sigma_{ls}=d\sigma_*. \tag{71}$$

We next deal with nonrelativistic electron energies. Here we assume that the velocities of electron oscillations in the field of both waves are small compared to the speed of light:

$$v_{i,f}\ll 1, \quad v_{F_{1,2}}=\eta_{1,2}\ll 1. \tag{72}$$

Under these conditions, the expressions (8) and (9) for the amplitude become much simpler, and after doing the necessary calculations as in the relativistic case, we obtain (66). In the process, energy and momentum conservation require that

$$\frac{\mathbf{p}_f^2}{2m}-\frac{\mathbf{p}_i^2}{2m}+b\mathbf{n}\cdot(\mathbf{p}_f-\mathbf{p}_i)+\omega'+l\omega_1+s\omega_2=0,$$
$$\mathbf{q}=\mathbf{p}_f-\mathbf{p}_i+\mathbf{k}', \tag{73}$$

where

$$b=\frac{1}{4}[(1+\delta_1^2)\eta_1^2+(1+\delta_2^2)\eta_2^2]. \tag{74}$$

The arguments $\chi_{1,2}$ and $\gamma_{1,2}$ of the functions I_{ls} are given by (23) and (24), in which

$$\mathbf{g}_{1,2}=\frac{\mathbf{p}_f-\mathbf{p}_i}{\omega_{1,2}m}, \tag{75}$$

and

$$\beta_{1,2}=\frac{(1-\delta_{1,2}^2)\,\eta_{1,2}^2}{8}\,\frac{\mathbf{n}\cdot(\mathbf{p}_f-\mathbf{p}_i)}{\omega_{1,2}},$$

$$\alpha_{\pm}=\eta_1\eta_2|d_{\mp}|\,\frac{\mathbf{n}\cdot(\mathbf{p}_i-\mathbf{p}_f)}{2(\omega_1\pm\omega_2)}. \tag{76}$$

Clearly, energy conservation in the form (73) holds for wave intensities

$$\eta_{1,2}^2\gtrsim v_i, \tag{77}$$

i.e., the velocities of electron oscillations in the waves are much greater than the velocity of the electron's translation motion ($\eta_{1,2}\gg v_i$). We write the law of energy conservation (73) in the form of a standard quadratic equation for the electron velocity in the final state:

$$v_f^2+2b_fv_f-c_i=0, \tag{78}$$

where

$$b_f=b\cos\theta_f,\quad c_i=v_i^2+2bv_i\cos\theta_i-\frac{2(l\omega_1+s\omega_2+\omega')}{m}, \tag{79}$$

with $\theta_{i,f}=\angle(\mathbf{n},\mathbf{p}_{i,f})$.

The solution of Eq. (78) depends heavily on the signs of the coefficients b_f and c_i. If $c_i>0$, there is only one physically meaningful root (for all b_f), but if $c_i<0$, there are two such roots (for $b_f<0$). Note that the sign of b_f depends on the electron scattering angle relative to the direction of propagation of both waves: $b_f>0$ for $\theta_f\leqslant\pi/2$ and $b_f<0$ for $\theta_f>\pi/2$. On the other hand, the sign of c_i depends on both the polar angle θ_i and the total number of emitted and absorbed photons of both waves. The final solution of Eq. (79) is

$$v_f=\begin{cases} v_1\equiv-b_f+\sqrt{b_f^2+c_i} & \text{for } c_i>0 \text{ and all } \theta_f,\\ v_{1,2}\equiv-b_f\pm\sqrt{b_f^2+c_i} & \text{for } c_i<0 \text{ and } \theta_f>\pi/2.\end{cases} \tag{80}$$

Bearing all this in mind, we see that the delta function in the amplitude of spontaneous bremsstrahlung, when Eq. (78) has one or two roots, can written

$$\delta(\ldots)=\frac{\delta(v_f-v_1)}{m\sqrt{b_f^2+c_i}},$$

$$\delta(\ldots)=\frac{\delta(v_f-v_1)+\delta(v_f-v_2)}{m\sqrt{b_f^2+c_i}}. \tag{81}$$

Allowing for the nonrelativistic limit (66), doing the necessary averaging and summing over photon and electron polarizations, and integrating over the momenta of the final electrons, we obtain an expression for the differential cross section of spontaneous bremsstrahlung for the cases of one and two roots of Eq. (79), respectively:

$$d\sigma_{ls}=\frac{1}{\pi^2}Z^2r_0^2\alpha_0Y_{ls}(v_1)\,\frac{d\omega'}{\omega'}\,d\Omega_{\mathbf{k}'}d\Omega, \tag{82}$$

$$d\sigma_{ls}=\frac{1}{\pi^2}Z^2r_0^2\alpha_0[Y_{ls}(v_1)+Y_{ls}(v_2)]\,\frac{d\omega'}{\omega'}\,d\Omega_{\mathbf{k}'}d\Omega, \tag{83}$$

where r_0 is the classical electron radius, α_0 is the fine structure constant, and

$$Y_{ls}(v_f)=|I_{ls}|^2\frac{(mv_f)^2}{v_i\sqrt{b_f^2+c_i}}\,\frac{(\mathbf{q}\times\mathbf{n}')^2}{\mathbf{q}^4}. \tag{84}$$

Equations (24), (75), and (76) clearly show that in the intensity range (77), the values of the quantum parameters are such that $\beta_{1,2}\sim\alpha_{\pm}\gtrsim mv_i^2/\omega_{1,2}\gg1$ and $\gamma_{1,2}\sim\beta_{1,2}\eta_{1,2}^{-1}\gg\beta_{1,2}$. Hence, in the given intensity range, the principal multiphoton processes are those in which the numbers of photons in the first and second waves are $l\sim mv_i^2/\omega_1\eta_1$ and $s\sim mv_i^2/\omega_2\eta_2$, respectively.

We now consider the intensity range complementary to (77):

$$\eta_{1,2}^2\ll v_i,\quad \text{but}\quad \eta_{1,2}\gtrsim v_i. \tag{85}$$

Energy conservation Eq. (73) requires

$$\frac{\mathbf{p}_f^2}{2m}-\frac{\mathbf{p}_i^2}{2m}+\omega'+l\omega_1+s\omega_2=0. \tag{86}$$

The expression for the differential probability can be obtained from (82), in which we must set $b_f=0$, while in the coefficient c_i in (79) we must ignore the second term. The result is

$$d\sigma_{ls}=\frac{1}{\pi^2}Z^2r_0^2\alpha_0|I_{ls}|^2\,\frac{|\mathbf{p}_f|}{|\mathbf{p}_i|}\,\frac{m^2(\mathbf{q}\cdot\mathbf{n}')^2}{\mathbf{q}^4}\,\frac{d\omega'}{\omega'}\,d\Omega_{\mathbf{k}'}d\Omega. \tag{87}$$

Here the Bunkin–Fedorov quantum parameters are less than in the region (77), but still they are large: $\gamma_{1,2}\sim\eta_{1,2}mv_i/\omega_{1,2}\gtrsim mv_i^2/\omega_{1,2}\gg1$. In the intensity range given by (85), where the velocities of electron oscillations in the waves are of the order of the velocity of translational motion, emission and absorption of a large number of photons of both waves are therefore dominant: $l\sim mv_i^2/\omega_1$ and $s\sim mv_i^2/\omega_2$. Note that if the wave intensities obey conditions more stringent than the first inequality in (85), i.e.,

$$\eta_{1,2}^2\ll\frac{\omega_{1,2}}{mv_i},\quad \eta_1\eta_2\ll\frac{\omega_{1,2}}{mv_i}, \tag{88}$$

both $\beta_{1,2}$ and $\alpha_{\pm}$ will be much less than unity, so that in (82) we can put $|I_{ls}|^2=J_l^2(\gamma_1)J_s^2(\gamma_2)$.

The inequalities (88) are thus the conditions for the applicability of the dipole approximation, in contrast to less stringent condition (72) commonly used in nonrelativistic approaches to such problems. The reason is that in fields so strong that $\eta_{1,2}\sim v_i$, the typical wavelength of a spontaneous photon is of order $(\gamma_{1,2}\omega_{1,2})^{-1}$, rather than $\omega_{1,2}^{-1}$. Note that under the conditions (88), the second inequality in (85) may hold, but only at high frequencies, $\omega_{1,2}\gg mv_i^3$. For example, at $v_i=0.1$, the frequencies $\omega_{1,2}$ are much higher than 0.5 keV, and only at the limit of validity of the Born approximation, where $v_i\sim Z\times10^{-2}$, will we have $\omega_{1,2}\gg0.5\times Z^3$ eV, i.e., we wind up at optical frequencies.

Now we examine the last case, where the velocities of electron oscillations in the waves are small compared to the velocity of translational motion:

$$\eta_{1,2} \ll v_i . \tag{89}$$

Here we can ignore the energies of the emitted and absorbed photons of both waves: $|l|\omega_1/mv_i^2 \lesssim \eta_1/v_i \ll 1$ and $|s|\omega_2/mv_i^2 \lesssim \eta_2/v_i \ll 1$. The energy conservation expression (86) then assumes the standard form

$$\frac{\mathbf{p}_f^2}{2m} - \frac{\mathbf{p}_i^2}{2m} + \omega' = 0, \tag{90}$$

while the differential cross section (87) can be factored, i.e., it assumes the form (68), where $d\sigma_*$ is the nonrelativistic limit of spontaneous bremsstrahlung by an electron scattered by a nucleus in vanishing external fields.

5. SPONTANEOUS INTERFERENCE BREMSSTRAHLUNG IN THE PLANE PERPENDICULAR TO THE WAVE POLARIZATION VECTOR

We examine spontaneous bremsstrahlung by an electron scattered by a nucleus in the interference region for linear polarization of both waves:

$$\delta_1^2 = \delta_2^2 = 0, \quad \Delta = 0, \quad \mathbf{e}_{1x} = \mathbf{e}_{2x} \equiv \mathbf{e}_x \tag{91}$$

[the first possible case; see text following (45)]. Let the scattering of the electron and the emission of the spontaneous photon take place in the plane perpendicular to the polarization vector of both waves (the yz plane). Then $\mathbf{e}_x \cdot \mathbf{k}' = \mathbf{e}_x \cdot \mathbf{p}_i = \mathbf{e}_x \cdot \mathbf{p}_f = 0$. In this case the conditions (44) and (45) hold, i.e., the Bunkin–Fedorov quantum parameters and the corresponding phases vanish. Thus, in the given region, for arbitrary wave intensities, the Bunkin–Fedorov parameters are small if the angles between the pre- and post-scattering electron momenta and the polarization vector [$\varphi_{i,f} = \angle(\mathbf{p}_{i,f}, \mathbf{e}_x)$], and between the momentum of the spontaneous photon and the polarization vector [$\psi' = \angle(\mathbf{k}', \mathbf{e}_x)$], are close to $\pi/2$:

$$\left|\varphi_{i,f} - \frac{\pi}{2}\right| \ll \frac{\omega_{1,2}}{mv_{i,f}\eta_{1,2}} \lesssim 1,$$
$$\left|\psi' - \frac{\pi}{2}\right| \ll \frac{\omega_{1,2}E_{i,f}}{m\omega'\eta_{1,2}} \lesssim 1. \tag{92}$$

The amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus under the conditions (92) is determined by the expressions (52)–(60) with allowance for (91). Note that the noninterference region, in which the main multiphoton parameters are the Bunkin–Fedorov quantum parameters, is given by conditions complementary (92).

We first consider spontaneous bremsstrahlung at relativistic electron energies under the conditions (92). Here multiphoton processes are determined by the functions $J_{l_+ - s_+, l_- - s_-}$ of (46), and the main multiphoton parameters are the quantum parameters α_+ and α_- [see (26) under conditions (91)]. Hence, in the interference region, in addition to electron scattering by the nucleus and emission of a spontaneous photon, the main processes will be stimulated, correlated emission (absorption) of equal numbers of photons of both waves or of numbers that differ by 1 [$l_\pm \lesssim \alpha_\pm$; see text following Eqs. (51)]. Consequently, the fraction of the energy emitted or absorbed by the electron in the initial or final state from both waves is of order $l_\pm(\omega_1 \pm \omega_2)/E_{i,f} \lesssim \zeta_{\rm int}$, where $\zeta_{\rm int}$ is the classical interference parameter that determines the constants of the motion of the process in the region given by (92),

$$\zeta_{\rm int} = \xi_1 \xi_2 v_i , \tag{93}$$

where the parameters $\xi_{1,2}$ are given by (62).

We characterize moderately strong fields for the interference region (92) by a condition on the classical interference parameter, $\zeta_{\rm int} \ll 1$. For the product of the wave intensities this condition becomes

$$\eta_1 \eta_2 \ll \begin{cases} 1 & \text{if} \quad E_{i,f} \sim m, \\ \left(\dfrac{E_{i,f}}{m}\right)^2 & \text{if} \quad E_{i,f} \gg m. \end{cases} \tag{94}$$

Note that when $\eta_{1,2} < 1$, the conditions (94) are less stringent than the analogous conditions in the noninterference region (63). Allowing for (94), in the expressions for the squares of the 4-momenta (50) and (51) we can ignore the energies of photons of combination frequencies in comparison to the electron energy. Hence, $\tilde{q}_i^2 \neq m_*^2$ and $\tilde{q}_f^2 \neq m_*^2$, with the result that the range of moderately strong fields given by (94) is the nonresonance region. If $\eta_1 \sim \eta_2$ under the conditions (94), then $\tilde{p}_{i,f} = p_{i,f}$, and the 4-momenta (50) and (51) assume the form (64). Note that if we use (52) and (53) to obtain the cross section, terms proportional to $\eta_{1,2}^0$, $\eta_{1,2}^2$, $\eta_1\eta_2$, and higher powers of the wave intensities emerge in $d\sigma^{(0)}_{l_+ l_-}$, and terms proportional to $\eta_{1,2}^2$, $\eta_1\eta_2$, and higher powers of wave intensities in $d\sigma^{(1)}_{l_+ l_-}$. By virtue of (94), the amplitude $H^{(1)}_{l_+ l_-}$ of (53) can therefore be ignored, while in the amplitude $H^{(0)}_{l_+ l_-}$ in (53) we can ignore the function $B^{(0)}_{rr'}$ of (59) and the 4-vector $D^{(1)}_{rr'}$ of (60), and sum over s_+ and s_-. As a result,

$$S^{(1)}_{l_+ l_-} = 0, \quad S^{(0)}_{l_+ l_-} = J_{l_+ l_-}(\beta_1, \beta_2; 0, 0, \alpha_+, \alpha_-) S_*, \tag{95}$$

where S_* is the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field,[22] and the functions $J_{l_+ l_-}$ are given by the expression (46) in whose arguments (25) and (26) we must put $\tilde{p}_1 = p_i$ and $\tilde{p}_2 = p_f$ and allow for (91):

$$\beta_j = \frac{\eta_j^2 m^2}{8\omega_j}\left(\frac{1}{\kappa_f} - \frac{1}{\kappa_i}\right), \quad j = 1,2, \tag{96}$$

$$\alpha_\pm = \frac{\eta_1 \eta_2 m^2}{2(\omega_1 \pm \omega_2)}\left(\frac{1}{\kappa_i} - \frac{1}{\kappa_f}\right), \quad \kappa_{i,f} = E_{i,f} - \mathbf{n}\cdot\mathbf{p}_{i,f} . \tag{97}$$

If we use (95), we can easily obtain an expression for the differential cross section of spontaneous bremsstrahlung by a relativistic electron scattered by a nucleus in the field of two moderately strong waves [the conditions (94)] in the interference region given by (92):

$$d\sigma^{(0)}_{l_+ l_-} = |J_{l_+ l_-}(\beta_1, \beta_2; 0, 0, \alpha_+, \alpha_-)|^2 d\sigma_* , \tag{98}$$

where $d\sigma_*$ is the differential cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus with emission of a photon of given frequency and direction and with the ejection of the final electron in a given direction (in a vanishing external field).[22] Note that in the differential cross section (98) the integer indices $l_\pm$ denote the emission (absorption) of equal numbers ($s=\pm l$) of photons of both waves. This is what makes (98) so different from the expressions (68) and (70) for the noninterference region, where there is no such correlation of the indices l and s. Equation (98) shows that in moderately strong fields [see (94)], the differential cross section becomes the product of the probability of correlated emission (absorption) of equal numbers of photons of both fields and the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field. What is important is that the given correlation of the numbers of photons of both waves in the interference region cannot be removed, i.e., the cross section (98) cannot be represented as the product of the probabilities of emission (absorption) in each of the two waves separately, in contrast to the situation in the noninterference region [see (70)].

Note that at wave intensities complementary to (69), i.e.,

$$\eta_{1,2}^2 \gtrsim \begin{cases} \omega_{1,2}/m & \text{if} \quad E_i \sim m, \\ \omega_{1,2}E_i/m^2 & \text{if} \quad E_i \gg m, \end{cases}$$
$$\eta_1\eta_2 \gtrsim \begin{cases} \omega_{1,2}/m & \text{if} \quad E_i \sim m, \\ \omega_{1,2}E_i/m^2 & \text{if} \quad E_i \gg m, \end{cases} \tag{99}$$

the quantum parameters are such that $\beta_{1,2}\gtrsim 1$ and $\alpha_\pm \gtrsim 1$, with the result that $\gamma_{1,2}\gg\alpha_\pm\gtrsim 1$ and $\gamma_{1,2}\gg\beta_{1,2}\gtrsim 1$ [see (61), (67), (96), and (97)]. Hence, at field intensities satisfying (99) and (63) we can easily show[30] that the differential cross section (68) in the noninterference region is much smaller than the corresponding cross section (98) in the interference region:

$$\frac{d\sigma_{ls}}{d\sigma^{(0)}_{l_+l_-}} = \frac{|I_{ls}|^2}{|J_{l_+l_-}|^2} \sim (\gamma_1\gamma_2)^{-1} \ll 1 \tag{100}$$

if $l_+ \sim l \ll \gamma_1$ and $l_- \sim s \ll \gamma_2$.

Thus, if we adjust the measuring device so it detects (in the plane perpendicular to the polarization vector of the waves) the spontaneous photon and the final electron simultaneously, we find that spontaneous bremsstrahlung with multiphoton correlated emission (absorption) of equal numbers of photons of both waves can dominate spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of both waves in the noninterference region. The differential cross section (98) can be summed over all emission and absorption processes. The result is similar to that for the noninterference region, i.e., under summation all essentially quantum contributions completely cancel [see (71)]:

$$d\sigma = \sum_{l_+=-\infty}^{\infty} \sum_{l_-=-\infty}^{\infty} d\sigma^{(0)}_{l_+l_-} = d\sigma_* . \tag{101}$$

Now we examine the range of nonrelativistic electron energies under the conditions (72). Here the fraction of the energy of photons of combination frequencies emitted or absorbed by the electron in the initial or final states from both waves is of order $2l_\pm(\omega_1 \pm \omega_2)/mv_{i,f}^2 \lesssim \zeta_{\rm int}$, where in the nonrelativistic limit of $\zeta_{\rm int}$ [see Eq. (93)] we have

$$\zeta_{\rm int} = \frac{\eta_1\eta_2}{v_i} . \tag{102}$$

Under the conditions (72), the interference amplitude given by (52) and (53) becomes much simpler, and after we have done calculations similar to those in the relativistic case, the amplitude assumes the form of the nonrelativistic limit (95):

$$S^{(1)}_{l_+l_-} = 0, \quad S^{(0)}_{l_+l_-} = J_{l_+l_-}(\beta_1,\beta_2;0,0,\alpha_+,\alpha_-) S_*^{v\ll 1}, \tag{103}$$

where $S_*^{v\ll 1}$ is formally similar to the nonrelativistic limit of the amplitude of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field[22] with allowance for energy and momentum conservation in the form

$$\frac{\mathbf{p}_f^2}{2m} - \frac{\mathbf{p}_i^2}{2m} + b\mathbf{n}\cdot(\mathbf{p}_f - \mathbf{p}_i) + \omega' + l_+(\omega_1+\omega_2) + l_-(\omega_1-\omega_2) = 0,$$
$$\mathbf{q} = \mathbf{p}_f - \mathbf{p}_i + \mathbf{k}', \tag{104}$$

where

$$b = \frac{1}{4}(\eta_1^2 + \eta_2^2). \tag{105}$$

The parameters $\beta_{1,2}$ and $\alpha_\pm$ in the functions $J_{l_+l_-}$ of (103) are

$$\beta_{1,2} = \eta_{1,2}^2 \frac{\mathbf{n}\cdot(\mathbf{p}_f - \mathbf{p}_i)}{8\omega_{1,2}}, \quad \alpha_\pm = \eta_1\eta_2 \frac{\mathbf{n}\cdot(\mathbf{p}_i - \mathbf{p}_f)}{2(\omega_1 \pm \omega_2)}. \tag{106}$$

Energy conservation in the form (104) holds for wave intensities

$$\eta_{1,2}^2 \gtrsim v_i , \tag{107}$$

i.e., when the velocities of electron oscillations in the waves are much greater than the electron velocity of translational motion ($\eta_{1,2}\gg v_i$). The reasoning and calculations that must follow are the same as for the nonrelativistic limit in the noninterference region [see Eqs. (78)–(84)]. The result is an expression for the differential cross section of spontaneous bremsstrahlung in the interference region (92) for the cases of one and two roots of the energy conservation expression (104), respectively:

$$d\sigma^{(0)}_{l_+l_-} = \frac{1}{\pi^2} Z^2 r_0^2 \alpha_0 T_{l_+l_-}(v_1) \frac{d\omega'}{\omega'} d\Omega_{\mathbf{k}'} d\Omega, \tag{108}$$

$$d\sigma^{(0)}_{l_+l_-} = \frac{1}{\pi^2} Z^2 r_0^2 \alpha_0 [T_{l_+l_-}(v_1) + T_{l_+l_-(v_2)}] \frac{d\omega'}{\omega'} d\Omega_{\mathbf{k}'} d\Omega, \tag{109}$$

where the velocities $v_{1,2}$ are given by (80), and

$$T_{l_+l_-}(v_f)=|J_{l_+l_-}(\beta_1,\beta_2;0,0,\alpha_+,\alpha_-)|^2 \times\frac{(mv_f)^2}{v_i\sqrt{b_f^2+c_i}}\frac{(\mathbf{q}\cdot\mathbf{n}')^2}{\mathbf{q}^4}. \quad (110)$$

Here in the expression for c_i in (79) we must replace $l\omega_1$ with $l_+(\omega_1+\omega_2)$ and $s\omega_2$ with $l_-(\omega_1-\omega_2)$. Equations (106) show that in the intensity range (107) the values of the quantum parameters are such that $\beta_{1,2}\sim\alpha_\pm\gtrsim mv_i^2/\omega_{1,2}\gg 1$ and $\gamma_{1,2}\sim\beta_{1,2}\eta_{1,2}^{-1}\gg\beta_{1,2}$. Hence, in the given intensity range, multiphoton processes have photon numbers $l_\pm\sim mv_i^2/\omega_{1,2}\eta_{1,2}\gg 1$. Note that the differential cross sections (108) and (109) in the interference region are much greater than the corresponding cross sections (82) and (83) in the noninterference region.

We now turn to the intensity range

$$\eta_{1,2}^2\ll v_i. \quad (111)$$

Under these conditions, $\zeta_{\rm int}=\eta_1\eta_2 v_i^{-1}\ll 1$, i.e., we are dealing with moderately strong fields. We can therefore neglect the energies of the emitted and absorbed photons, and also set $b_f=0$ and $c_i=v_i$. As a result, the energy conservation expression (104) assumes the standard form (90), and the expression for the differential probability can be found from (108):

$$d\sigma_{l_+l_-}^{(0)}=|J_{l_+l_-}(\beta_1,\beta_2;0,0,\alpha_+,\alpha_-)|^2 d\sigma_*^{v\ll 1}, \quad (112)$$

where $d\sigma_*^{v\ll 1}$ is the nonrelativistic limit of spontaneous bremsstrahlung by an electron scattered by a nucleus in vanishing external fields. We see that for intensities (111) the differential cross section becomes the product of the probability of correlated emission (absorption) of equal numbers of photons of both waves and the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field. The differential cross section can easily be summed over all emission and absorption processes. As a result, all the essentially quantum contributions completely cancel [see (71)]. We also note that for field intensities

$$\eta_{1,2}^2\gtrsim\frac{\omega_{1,2}}{mv_i}, \quad \eta_1\eta_2\gtrsim\frac{\omega_{1,2}}{mv_i}, \quad (113)$$

$\beta_{1,2}\gtrsim 1$ and $\alpha_\pm\gtrsim 1$, and hence $\gamma_{1,2}\gg\alpha_\pm\gtrsim 1$ and $\gamma_{1,2}\gg\beta_{1,2}\gtrsim 1$ [see (61) and (106)]. For wave intensities satisfying the conditions (113) and (72), we can thus easily show [see (100)] that the nonrelativistic limit of the differential cross section (68) in the noninterference region is much smaller than the corresponding cross section (112) in the interference region, i.e., the latter can be dominant.

6. SPONTANEOUS INTERFERENCE BREMSSTRAHLUNG FOR ELLIPTICAL WAVE POLARIZATION

In this section we examine spontaneous bremsstrahlung by an electron scattered by a nucleus in the interference region (43) with both waves being elliptically polarized, i.e., in the general case [see text following formula (45)]. Here Eqs. (44) and (45) are valid, i.e., the Bunkin–Fedorov quantum parameters and the corresponding phases vanish. With both waves having arbitrary wave intensities, the Bunkin–Fedorov quantum parameters in this region are small if the angles between the vectors $\mathbf{g}_i$ [see (23)] and the directions of propagation of the two waves are small:

$$\psi_j\ll\frac{\omega_{1,2}}{mv_{i,f}\eta_{1,2}}\lesssim 1. \quad (114)$$

Equations (43) lead to four conditions on the kinematics of electron scattering and spontaneous-photon ejection at arbitrary wave intensities:

$$\mathbf{g}_j^2(\tilde p_f,\tilde q_i)=(\mathbf{n}\cdot\mathbf{g}_j(\tilde p_f,\tilde q_i))^2, \quad \mathbf{g}_j^2(\tilde q_i,\tilde p_i)=(\mathbf{n}\cdot\mathbf{g}_j(\tilde q_i,\tilde p_i))^2, \quad j=1,2, \quad (115a)$$

for the amplitude of process a (Fig. 1), and

$$\mathbf{g}_j^2(\tilde q_f,\tilde p_i)=(\mathbf{n}\cdot\mathbf{g}_j(\tilde q_f,\tilde p_i))^2, \quad \mathbf{g}_j^2(\tilde p_f,\tilde q_f)=(\mathbf{n}\cdot\mathbf{g}_j(\tilde p_f,\tilde q_f))^2, \quad j=1,2, \quad (115b)$$

for the amplitude of process b. Note that $k_jg_j=0$ and therefore $g_0=g_z$, so the conditions (43) can be written in the Lorentz-invariant form

$$g_j^2(\tilde p_2,\tilde p_1)=0, \quad (116)$$

where the 4-momenta $\tilde p_{1,2}$ assume the four possible sets of values in (115a) and (115b). We also note that the first equations in (115) correspond to the amplitude of electron scattering by the nucleus in the field of the waves [see (9) and (16)], while the second equations correspond to the amplitude of photon emission by the electron in the field of the waves [see (9) and (17)]. Substituting the explicit form of the vectors $\mathbf{g}_j$ from (23) together with the corresponding momenta p_1 and p_2 into (115), we find that the kinematics of electron scattering and spontaneous-photon ejection for the amplitudes of a and b is the same. Here the scattering of the electron and the emission of the spontaneous photon occur in a single plane formed by the initial electron momentum and the direction of propagation of both waves. The corresponding azimuthal angles are equal, while the polar angles are given by

$$a_f=a_i, \quad a_{i,f}=\frac{|\mathbf{p}_{i,f}|}{\kappa_{i,f}}\sin\theta_{i,f} \quad (117)$$

for electron scattering, and

$$\cot\frac{\theta'}{2}=a_i, \quad \theta'=\angle(\mathbf{n},\mathbf{k}') \quad (118)$$

for the exit angle of the spontaneous photon. In (117), the $\kappa_{i,f}$ are given by (97). Note that the relations (117) and (118) coincide, respectively, with the conditions in the interference region for stimulated bremsstrahlung and inverse bremsstrahlung by an electron scattered by a nucleus in the field of two waves,[11] and for spontaneous photon emission by an electron in the field of two waves.[20]

We first consider spontaneous bremsstrahlung in the interference region (114) at relativistic electron energies and in moderately strong fields, $\zeta_{\rm int}\ll 1$, i.e., under the conditions (94). Then the reasoning in Sec. 5 [see text between formulas (92) and (95)] is valid. Hence, we can ignore the corrections

associated with the correlated emission and absorption of photons of both waves, and the expressions (50) and (51) for the 4-momentabecome (64), i.e., energy conservation requires that

$$E_f = E_i - \omega'. \tag{119}$$

Analyzing the kinematic conditions (117), we see that

$$v_f = \frac{|\mathbf{p}_f|}{E_f} = \frac{a_i}{\sin\theta_f + a_i \cos\theta_f}, \quad \kappa_f = \frac{\tan\theta_f}{a_i + \tan\theta_f} E_f. \tag{120}$$

Combining the expression for the final electron velocity in (120) with (119) and doing simple transformations, we obtain an equation for the scattering angle of the electron:

$$a_i\left(\frac{1}{v_f}+1\right)\tan^2\frac{\theta_f}{2} - 2\tan\frac{\theta_f}{2} + a_i\left(\frac{1}{v_f}-1\right) = 0, \tag{121}$$

where

$$v_f = \sqrt{1-\left(\frac{m}{E_i-\omega'}\right)^2}. \tag{122}$$

Solving Eq. (121), we obtain two possible scattering angles for the final electron:

$$\tan\frac{\theta_f}{2} = \frac{1}{a_i(v_f^{-1}+1)} \times\left[1 \pm \sqrt{\frac{(\omega'_{\max}-\omega')(2E_i-\omega'_{\max}-\omega')}{(E_i-\omega')^2-m^2}}\right], \tag{123}$$

where

$$\omega'_{\max} = E_i - m\sqrt{1+a_i^2} \tag{124}$$

is the maximum frequency of the spontaneous photon. The solution (123) shows that the radiative spectrum in the interference region (114) is bounded from above by $\omega'_{\max}$, in contrast to the spectra in the interference region (92) and the noninterference region (see Sec. 4), where the maximum frequency of the spontaneous photon is $E_i - m$. Note that the upper limit (124) of the frequency of the spontaneous photon depends heavily on the energy and incidence angles of the initial electron. If we consider it a function of the polar angle of the initial electron, it has minimum value 0 at the polar angle

$$\theta_{i*} = \arccos v_i \tag{125}$$

(θ_{i*} is the critical angle, at which the electron ceases to emit photons), and a maximum value $E_i - m$ at $\theta_i = 0, \pi$ (Fig. 2). Hence, as $\theta_i \to \theta_{i*}$ (from both the right and left), emission rapidly decreases ($\omega'_{\max} \to 0$), and in a narrow range of angles near the critical angle $[(\theta_i - \theta_{i*})^2 \ll 1]$ it essentially vanishes ($\omega'_{\max}/m \ll 1$). As we will shortly show, we then have forward electron scattering without emission or absorption of photons of both waves [see text following Eq. (137)]. Thus, the electron emits photons if its polar incidence angle lies in the range

$$0 \leqslant \theta_i < \theta_{i*} \quad \text{or} \quad \theta_{i*} < \theta_i \leqslant \pi, \tag{126}$$

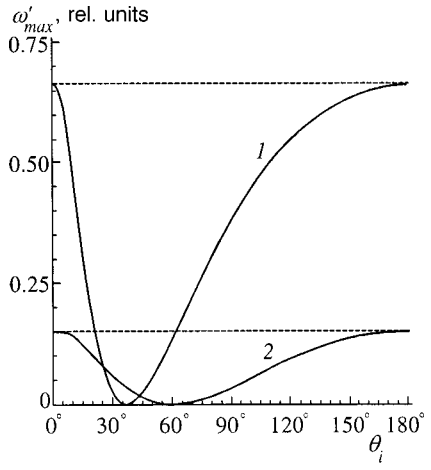


FIG. 2. Maximum frequency of the spontaneous photon (in units of electron rest energy) vs. polar incidence angle of the initial electron [Eq. (124)]. Curve *1* corresponds to electron velocity $v_i = 0.8$ and curve *2*, to $v_i = 0.5$. The minimum is at the point where $\cos\theta_i = v_i$. The horizontal dashed lines correspond to the maximum frequency of the spontaneous photons for the corresponding electron velocities in a vanishing external field.

which is not very close to the critical angle. Note that the critical angle θ_{i*} increases with decreasing initial electron velocity, taking values from $\theta_{i*} \approx \sqrt{2(1-v_i)} \ll 1$ (for ultrarelativistic electron energies) to $\theta_{i*} \approx \pi/2$ (for nonrelativistic electron energies).

We now ascertain the angles at whichthe spontaneous photon emerges. From (118) we see that the exit angle of the spontaneous photon depends heavily on the energy and incidence angles of the initial electron (Fig. 3). If we consider the exit angle of the spontaneous photon as a function of the polar angle of the initial electron, at the critical angle $\theta_i = \theta_{i*}$ it has a minimum value (at which there is no emission) equal to

$$\theta'_{\min} = 2\arcsin\frac{m}{E_i} = 2\theta_{i*}. \tag{127}$$

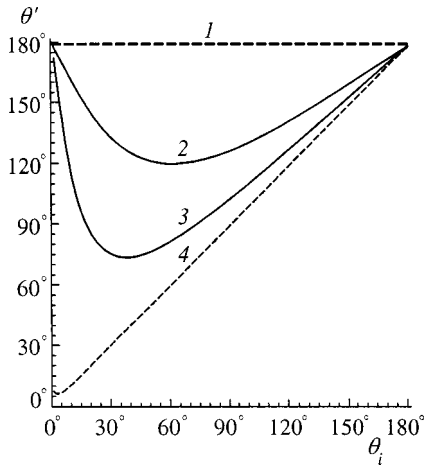


FIG. 3. Exit angle θ' vs. polar incidence angle θ_i of the initial electron [Eq. (118)] for various initial electron velocities. The straight dashed lines correspond to nonrelativistic electrons with $v_i = 0.01$ (line *1*) and ultrarelativistic electrons with $v_i = 0.999$ (line *4*). The solid curves correspond to electron velocities $v_i = 0.5$ (curve *2*) and $v_i = 0.8$ (curve *3*).

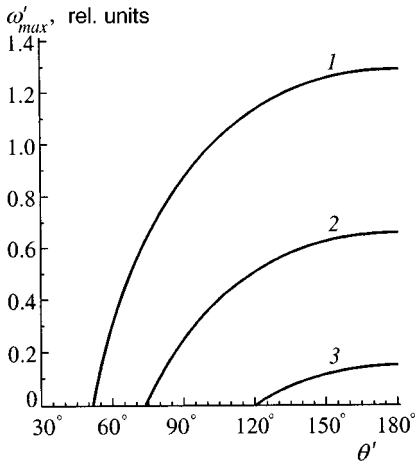


FIG. 4. Maximum frequency of the spontaneous photon (in units of electron rest energy) vs. polar exit angle of the spontaneous photon [Eq. (129)]. Curve *1* corresponds to the electron velocity $v_i=0.9$ and $\theta'_{\min}=51.7°$; curve *2*, to $v_i=0.8$ and $\theta'_{\min}=73.7°$; and curve *3*, to $v_i=0.5$ and $\theta'_{\min}=120°$.

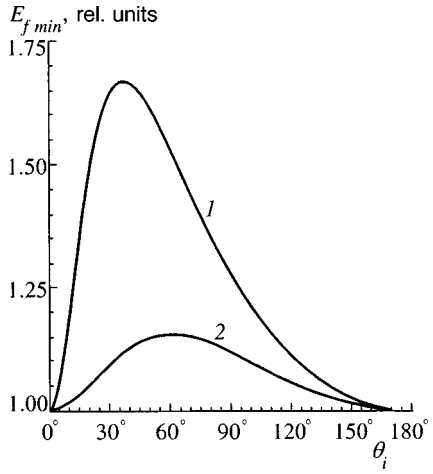


FIG. 5. Minimum energy of the final electron (in units of the electron rest energy) vs. polar incidence angle of the initial electron [Eq. (130)]. Curve *1* corresponds to the electron velocity $v_i=0.8$ and curve *2*, to $v_i=0.5$. The maximum is at the point where $\cos\theta_i=v_i$.

Hence, for a given energy E_i of the initial electron and various incidence angles (126), the exit angle of the spontaneous photon satisfies

$$\theta'_{\min}<\theta'\leqslant\pi, \tag{128}$$

i.e., has a lower bound determined by the energy of the initial electron. It is important to note here that to each direction of exit of the spontaneous photon there corresponds a specific upper bound of the emission spectrum. This dependence can easily be obtained from (118) and (125):

$$\omega'_{\max}=E_i-\frac{m}{\sin(\theta'/2)}, \tag{129}$$

where for a given energy of the initial electron the exit angle θ' can be chosen to be in the range (128). Combining (118), (124), and (129), we see that by continuously varying the polar incidence angle of the initial electron (with given energy), we can continuously change the upper bound of the frequency and exit angle of the spontaneous photon (Fig. 4).

The frequency of the spontaneous photon has an upper bound, so the energy of the final electron is bounded from below:

$$E_{f\,\min}=E_i-\omega'_{\max}=m\sqrt{1+a_i^2}. \tag{130}$$

Obviously, at the critical angle $\theta_i=\theta_{i*}$ (where $\omega'_{\max}=0$), the minimum energy (130) has a maximum (Fig. 5). Hence, the energy of the final electron lies in the range

$$E_{f\,\min}\leqslant E_f\leqslant E_i\,. \tag{131}$$

What are the angles at which the final electron exits? Figure 6 shows the dependence of the exit polar angle of the final electron on the frequency of the spontaneous photon at given energy and polar incidence angle of the initial electron. We see that to each frequency there correspond two possible exit angles of the electron, and as the frequency of the spontaneous photon increases, the exit angle of the final electron first varies only slightly, but then near the maximum frequency the variation of the exit angle of the electron becomes significant. Note that this tendency becomes stronger as the electron energy increases.

Let us find the exit angle of the final electron at $\omega'=\omega'_{\max}$. If we allow for (123), we see that

$$\cot\theta_f=a_i\,. \tag{132}$$

Taking (132) and (118) into account yields

$$\theta_f=\frac{1}{2}\,\theta'. \tag{133}$$

This equation relates the exit angles of the final electron and the spontaneous photon emitted at the maximum frequency, for arbitrary energies and angles of the initial electron. Note that the diagram representing the dependence of the exit angle of the final electron on the incidence angle of the initial electron [Eq. (132)] has exactly the same shape as the corresponding diagram for the exit angle of the spontaneous pho-

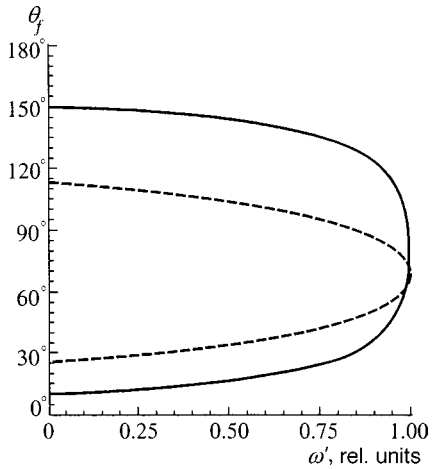


FIG. 6. Exit polar angle of the final electron vs. frequency of the spontaneous photon (in units of $\omega'_{\max}$) [Eq. (123)]. The dashed curve corresponds to the initial parameters $v_i=0.5$ and $\theta_i=25°$ and the solid curve, to $v_i=0.5$ and $\theta_i=150°$.

ton (Eq. (118); see Fig. 3), the only difference being that it is scaled down along the vertical axis by a factor of two.

In the moderately strong fields given by (94), the differential cross section of spontaneous bremsstrahlung for the relativistic case has the form

$$d\sigma^{(0)}_{l_+l_-}=|J_{l_+l_-}(\beta_1,\beta_2;\chi_+,\chi_-,\alpha_+,\alpha_-)|^2 d\sigma_*, \quad (134)$$

where $d\sigma_*$ is the differential cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing field,[22] which amounts to emission of a photon of given frequency (with upper bound $\omega'_{\max}$) and direction [see (118)], and escape of the final electron in the given direction [see (123)]. Here the functions $J_{l_+l_-}$ are given by (46); their arguments are

$$\beta_j=\frac{(1-\delta_j^2)\eta_j^2 m^2}{8\omega_j}\left[\frac{a_i+\tan\theta_f}{(E_i-\omega')\tan\theta_f}-\frac{1}{\kappa_i}\right], \quad j=1,2, \quad (135)$$

$$\alpha_\pm=\frac{\eta_1\eta_2|d_\mp|m^2}{2(\omega_1\pm\omega_2)}\left[\frac{1}{\kappa_i}-\frac{a_i+\tan\theta_f}{(E_i-\omega')\tan\theta_f}\right]. \quad (136)$$

When both waves are circularly polarized ($\delta_1^2=\delta_2^2=1$), the quantum parameters and the functions $J_{l_+l_-}$ become the Bessel functions $J_{l_\pm}$ of integer order. Hence the expression (134) for the differential cross section becomes much simpler:

$$d\sigma^{(0)}_{l_\pm}=J^2_{l_\pm}(\alpha_\pm)d\sigma_*, \quad (137)$$

where the plus and minus signs correspond to opposite ($\delta_1=-\delta_2=1$, whereupon $|d_-|=2$ and $|d_+|=0$) and identical ($\delta_1=\delta_2=1$, yielding $|d_-|=0$ and $|d_+|=2$) polarizations of the waves, and the index $l_\pm$ corresponds to correlated emission and absorption of equal numbers of photons of both waves ($s=\pm l$). From (134) and (137) it follows that in moderately strong fields [the conditions (94)], the differential cross section of spontaneous bremsstrahlung in the interference region (114) becomes the product of the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing field and the probability of correlated emission (absorption) of equal numbers of photons of both waves. When both waves are circularly polarized, the latter is given by the square of a Bessel function, and the quantum interference parameter $\alpha_\pm$ is the multiphoton parameter. Note that the probabilities (134) and (137) cannot be represented as a product of probabilities of emission (absorption) of photons of each wave separately.

We emphasize that when the incidence angle of the initial electron ($\theta_i=\theta_{i*}$) is critical ($\omega'_{\max}=0$), we have $\omega'_{\max}=0$, and from (123) or (127) and (133) it follows that the electron is then scattered in the forward direction ($\theta_f=\theta_i$). Bearing this in mind, (135) and (136) imply that $\alpha_\pm=\beta_{1,2}=0$ here, i.e., the only processes that occur are those with no emitted (absorbed) photons ($l_\pm=0$). Hence examining multiphoton correlated emission (absorption) of equal numbers of photons of both waves has meaning only if the incidence angle of the initial electron is not close to the critical angle, and $\omega'_{\max}\sim m$. For given initial electron energy and polar incidence angles not close to the critical angle (125), the differential cross sections (134) and (137) determine the multiphoton correlated emission (absorption) of equal numbers of photons of both waves, with the emission spectrum of the spontaneous photon bounded from above by the frequency $\omega'_{\max}$ (124), and the correlated exit of the spontaneous photon and final electron in the plane specified by the initial electron momentum of the wave vector at given angles (118) (bounded from above by $\theta'_{\min}$) and (123), respectively. Here, irrespective of the energy and incidence angle of the initial electron, the spontaneous photon of maximum frequency exits at twice the exit angle of the final electron given by (133) (the exit angles of the final electron are bounded from above in this case by $\pi/2$). Most importantly, for wave intensities satisfying the conditions (99) and (63), the differential cross sections (134) and (137) in the interference region are much greater than the corresponding cross sections (68) in the noninterference region [see text near Eqs. (99) and (100)].

We now consider ultrarelativistic energies of the initial electron, $E_i\gg m$. Here $a_i\approx\cot(\theta_i/2)$, and (124) and (130) yield expressions for the maximum energy of the spontaneous photon and the minimum energy of the final electron:

$$\omega'_{\max}=E_i-\frac{m}{\sin(\theta_i/2)}, \quad E_{f\,\min}=\frac{m}{\sin(\theta_i/2)}. \quad (138)$$

Note that these expression were obtained under the assumption that $\theta_i^2\gg m^2/E_i^2$, i.e., far from the critical angle $\theta_{i*}\approx m/E_i$. We also note that for frequencies of the spontaneous photon far from the maximum frequency in (138), in the main region of the emission spectrum the final electron scatters into a narrow cone along the wave vector and in the direction of the initial electron momentum:

$$\theta_f=\begin{cases}\theta_i\pm\dfrac{m^2\sin\theta_i}{2E_i^2(1-\omega'/\omega'_{\max})^2}\approx\theta_i,\\ \dfrac{m^2\tan(\theta_i/2)}{2E_i^2(1-\omega'/\omega'_{\max})^2}\ll 1.\end{cases} \quad (139)$$

Here the spontaneous photon scatters into a narrow cone in the direction of the initial electron momentum [see (118); $a_i^{-1}=\tan(\theta_i/2)+(1-v_i)/\sin\theta_i\approx\tan(\theta_i/2)$]:

$$\theta'=\theta_i\pm\frac{m^2}{2E_i^2}\cot\theta_i\cos\theta_i\approx\theta_i. \quad (140)$$

But if the spontaneous photon is emitted at the maximum frequency,

$$\theta_f=\frac{1}{2}\theta_i, \quad (141)$$

which also follows from (133) and (140). Thus, in the ultrarelativistic case, the energy and incidence angle of the initial electron affect only the vertex angle of the narrow exit cone of the spontaneous photon and the final electron. Here the spontaneous photon is always emitted into a narrow cone about the initial electron momentum, while for the final electron, which may be either ultrarelativistic or nonrelativistic [see (138)], there are three alternative exit geometries:

a) the final electron exits in a narrow cone along the initial electron momentum;

b) the final electron exits in a narrow cone along the wave vector;

c) the final electron exits at the angle given by (141).

Note that the first two cases, described by (139), apply if the frequency of the spontaneous photon is not near the upper bound, i.e.,

$$\left(1-\frac{\omega'}{\omega'_{\max}}\right)^2 \gg \frac{m^2}{E_i^2}, \tag{142}$$

while case (c), described by (141), apply if the inequality when the opposite of (142) holds, i.e., when the spontaneous photon is emitted at the maximum frequency. The differential cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus is given by the expressions (134) and (137) with allowance for (139)–(141). These expressions hold in the cases considered if the angles between the momentum of the spontaneous photons and the momenta of the ultrarelativistic initial and final electrons are not very small,

$$(\theta'-\theta_{i,f})^2 \gg \eta_1\eta_2\frac{m^2}{\omega' E_{i,f}} \ll 1, \tag{143}$$

and if the angle between the wave vector and the momentum of ultrarelativistic final electron is not very small,

$$\theta_f^2 \gg \eta_1\eta_2\frac{m^2}{E_f^2} \ll 1. \tag{144}$$

Combining (139), (140), (143), and (144), we obtain a constraint on the product of the field intensities:

$$\eta_1\eta_2 \ll \frac{m^2\omega'}{E_i^3}. \tag{145}$$

Note that this inequality is much more stringent than the corresponding conditions (94), which characterize moderately strong fields. Nevertheless, under the condition (145) the following inequalities hold: $\alpha_\pm \gtrsim 1$ and $\beta_{1,2} \gtrsim 1$ [see (135) and (136)]. Hence, the differential cross sections (134) and (137) can be, in the ultrarelativistic limit of electron energy, much greater than the corresponding cross section in any other geometry [see (100)]. Thus, if we adjust the measuring device so it simultaneously detects (in the plane determined by the initial electron momentum and the wave vector) the spontaneous photon in a narrow cone along the momentum of the ultrarelativistic initial electron [condition (140)] and a frequency spectrum bounded from above by the frequency $\omega'_{\max}$ given by (138), and the final electron at angles θ_f given by (139) [under the conditions (142)] or by (141) (where $\omega'=\omega'_{\max}$), we find that spontaneous bremsstrahlung with multiphoton correlated emission (absorption) of equal numbers of photons of both waves may dominate over spontaneous bremsstrahlung by an ultrarelativistic electron scattered by a nucleus in the field of both waves in the noninterference region.

We now examine the nonrelativistic limit of electron energies [the conditions (72)]. We begin with the kinematics of emission of a spontaneous photon and exit of the final electron. The condition (118) shows than in the nonrelativistic limit of electron velocities, $a_i \approx v_i \sin\theta_i \ll 1$, with the result that irrespective of the energy and incidence angle of the initial electron, the spontaneous photon exits in a narrow cone pointing opposite to the wave vector $\mathbf{k}_1$:

$$\theta' = \pi - 2v_i \sin\theta_i \approx \pi. \tag{146}$$

The interference condition (117) becomes

$$v_f \sin\theta_f = v_i \sin\theta_i. \tag{147}$$

We limit ourselves to such wave intensities that the squared velocity of electron oscillations in the field of both waves is much less than the squared velocity of the electron's translational motion (the condition (111); note that here the velocities of oscillations and translational motion of the electron may be of the same order of magnitude: $\eta_{1,2} \sim v_i$). In this case we are dealing with moderately strong fields ($\zeta_{\text{int}} = \eta_1\eta_2 v_i^{-1} \ll 1$). Hence we can ignore the energies of emitted and absorbed photons of combination frequencies, and the law of energy conservation in (104) assumes the standard form (90), which yields

$$v_f = v_i\sqrt{1-\frac{2\omega'}{mv_i^2}}. \tag{148}$$

Hence, we find from (117) that the scattering angles of the electron obey the relation

$$\sin\theta_f = \sin\theta_i \Big/ \sqrt{1-\frac{2\omega'}{mv_i^2}}. \tag{149}$$

Bearing in mind that the left-hand side of Eq. (149) cannot be greater than unity, we obtain an upper bound on the frequency of the spontaneous photon:

$$\omega'_{\max} = \frac{1}{2}mv_i^2\cos^2\theta_i. \tag{150}$$

Note that a similar result follows from the nonrelativistic limit (124). The dependence of the exit angle of the final electron on the frequency of the spontaneous photon at given incidence angles of the initial electron [see (149)] is similar to such a dependence in the relativistic case (see Fig. 6). What is important is that the critical incidence angle of the initial electron (at this angle $\omega'_{\max}=0$) is $\theta_{i*}=\pi/2$, so that as noted earlier for relativistic energies, we must select the incidence angle of the initial electron far from the critical angles, i.e., when $\omega' \sim mv_i^2/2$. If the spontaneous photon is emitted at the maximum frequency (for any incidence angle of the initial electron), the electrons exit at an angle $\theta_f=\pi/2$ [this follows from (149)].

The expression for the differential probability in the given case can be found from (134)–(136):

$$d\sigma^{(0)}_{l_+l_-} = |J_{l_+l_-}(\beta_1,\beta_2;\chi_+,\chi_-,\alpha_+,\alpha_-)|^2 d\sigma_*^{v\ll 1}, \tag{151}$$

where $d\sigma_*^{v\ll 1}$ is the differential cross section of spontaneous bremsstrahlung by nonrelativistic electrons in a vanishing field,[22] which amounts to emission of a photon of given frequency (with an upper bound $\omega'_{\max}$) and direction [see (146)] and is accompanied by exit of the final electron in the given direction [see (149)]. Here the functions $J_{l_+l_-}$ are given by (46) and their arguments are

$$\beta_j=\frac{(1-\delta_j^2)\,\eta_j^2 m v_i}{8\omega_j}\left[\sqrt{1-\frac{2\omega'}{m v_i^2}}\cos\theta_f-\cos\theta_i\right],\quad j=1,2, \tag{152}$$

$$\alpha_{\pm}=\frac{\eta_1\eta_2|d_{\mp}|m v_i}{2(\omega_1\pm\omega_2)}\left[\cos\theta_i-\sqrt{1-\frac{2\omega'}{m v_i^2}}\cos\theta_f\right]. \tag{153}$$

When both waves are circularly polarized ($\delta_1^2=\delta_2^2=1$), the quantum parameters $\beta_{1,2}$ vanish and the expression (151) for the differential cross section becomes much simpler:

$$d\sigma_{l_{\pm}}^{(0)}=J_{l_{\pm}}^2(\alpha_{\pm})\,d\sigma_*^{v\ll 1}. \tag{154}$$

From (151) and (154) it follows that for field intensities given by (111) the differential cross section of spontaneous bremsstrahlung becomes the product of the probability of correlated emission (absorption) of equal numbers of photons of both waves by the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing external field. Here the differential cross sections can easily be summed over all processes of emission and absorption of photons of the waves. As a result, the summing balances the important quantum contributions perfectly [see (71)]. We also note that for wave intensities satisfying the inequalities (113), $\beta_{1,2}\gtrsim 1$ and $\alpha_{\pm}\gtrsim 1$, with the result that $\gamma_{1,2}\gg\alpha_{\pm}\gtrsim 1$ and $\gamma_{1,2}\gg\beta_{1,2}\gtrsim 1$ [see (61) and (106)]. Hence, for wave intensities satisfying the conditions (113) and (72), we can easily show [see (100)] that the nonrelativistic limit of the differential cross section (68) in the noninterference region is much smaller than the corresponding cross section (151) in the interference region. Thus, if we adjust the measuring device so it simultaneously detects (in the plane determined by the initial electron momentum and the wave vector) the spontaneous photon in a narrow cone opposite to the wave vector $\mathbf{k}_1$ [the condition (146)] and a frequency spectrum bounded from above by the frequency $\omega'_{\max}$ given by (150), and the final electron at angles (149) or $\theta_f=\pi/2$ (when $\omega'=\omega'_{\max}$), we find that spontaneous bremsstrahlung with multiphoton correlated emission (absorption) of equal numbers of photons of both waves can dominate spontaneous bremsstrahlung by a nonrelativistic electron scattered by a nucleus in the field of both waves in the noninterference region.

7. CONCLUSION

We have studied spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two light waves propagating in the same direction. The following results have been achieved.

1. Spontaneous bremsstrahlung by an electron scattered by a nucleus in the field of two waves depends heavily on the kinetics of the scattered electron and the spontaneously emitted photon. Hence, two kinematic regions can be identified: the noninterference region (if the conditions (43) are not met), in which the main multiphoton parameters are the Bunkin–Fedorov quantum parameters (24), and the interference region (if the conditions (43) hold) where by the quantum interference parameters $\alpha_{\pm}$ [Eq. (26)] act as multiphoton parameters.

2. In the noninterference region with moderately strong fields [$\xi_{1,2}\ll 1$; see (63)], the differential cross section is the product of the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing field and the probability of emission (absorption) of definite numbers of photons of the first and second waves [see (68)]. For the nonrelativistic limit of electron energies we have found the condition for the applicability of the dipole approximation [see (88)], which differs from the one commonly used in such problems.

3. We have discovered a spontaneous interference bremsstrahlung effect for equal linear polarizations of both waves. The effect occurs in the scattering of the electron and emission of a spontaneous photon in the plane perpendicular to polarization vector of the waves. In this case the electron, in the course of deceleration by the nucleus and spontaneous emission of a photon, emits (absorbs) equal numbers of photons of both waves in a correlated manner. In moderately strong fields [$\zeta_{\rm int}\ll 1$; see (94)], the differential cross section is the product of the cross section of spontaneous bremsstrahlung by an electron scattered by a nucleus in a vanishing field and the probability of correlated emission (absorption) of equal numbers of photons of both waves [Eq. (98)] [which is given by the square of the absolute value of the functions $J_{l_+l_-}$ in (46)]. The latter factor cannot be represented as a product of probabilities of emission (absorption) of photons of each of the waves separately, as it can in the noninterference region. For wave intensities satisfying (63) and (99) (in the relativistic case) or (72) and (113) (in the nonrelativistic case), the differential cross section in the interference region is much greater than the corresponding cross section in the noninterference region.

4. We have also discovered a spontaneous interference bremsstrahlung effect in the general case (for elliptical polarization of both waves), which occurs in the scattering of an electron and emission of a spontaneous photon in the plane specified by the initial electron momentum and the wave vector at given angles [see (117) and (118)]. Here, due to interference, the electron spontaneously emits equal numbers of correlated photons of both waves. The given effect has been analyzed in moderately strong fields ($\zeta_{\rm int}\ll 1$). We have also found that the frequency of the spontaneous photon has an upper bound, $\omega'_{\max}$ [see (124)], which depends on the energy and polar incidence angle of the initial electron. The exit angles of the electron and spontaneous photon are highly correlated and depend heavily on the energy of the initial electron [see (133)]. Finally, we have shown that the differential cross sections with correlated emission (absorption) of equal numbers of photons of both waves, with emission of a spontaneous photon of given energy and in a given direction, and with the scattering of the electron to a given angle (correlated with the exit angle of the spontaneous photon), can be much greater than the corresponding cross section in any other geometry.

*)E-mail: rosh_sp@ssu.sumy.ua

[1] I. V. Lebedev, Opt. Spektrosk. **32**, 120 (1972) [Opt. Spectrosc. **32**, 59 (1972)].
[2] A. V. Borisov and V. Ch. Zhukovskiĭ, Zh. Éksp. Teor. Fiz. **70**, 477 (1976) [Sov. Phys. JETP **43**, 247 (1976)].
[3] R. V. Karapetyan and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **75**, 816 (1978) [Sov. Phys. JETP **48**, 412 (1978)].
[4] A. V. Borisov, V. Ch. Zhukovskiĭ, and P. A. Éminov, Zh. Éksp. Teor. Fiz. **78**, 530 (1980) [Sov. Phys. JETP **51**, 267 (1980)].
[5] V. P. Kraĭnov and S. P. Roshchupkin, Zh. Éksp. Teor. Fiz. **84**, 1302 (1983) [Sov. Phys. JETP **57**, 754 (1983)].
[6] S. P. Roshchupkin, Izv. Vyssh. Uchebn. Zaved. Fiz. **No. 26**, 18 (1983).
[7] S. P. Roshchupkin, Yad. Fiz. **41**, 1244 (1985) [Sov. J. Nucl. Phys. **41**, 796 (1985)].
[8] S. P. Roshchupkin, Laser Phys. **5**, 837 (1996).
[9] R. V. Karapetyan and M. V. Fedorov, Kvant. Elektron. (Moscow) **4**, 2203 (1977) [Sov. J. Quantum Electron. **7**, 1260 (1977)].
[10] R. L. Gorodinskii and S. P. Roshchupkin, Laser Phys. **2**, 602 (1992).
[11] S. P. Roshchupkin, Zh. Éksp. Teor. Fiz. **106**, 102 (1994) [JETP **79**, 54 (1994)].
[12] S. P. Roshchupkin, Zh. Éksp. Teor. Fiz. **109**, 337 (1996) [JETP **82**, 177 (1996)].
[13] S. P. Roshchupkin and A. I. Voroshilo, Laser Phys. **7**, 873 (1997).
[14] O. I. Denisenko and S. P. Roshchupkin, Phys. Scr. **50**, 339 (1994).
[15] I. M. Ternov, V. G. Bagrov, and Yu. I. Klimenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **No. 2**, 50 (1968).
[16] I. V. Lebedev, Opt. Spektrosk. **29**, 948 (1970).
[17] Yu. I. Klimenko, V. V. Kulish, and A. I. Khudom'yasov, Izv. Vyssh. Uchebn. Zaved. Fiz. **No. 11**, 12 (1974).
[18] Yu. I. Klimenko, V. V. Kulish, I. I. Fedosov, and A. I. Khudom'yasov, Izv. Vyssh. Uchebn. Zaved. Fiz. **No. 4**, 136 (1974).
[19] V. I. Radionov, Zh. Éksp. Teor. Fiz. **81**, 187 (1981) [*sic*].
[20] A. I. Voroshilo and S. P. Roshchupkin, Laser Phys. **7**, 466 (1997).
[21] S. P. Roshchupkin and A. I. Voroshilo, Visnik Sumy State Univ. **1**(7), 126 (1997).
[22] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[23] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[24] L. S. Brown and T. W. B. Kibble, Phys. Rev. **133**, A705 (1964).
[25] H. R. Reiss and J. H. Eberly, Phys. Rev. **151**, 1058 (1966).
[26] V. P. Yakovlev, Zh. Éksp. Teor. Fiz. **51**, 617 (1966)[*sic*].
[27] F. V. Bunkin and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **49**, 1215 (1965) [Sov. Phys. JETP **22**, 844 (1966)].
[28] M. M. Denisov and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **53**, 1340 (1967) [Sov. Phys. JETP **26**, 779 (1968)].
[29] F. V. Bunkin, A. E. Kazakov, and M. V. Fedorov, Usp. Fiz. Nauk **107**, 559 (1972) [Sov. Phys. Usp. **15**, 416 (1972)].
[30] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).

Translated by Eugene Yankovsky

Resonances and dichroism in the scattering of an electron in an intense laser field by the Coulomb potential

L. P. Rapoport and A. S. Kornev*)

Voronezh State University, 394693 Voronezh, Russia
(Submitted 22 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1241–1249 (October 1999)

We use a new type of Hamiltonian representing the electron–proton interaction in an intense laser field for an *ab initio* calculation of the differential scattering cross section. We give the diagrams of the results of calculations of the shapes and widths of the resonances in the cross section of electron scattering by the ''field-dressed'' Coulomb potential. The resonances emerge because of re-emission of photons by the electron. We also give the angular distribution of the scattered electrons as a function of circular dichroism for different values of the laser field strength and frequency. © *1999 American Institute of Physics.* [S1063-7761(99)00810-0]

1. INTRODUCTION

The theoretical and experimental investigation of the scattering of electrons by atoms in the presence of an intense laser field has attracted much attention, since it introduces new parameters into the scattering process such as the photon energy $\hbar\omega$, the field intensity I, the field's polarization, and statistics (see the reviews in Refs. 1–3 and the literature cited therein). However, the calculation of the electron–atom scattering cross sections in the presence of an intense radiation field is extremely complicated and can be done easily only in the Born approximation. The key issue of more accurate calculations is the formula for the cross section of electron scattering in an intense field by the Coulomb potential, which in the absence of a light field is given by an exact analytic formula. An exact solution for the cross sections of photoionization and scattering in a circularly polarized field is known for a short-range delta-function potential.[4,5]

The differential cross section of elastic and inelastic electron scattering by the Coulomb potential in the presence of a strong circularly polarized electromagnetic field were considered by Dimou and Faisal.[6] Since the variables in the corresponding time-dependent Schrödinger equation do not separate, the problem was solved by direct numerical integration of the system of equations for the strongly coupled channels.

In the present paper we solve the same problem by using a new representation (via the unitary transformations method) for the Hamiltonian of the interaction between intense circularly polarized light and an electron in the Coulomb field. This representation is in the form of a multipole expansion with the field parameter $a_0 = eF/\mu\omega^2$ (F is the field strength, and ω is the field's frequency) taken into account.[7] In this case the system of equations of the method of strongly coupled channels can be analytically integrated with respect to the angles θ and ϕ, which results in equations that are one-dimensional. Such equations can be conveniently integrated by numerical methods.

We study the resonances (shape and position) that emerge in the differential scattering cross section as a function of the electron energy. Due to the coupling between elastic and inelastic channels, there emerges an interference term caused by energy dissipation, which leads to dichroism in the position and shape of the resonances in the scattering cross section as a function of the electron energy. We build the angular distribution of the scattered electrons for left- and right-hand circularly polarized light. We also show that in a strong field, depending on the number of photons involved in the process, the angular distribution of the scattered electrons differs (not only qualitatively but also quantitatively) from the effect of dichroism on the angular distribution of scattering calculated by perturbation-theory techniques.[8] Finally, we discuss the effect of the field strength F on the position and shape of the resonances.

2. UNITARY TRANSFORMATIONS OF THE INTERACTION HAMILTONIAN OF AN ELECTRON IN AN INTENSE LASER FIELD SCATTERED BY THE COULOMB POTENTIAL

Let us consider the time-dependent Schrödinger equation in a light field described by a vector potential $\mathbf{A}(t)$ that depends only on time (the dipole approximation; everywhere we use the system of units in which $\hbar = e = m = 1$):

$$i\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = \left[-\frac{1}{2}\nabla^2 - \frac{\mathbf{A}(t)}{2}\cdot\hat{\mathbf{p}} + \frac{1}{2c^2}\mathbf{A}^2(t) - \frac{Z}{r}\right]\Psi(\mathbf{r},t). \tag{1}$$

We assume that the field is circularly polarized:

$$\mathbf{A}(t) = -A_0(\mathbf{e}_x \sin\omega t + \eta\mathbf{e}_y \cos\omega t), \tag{2}$$

where $A_0 = Fc/\omega$, with c the speed of light, and $\eta = \pm 1$ for left- and right-hand circular polarizations, with F the field strength, and ω the frequency of the field. The wave vector is directed along the z axis.

We write the solution of Eq. (1) in the form $\Phi = \hat{U}(t)\Psi$, where $\hat{U}^\dagger = \hat{U}^{-1}$ is a unitary operator. Then for Φ we obtain an equation of type (1) with the operator

$$\hat{H}' = i\left(\frac{\partial \hat{U}}{\partial t}\right)\hat{U}^{\dagger} + \hat{U}\hat{H}\hat{U}^{\dagger}, \tag{3}$$

where $\hat{H}$ is the operator of the right-hand side of Eq. (1).

We transform the Hamiltonian $\hat{H}$ via two unitary operators applied successively:[7] $\hat{U}_{\text{K–H}}$ (see Refs. 9 and 10) and $\hat{U}_{\text{rot}}$. The analytic expressions for these operators are

$$\hat{U}_{\text{K–H}} = \exp\left\{ i\int^{t} dt' \left[\frac{\mathbf{A}(t')}{c}\cdot\hat{\mathbf{p}} - \frac{1}{2c^2}\mathbf{A}^2(t') \right] \right\} \tag{4}$$

and

$$\hat{U}_{\text{rot}} = \exp\{-i\eta\omega t\hat{L}_z\}, \tag{5}$$

where $\hat{L}_z = -i\partial/\partial\phi$ is the operator of the projection of angular momentum in a spherical system of coordinates.

Applying the first transformation, we obtain the Hamiltonian (1) in a reference frame that oscillates at the frequency of the field:

$$\hat{H}_{\text{vib}} = \hat{U}_{\text{K–H}}\hat{H}\hat{U}^{\dagger}_{\text{K–H}} = -\frac{1}{2}\nabla^2 - \frac{Z}{|\mathbf{r}-\mathbf{a}(t)|}, \tag{6}$$

where $\mathbf{a}(t) = c^{-1}\int^t dt'\,\mathbf{A}(t')$.

The second transformation by the operator (5) transforms the new Hamiltonian into a rotating reference frame:

$$\hat{H}_{\text{rot}} = \hat{U}_{\text{rot}}\hat{H}_{\text{vib}}\hat{U}^{\dagger}_{\text{rot}} = -\frac{1}{2}\nabla^2 - \frac{Z}{|\mathbf{r}-\mathbf{a}_0|} + \eta\omega\hat{L}_z, \tag{7}$$

where $\mathbf{a}_0 = a_0\mathbf{e}_x$, with $a_0 = F/\omega^2$.

The operator (7) does not depend on time, so that Eq. (1) with the operator $\hat{H}_{\text{rot}}$ given by formula (7) becomes a time-independent Schrödinger equation with an exact quasienergy E. What is important is that $\hat{H}_{\text{rot}}$ has the same asymptotic behavior as $r\to\infty$ as the Hamiltonian of Eq. (1) at $\mathbf{A}(t) = 0$. The field parameter a_0 determines the dynamics of the electron in the field and the quasienergy spectrum in the system consisting of the field and the atom.

In the operator $\hat{H}_{\text{rot}}$ [Eq. (7)], the potential $Z|\mathbf{r}-\mathbf{a}_0|^{-1}$ is the generating function of the Legendre polynomials. If we introduce the unnormalized spherical harmonics $C_{LM}(\hat{\mathbf{r}})$ and $C_{LM}(\hat{\mathbf{a}}_0)$ (these differ from the functions Y_{LM} by the factor $\sqrt{4\pi/(2L+1)}$; $\hat{\mathbf{r}} = \mathbf{r}/r$ and $\hat{\mathbf{a}}_0 = \mathbf{a}_0/a_0$), we can write this function in the form of a series:

$$\frac{Z}{|\mathbf{r}-\mathbf{a}_0|} = Z\sum_{L=0}^{\infty}\sum_{M=-L}^{L}\xi_L(r,a_0)C_{LM}(\hat{\mathbf{r}})C^{*}_{LM}(\hat{\mathbf{a}}_0), \tag{8}$$

where

$$\xi_L(r,a_0) = \frac{r_<^L}{r_>^{L+1}}, \quad r_< = \min(r,a_0), \quad r_> = \max(r,a_0).$$

Now we employ the inverse unitary transformation $\hat{U}^{-1}_{\text{rot}} = \hat{U}^{\dagger}_{\text{rot}}$ [see Eq. (5)] and return to the oscillating reference frame. We obtain

$$\hat{H}^{(\eta)}_{\text{vib}}(\mathbf{r},\mathbf{a}_0;t) = -\frac{1}{2}\nabla^2 - Z\xi_0(r,a_0) - Z\sum_{L=1}^{\infty}\sum_{M=-L}^{L}\xi_L(r,a_0)C_{LM}(\hat{\mathbf{r}})C^{*}_{LM}(\hat{\mathbf{a}}_0)\times\exp\{i\eta M\omega t\}. \tag{9}$$

In (9) we wrote the term with the centrally symmetric part of the interaction potential ($L=0$) explicitly, i.e., the potential

$$\xi_0(r,a_0) = \begin{cases} \dfrac{1}{a_0}, & r < a_0, \\ \dfrac{1}{r}, & r > a_0, \end{cases}$$

which describes a ''field-dressed'' proton. The part of the sum in (9) with $M=0$ corresponds to the time-independent noncentrally symmetric potential. The remaining part of the potential (with $|M|\geqslant 1$) is a multipole expansion that harmonically depends on time.

3. SOLUTION OF THE SCHRÖDINGER EQUATION BY THE METHOD OF STRONG CHANNEL COUPLING

The time-dependent Schrödinger equation that we must solve with the Hamiltonian (9) has the form

$$i\frac{\partial\Phi(\mathbf{r},t)}{\partial t} = \left\{ -\frac{1}{2}\nabla^2 - Z\xi_0(r,a_0) - Z\sum_{L=1}^{\infty}\sum_{M=-L}^{L}\xi_L(r,a_0)C_{LM}(\hat{\mathbf{r}})C^{*}_{LM}(\hat{\mathbf{a}}_0)\times\exp\{i\eta M\omega t\}\right\}\Phi(\mathbf{r},t). \tag{10}$$

We seek a solution of Eq. (10) in the form of a quasienergy and partial-wave expansion:

$$\Phi(\mathbf{r},t) = \sum_{n=-\infty}^{\infty}\sum_{l'm'}\frac{\exp\{-iEt+in\omega t\}}{r}F_{nl'm'}(r)Y_{l'm'}(\hat{\mathbf{r}}). \tag{11}$$

Substituting (11) into (10), we obtain

$$\sum_{n=-\infty}^{\infty}\sum_{l'm'}(E-n\omega)\exp\{in\omega t\}F_{nl'm'}(r)Y_{l'm'}(\hat{\mathbf{r}}) = \sum_{n=-\infty}^{\infty}\sum_{l'm'}\left\{ -\frac{1}{2}\frac{d^2}{dr^2} + \frac{l'(l'+1)}{2r^2} - Z\xi_0(r,a_0) - Z\sum_{L=1}^{\infty}\sum_{M=-L}^{L}\xi_L(r,a_0)C_{LM}(\hat{\mathbf{r}})C^{*}_{LM}(\hat{\mathbf{a}}_0)\exp\{i\eta M\omega t\}\right\}\times\exp\{in\omega t\}F_{nl'm'}(r)Y_{l'm'}(\hat{\mathbf{r}}). \tag{12}$$

Since the right-hand side of Eq. (12) is an expansion in spherical harmonics, we can integrate over angles and reduce (12) to a one-dimensional system of equations. Multiplying

(12) into $Y^*_{lm}(\hat{\mathbf{r}})$, integrating over the solid angle $d\Omega$, and equating the coefficients of the like factors $\exp\{in\omega t\}$, we obtain a system of equations for $F_{nlm}(r)$:

$$\left\{\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}+2Z\xi_0(r,a_0)+k_n^2\right\}F_{nlm}(r)$$

$$=2\sum_{n'l'm'}V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)F_{n'l'm'}(r), \tag{13}$$

where

$$V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)\equiv -Z\sum_{L=1}^{\infty}\xi_L(r,a_0)(-1)^m[(2l+1)(2l'+1)]^{1/2}C^*_{L,m-m'}(\hat{\mathbf{a}}_0)\begin{pmatrix} l' & L & l\\ 0 & 0 & 0\end{pmatrix}\times\begin{pmatrix} l' & L & l\\ m' & m-m' & -m\end{pmatrix}\delta_{n',n+\eta(m'-m)}, \tag{14}$$

$$k_n^2=2(E-n\omega),$$

$$C^*_{LM}(\hat{\mathbf{a}}_0)=C^*_{LM}\left(\frac{1}{2}\pi,0\right)=\begin{cases}(-1)^{(L+M)/2}\Xi(L,M) & \text{if } L+M \text{ is even},\\ 0 & \text{if } L+M \text{ is odd},\end{cases}$$

$$\Xi(L,M)=\sqrt{\frac{(L-M-1)!!}{(L-M)!!}\,\frac{(L+M-1)!!}{(L+M)!!}}$$

(see Ref. 11),

with $\left(\begin{smallmatrix} l_1 & l_2 & l_3\\ m_1 & m_2 & m_3\end{smallmatrix}\right)$ the $3jm$-symbol. The radial dependence of the channel coupling ''potential'' $V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)$ for several values of n, l, and m is depicted in Fig. 1.

We see that the right-hand side of Eq. (13) mixes the different channels, each of which is characterized by the number n of photons of polarization η absorbed ($n<0$) or emitted ($n>0$) by an electron with orbital angular momentum L and its projection M on the z axis. The ''potentials'' $V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)$ given by Eq. (14) consist of a finite number of terms for each set (n,l,m) and (n',l',m'), are expressed in terms of $3jm$-symbols, and therefore can be calculated exactly. Figure 1 shows that the $V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)$ are alternating-sign quantities, a property that significantly improves the convergence of the sum on the right-hand side of Eq. (13). In what follows we denote the set of quantum numbers (n,l,m) by a single channel symbol i, and by n_i we mean the value of n in channel i, etc.

It is convenient to seek a solution of the system of equations (13) in the form of a combination of linearly independent solutions in a limited number of channels (to a given accuracy). Let $F^{(j)}_i(r)$ be the jth solution in the ith channel. If $k_i^2>0$, the ith channel is open, but if $k_i^2<0$, the ith channel is closed. When selecting the boundary conditions for solutions of Eq. (13), one must bear in mind that $V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)\to 0$ as $r\to 0$ and $r\to\infty$, and the channels cease to be coupled. Thus, the asymptotic behavior of the wave function of open channels is given by the expression

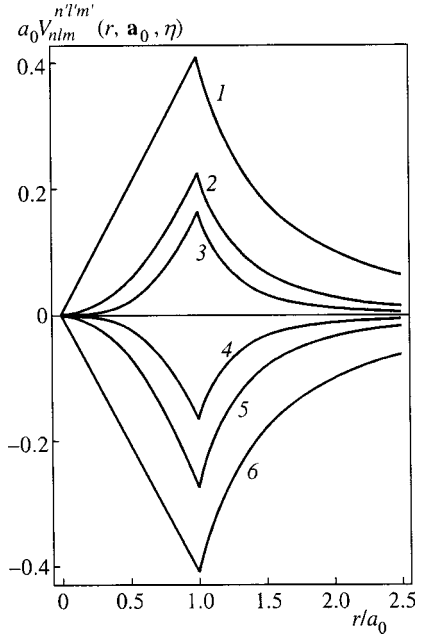


FIG. 1. Curves representing the radial dependence of $a_0V^{n'l'm'}_{nlm}(r,\mathbf{a}_0,\eta)$ for a proton at $n=l=m=0$ and $\eta=-1$. Curve *1*: $n'=-1$ and $l'=m'=1$; curve *2*: $n'=0$, $l'=2$, and $m'=0$; curve *3*: $n'=1$, $l'=3$, and $m'=-1$; curve *4*: $n'=-1$, $l'=3$, and $m'=1$; curve *5*: $n'=\mp 2$, $l'=2$, and $m'=\pm 2$; and curve *6*: $n'=1$, $l'=1$, and $m'=-1$.

$$F^{(j)}_i(r)|_{r\to\infty}=\frac{\delta_{ij}F_{l_i}(-Z/k_i,k_ir)+K^{(\eta)}_{ij}G_{l_i}(-Z/k_i,k_ir)}{\sqrt{k_i}},$$

$$i,j=1,\ldots,n_{\text{open}}, \tag{15}$$

where F and G are the Coulomb wave functions with the asymptotic behavior

$$F_{l_i}(r)|_{r\to\infty}\propto\sin\left(k_ir-\frac{\pi l_i}{2}+\frac{Z}{k_i}\ln(2k_ir)+\sigma_{l_i}\right),$$

$$G_{l_i}(r)|_{r\to\infty}\propto\cos\left(k_ir-\frac{\pi l_i}{2}+\frac{Z}{k_i}\ln(2k_ir)+\sigma_{l_i}\right), \tag{16}$$

$\sigma_{l_i}=\arg\Gamma(l_i+1-iZ/k_i)$ is the Coulomb scattering phase; n_{open} is the number of open channels involved; and $K=\{K_{ij}\}$ is the real-valued reaction K matrix, related to the complex-valued scattering S matrix by the formula

$$S_{ij}=[(I+iK)(I-iK)^{-1}]_{ij},$$

with I the identity matrix.

The boundary conditions selected for closed channels ($k_i^2<0$) are represented by exponentially decaying functions as $r\to\infty$:

$$F^{(j)}_i(r)|_{r\to\infty}=\delta_{ij}\exp\{-|k_i|r\},$$

$$i=1,\ldots,n_{\text{tot}},\quad j=n_{\text{open}}+1,\ldots,n_{\text{tot}}, \tag{17}$$

where n_{tot} is the total number of channels involved in the process.

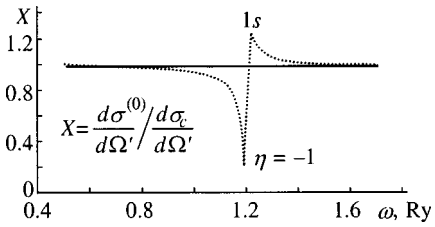


FIG. 2. Differential elastic electron–proton scattering cross section in a laser field, referred to the Rutherford scattering cross section, at different values of frequency ω and a fixed value of the field parameter, $a_0=0.1795$ a.u. The direction of the momentum of the incoming electron is specified by the angles $\theta_0=\pi/2$ and $\varphi_0=0$ and that of the scattered electron, by the angles $\theta'=\varphi'=\pi/2$. The electron energy E is 0.2 Ry. A resonance is observed at the $1s$ level in the field of the ''dressed'' atom ($E_{1s}=-0.9688$ Ry).

Near zero the boundary conditions for a state in a centrifugal potential are

$$F_i^{(j)}(r)|_{r\to 0}=\delta_{ij}r^{l_i+1}, \quad i,j=1,\ldots,n_{\rm tot}. \tag{18}$$

Knowing the S matrix, we can find the scattering amplitude. We denote the momentum of an incoming electron by $\mathbf{k}_0$ and that of an electron whose scattering is accompanied by an exchange of N photons, by $\mathbf{k}'_N$. Then the scattering amplitude in open channels is[12]

$$f^{(\eta)}_{0\to N}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_N)=f_C(\theta)\,\delta_{N,0}+f^{(N,\eta)}_{\rm rad}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_N), \tag{19}$$

where

$$f^{(N,\eta)}_{\rm rad}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_N)=\sum_{\substack{l_0,m_0\\ l,m}}\frac{i^{l_0-l+1}}{2\sqrt{k_0k'_N}}[(2l_0+1)(2l+1)]^{1/2}$$
$$\times\exp\{i(\sigma_{l_0}+\sigma_l)\}C^*_{l_0m_0}(\hat{\mathbf{k}}_0)$$
$$\times C_{lm}(\hat{\mathbf{k}}'_N)[\,\delta_{N,0}\delta_{ll_0}\delta_{mm_0}-S^{(\eta)}_{Nlm,0l_0m_0}], \tag{20}$$

$f_C(\theta)$ is the scattering amplitude in a purely Coulomb field (for $N=0$,)[12] $\theta=\widehat{(\mathbf{k}_0,\mathbf{k}'_0)}$ is the scattering angle, $\hat{\mathbf{k}}_0=\mathbf{k}_0/k_0$, and $\hat{\mathbf{k}}'_N=\mathbf{k}'_N/k'_N$.

The differential elastic scattering cross section for an electron ($N=0$), corresponding to $f^{(\eta)}_{0\to N}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_N)$, in an intense laser field of polarization η is given by the formula

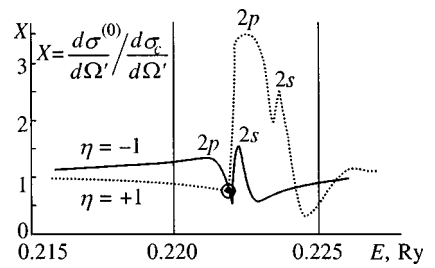


FIG. 3. Differential elastic electron–proton scattering cross section in a laser field, referred to the Rutherford scattering cross section, as a function of the electron energy at $F=0.01$ a.u. and $\omega=0.472$ Ry$=6.419$ eV ($a_0=0.1795$ a.u.). The geometry is the same as in Fig. 2. Resonances are observed at the $2s$ and $2p$ levels in the field of the ''dressed'' atom ($E_{2s}=-0.2461$ Ry and $E_{2p}=-0.2500$ Ry).

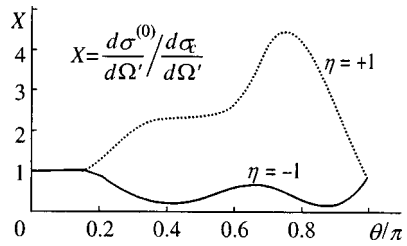


FIG. 4. Angular distribution of electrons ($E=0.2218$ Ry) elastically scattered in a plane perpendicular to the direction of the laser beam. Here θ is the scattering angle. The other parameters and notation are the same as in Fig. 3. The number of photons involved is $n=0$ and ± 1.

$$\frac{d\sigma^{(0)}}{d\Omega}=|f_C(\theta)|^2+|f^{(0,\eta)}_{\rm rad}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_0)|^2$$
$$+2\,{\rm Re}[f^*_C(\theta)f^{(0,\eta)}_{\rm rad}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_0)]. \tag{21}$$

The differential cross section of inelastic processes related to stimulated bremsstrahlung ($E>0$) or absorption ($E<0$) is given by the formula

$$\frac{d\sigma^{(N)}}{d\Omega}=\frac{k'_N}{k_0}|f^{(N,\eta)}_{\rm rad}(\hat{\mathbf{k}}_0,\hat{\mathbf{k}}'_N)|^2. \tag{22}$$

4. RESULTS OF CALCULATIONS

The numerical calculations of the cross sections of the elastic scattering of an electron by the Coulomb potential in the presence of a laser field were done for different energies E of the incident electron and different frequencies ω and strengths F of the field. The scattering plane was chosen perpendicular to the direction of light propagation. As shown by Manakov,[8] this is the optimum geometry for observing dichroism in the angular distribution of scattered electrons.

Figure 2 depicts a resonance as a function of the frequency ω for $\eta=-1$. Obviously, there can be no resonance if we do not allow for the inelastic channel.

Figures 3–5 depict the energy spectra and the angular distributions of the scattered electrons for different numbers of photons involved in the process and different values of the polarization index $\eta=\pm 1$ (circular dichroism). Figure 3 depicts the differential elastic electron–proton scattering cross section as a function of the electron energy E. Resonances are observed at certain values of E, ω, and F, and their position and shape were found to depend on the energy levels of the ''dressed atom,'' i.e., on the energy spectrum in the

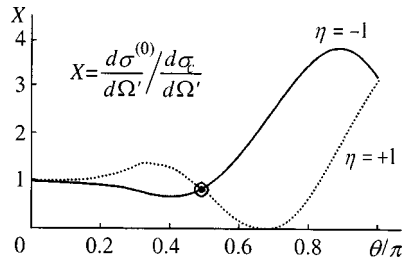


FIG. 5. Angular distribution of elastically scattered electron with the number of photons involved is $n=0$, ± 1, and ± 2. The notation and parameters are the same as in Fig. 4.

potential $-Z\xi_0(r,a_0)$, and on the direction of polarization. Here the resonances at the $2s$ and $2p$ levels are resolved, since the potential of the ''dressed'' atoms lifts the Coulomb degeneracy. Figure 3 also shows that the width and shape of the resonances at the levels $2s$ and $2p$ strongly depend on the polarization of the light.

Figures 4 and 5 depict the dependence of the differential scattering cross sections on the electron scattering angle for different polarizations. The channels that are involved in Fig. 4 are $n=0$ and ± 1. In this case the angular distribution of the electrons for $\eta=\pm 1$ differs from the similar distribution calculated by perturbation-theory techniques[8] only quantitatively (i.e., the shape of the curves is different). For the channels involved in the process with an exchange of $n=0$, ± 1, and ± 2 photons the difference is not only quantitative but also qualitative: the curves for different η intersect (see Fig. 5). The corresponding points of intersection in Figs. 3 and 5 are indicated by small circles.

As the field strength F grows (but the frequency ω remains fixed), the number of channels involved in the process increases (in our case satisfactory convergence was achieved at $l_{\max}=3$ and $|n_{\max}|=2$) and so does the resonance width. Here the ''center of gravity'' of the resonance is shifted to higher electron energies. This is quite obvious since at a fixed frequency the value of the parameter a_0 is proportional to F, and as a_0 increases the depth of the potential $-Z\xi_0(r,a_0)$ decreases, which raises the levels of the discrete spectrum. There is reason to believe that in ultrastrong fields ($a_0 \gtrsim 50$), where the bound states become crowded, no resonances will be observed.

At a fixed field strength F, the widths of the resonances of the scattering cross section as a function of the frequency ω increase with the ''depth'' of the levels.

Let us examine more thoroughly the effect of the light polarization on the cross section of various multiphoton processes. As noted by Manakov,[8] for all multiphoton processes the general structure of circular dichroism can be determined on the basis of arguments involving the space–time invariance of the process. Specifically, the dependence of the amplitude of a quantum transition on the light polarization manifests itself in all cases where there is interference, i.e., coupling between the given transition and the inelastic channels in the total amplitude of the process (channels of light energy dissipation). In our case, Eq. (21) is a clear indication of this. Here dichroism is not determined by the field strength. In particular, Manakov[8] did his calculations for a weak field in a perturbation-theory setting. It goes without saying, however, that exact studies of polarization effects in an intense laser field, studies that do not use perturbation-theory techniques, are important.

Polarization effects have yet to be studied in free–free transition of electron–atom systems in a strong field (stimulated bremsstrahlung and absorption). However, circular dichroism plays an important role in scattering processes and will certainly be detected in experiments, since, as Ref. 7 indicates, extending this theory to complicated atoms requires no serious modifications.

This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant No. 98-02-16084).

*)E-mail: kornev@tooth.vsu.ru

[1] L. Rosenberg, Adv. At. Mol. Phys. **18**, 1 (1982).
[2] M. H. Mittleman, Comments At. Mol. Phys. **11**, 91 (1982).
[3] N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 6 (1986).
[4] N. L. Manakov and L. P. Rapoport, Zh. Éksp. Teor. Fiz. **69**, 842 (1975) [Sov. Phys. JETP **42**, 430 (1976)].
[5] I. J. Berson, J. Phys. B **8**, 3078 (1975).
[6] L. Dimou and F. H. M. Faisal, Phys. Rev. Lett. **59**, 872 (1987).
[7] L. P. Rapoport, JETP Lett. **68**, 198 (1998).
[8] N. L. Manakov, in *Super-Intense Laser–Atom Physics IV (1995)*, H. G. Muller and M. V. Fedorov (eds.), Kluwer Academic, Dordrecht (1996), p. 153.
[9] H. A. Kramers, *Quantum Mechanics*, North-Holland, Amsterdam (1956).
[10] W. C. Henneberger, Phys. Rev. Lett. **21**, 838 (1968).
[11] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[12] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions*, 3rd ed., Clarendon Press, Oxford (1965)

Translated by Eugene Yankovsky

Photoinduced chirality of hydrogen peroxide molecules

B. A. Grishanin*) and V. N. Zadkov

Institute of Physics and International Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 20 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1250–1263 (October 1999)

The feasibility of using nonlinear optical techniques to control the chiral states of molecules is examined with the hydrogen peroxide molecule as an example. Raman excitation of optical activity owing to a transition among states with different chiral symmetries is proposed, along with an experimental scheme for detecting the corresponding photoinduced optical rotation in hydrogen peroxide vapor. © *1999 American Institute of Physics.* [S1063-7761(99)00910-5]

1. INTRODUCTION

One of the most intriguing mysteries of nature is the chiral purity of the biological world. For example, with enviable consistency, nature chooses levorotatory helical DNA molecules and dextrorotatory sugar molecules, although the levo- and dextro-configurations are energetically equivalent.[1] Research on the chiral purity of the biological world and, as a whole, on the nature of the chiral symmetry of molecules, the construction of chiral properties and their experimental study, and attempts at controlling chirality are currently among the most pressing areas of modern physics.[2–4] In particular, the physical consequences of the existence of a universal mechanism for breaking of chiral symmetry[5,6] owing to the failure of reflection symmetry as the result of a weak interaction through neutral currents[7,8] are the subject of ongoing discussion. Besides the spontaneous breakdown of chiral symmetry, there is practical interest in research on the feasibility of deliberately controlling changes in the chiral symmetry, induced, for example, by light. If this is possible, then in optics we have at our disposal a powerful arsenal of experimental methods, both for exciting and for probing molecular systems,[9] which may be used in setting up an experiment on photoinduced chirality.

In this paper we study the feasibility of exciting chirally asymmetric states of the hydrogen peroxide molecule (H_2O_2) by means of a laser pulse with specially selected parameters. The hydrogen peroxide molecule (Fig. 1a) is the simplest chiral molecule whose geometry is not invariant under the spatial inversion $(x,y,z)\rightarrow(x,y,-z)$. After this transformation, the initial position of the atoms of the molecule cannot be recovered by a rotational transition, since the torsional angles $\pm\theta_{1,2}$ are inequivalent. The choice of direction of the vector $\mathbf{n}_O$ from one oxygen atom of the molecule to the other makes it possible to assign a definite sign to the torsion angle $\angle$HOOH after choosing between the right- and left-handed coordinate systems. According to theoretical calculations and experiment,[10–13] the equilibrium torsion angle in the gaseous phase is $\theta\simeq\pm 120°$, where a positive sign corresponds to the so-called *d*-state (''dextro'' or right-handed) indicated in Fig. 1a and a negative sign, to the so-called *l*-state (''levo'' or left-handed).

Let us examine qualitatively what happens when a chiral H_2O_2 molecule interacts with an electromagnetic field. The existence of reflection symmetry between the minima of the torsional potential means that the eigenstates of the torsional Hamiltonian are split as a result of tunneling through the lower barrier and are described by even and odd wave functions ψ_S, ψ_A (tunneling through the upper barrier and the additional splitting owing to it are negligible[13]). The relatively large splitting $\Delta E_0=11.4\ \mathrm{cm}^{-1}$ corresponds to a lack of stationary chirally asymmetric states, which, prior to the establishment of equilibrium, oscillate at the tunneling frequency relative to the stationary states ψ_S and ψ_A. Thus, in the H_2O_2 molecule, in principle it is possible only to obtain an oscillatory optical rotation effect, as opposed to heavy molecules, for which the period of these oscillations can correspond to arbitrarily long times and the initial chirally asymmetric state is stable.

The Hamiltonians for the interaction of the molecule with an electromagnetic field

$$\hat{H}_I=-\sum \mathbf{d}_k\mathbf{E}(\mathbf{r}_k)$$

in the dipole (H_D) and quadrupole (H_Q) approximations have a qualitative difference owing to the fact that in the dipole approximation [for $\mathbf{E}(\mathbf{r}_k)\rightarrow\mathbf{E}(\mathbf{r}_0)$] the contributions of the protons add, while in the quadrupole approximation $[\mathbf{E}(\mathbf{r}_k)\rightarrow(\mathbf{r}_k-\mathbf{r}_0)\nabla\mathbf{E}(\mathbf{r}_0)]$ they subtract, so that H_D is an even function, while H_Q is odd. For the corresponding off-diagonal matrix elements H_D^{12} and H_Q^{12} of the $S\rightarrow A$ transition for the dipole and quadrupole Hamiltonians, we obtain

$$H_D^{12}=0,\quad H_Q^{12}\neq 0. \tag{1}$$

This implies that the quadrupole interaction excites precession between states of a chiral molecule that are split owing to the $d\leftrightarrow l$ tunneling transition between the right- and left-handed chiral configurations, while the dipole interaction only produces a modulation in the energy of the eigenstates, which therefore oscillates at the frequency of the exciting field. The existence of the quadrupole contribution offers the

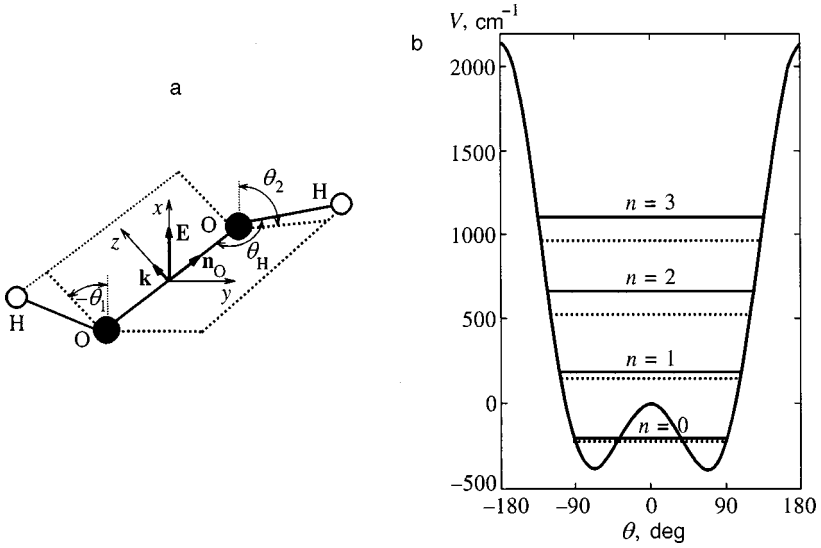


FIG. 1. (a) Geometry of an H_2O_2 molecule in the *d*-configuration. The wave vector **k** of the laser field **E** is directed along the *z* axis, $\mathbf{n}_O$ is a vector in the direction of the O–O bond, $\theta_{1,2}$ are the torsion angles relative to the $x\mathbf{n}_O$ plane, the equilibrium valence angles are $\theta_H \approx 100°$, and $a_O \approx 1.461$ Å and $a_H = 0.964$ Å are the lengths of the O–O and O–H bonds, respectively. (b) Model function for the torsional potential *V* as a function of the torsion angle θ and the position of the lowest energy levels for the symmetric (dots) and antisymmetric (continuous line) states.

opportunity, in principle, to selectively excite the *d*- or *l*-enantiomers using a fundamental quantum optical effect—the coherent precession of a two-level system driven by a coherent electromagnetic field pulse.

In this paper we discuss the conditions under which this photoinduced chirality effect (optical activity) can be detected experimentally. The excitation mechanism is first discussed using a single-photon model, and then a more realistic scheme of two-photon excitation using two lasers with orthogonal polarizations is proposed. As the following analysis shows, if a molecule is initially oriented along the *z* axis, then in principle a laser pulse of some duration can be used to bring it from an initial *S*- or *A*-state uniformly distributed across the *d*- and *l*-configurations, and which has no optical activity, into a state with a definite configuration and a sign corresponding to rotation of the plane of polarization of the incident field. This photoinduced chirality effect can be observed experimentally in its purest form in hydrogen peroxide vapor, since the interpretation of data for the liquid phase is made more complicated by the strong intermolecular interaction.

Let us estimate the order of magnitude of the effect in a single-photon excitation scheme. The condition for complete orientation of the dipole moment of the molecules along the *z* axis under the influence of a field of strength *E* is $Eea_H \gg kT$, which corresponds to $E \gg 10^6$ V/cm at room temperature. If we do not require 100% orientation of the molecules, then much lower fields can be used to obtain a substantially lower degree of orientation, $\varkappa = Eea_H/kT$, which can nevertheless be used to establish experimentally the presence of one or the other chiral configuration of the molecules by measuring the optical activity of the medium. In the frequency range of the electronic susceptibility, it can be estimated to lowest order by assuming that $\varkappa$ describes the fraction of the molecules strictly oriented in a given direction, while the remainder of the molecules is not oriented along the field. Then, for an upper bound estimate, on multiplying this small factor $\varkappa$ by the characteristic magnitude of the specific rotation $[\alpha] \sim 10^2$ deg·cm^3/g·dm for materials with the most distinct optical activity, we obtain $\sim 10^{-1}$ deg·cm^3/g·dm for resonant excitation of optical activity at $E \approx 10^3$ V/cm. A 1-mm thick layer of the vapor at standard temperature and pressure yields a rotation angle of the order of 10^{-1} sec. This value is at the limit of sensitivity for linear polarization spectroscopic techniques. Thus, even neglecting the other complicating factors in a single-photon excitation scheme with initial orientation of the molecules by a constant electric field, observing the laser induced optical rotation will be difficult.

Another important point is the finite lifetime τ_r of the chiral state owing to intermolecular collisions. It should at least be longer than the time the field propagates in the active region, $\tau_c = L/c \geqslant 10^{-10}$ s^{-1}, and a lower bound estimate is $\tau_r = (\mathcal{N} v \sigma)^{-1}$, where $\mathcal{N}$ is the concentration of the molecules, *v* is the thermal speed, and σ is the collision cross section. For atmospheric pressure and room temperature we obtain $\tau_r \sim 10^{-9}$ s$^{-1} > \tau_c$.

The biharmonic excitation scheme using two laser pulses proposed here satisfies this restriction, and furthermore can be used to efficiently induce a given chiral state in the excited molecules.

2. THEORETICAL MODEL FOR PHOTOINDUCED CHIRALITY OF H_2O_2 MOLECULES

2.1. Reduced model for the dynamics of a free molecule

For an approximate description of the dynamics of molecules in a laser radiation field, it is appropriate to simplify the complete Hamiltonian by taking advantage of the small ratio of the masses of the protons and oxygen atoms. Then the dynamics of the protons can be treated in the adiabatic approximation with respect to the coordinates of the oxygen atom, and the dynamics of the latter reduces, in the simplest case, to averaging over the direction of the unit vector $\mathbf{n}_O = (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta)$, which specifies the orientation of the O–O-bond. Here the average over the *z*-direction (Fig. 1a) can be taken considering only the rotational dynamics of the protons relative to the O–O bond, given that the potential of the proton bond depends only on the torsion angle $\theta = \theta_2 - \theta_1$.

Given these considerations, the characteristic Hamiltonian in the reduced model, which accounts for the rotation of a molecule relative to the O–O bond as a whole and the torsional vibrations, has the form

$$\hat{H}=\hat{H}_{\mathrm{H}}+\hat{H}_{\theta}, \tag{2}$$

where

$$\hat{H}_{\mathrm{H}}=-\frac{\hbar^2}{4m_{\mathrm{H}}r_{\mathrm{H}}^2}\frac{\partial^2}{\partial\tilde{\theta}^2}, \tag{3}$$

$$\hat{H}_{\theta}=-\frac{\hbar^2}{m_{\mathrm{H}}r_{\mathrm{H}}^2}\frac{\partial^2}{\partial\theta^2}+V(\theta), \tag{4}$$

where $\tilde{\theta}=(\theta_1+\theta_2)/2$, m_{H} is the proton mass, and $V(\theta)$ is the torsional potential. The vibrations of the valence angles $\angle$HOO are neglected here for simplicity.

Besides the uncertainty in the direction of $\mathbf{n}_{\mathrm{O}}$ mentioned above, the initial state of the molecule with respect to the average angle $\tilde{\theta}$ includes an uncertainty associated with the rotation of the molecule relative to that direction, which corresponds to the third rotational degree of freedom of the molecule and supplements the two angles that specify the direction of $\mathbf{n}_{\mathrm{O}}$. Here the corresponding frequencies of the transitions between the levels for free rotation about the O–O axis are

$$\omega_{n\to n+1}=7.84\cdot(n+1/2)\ \mathrm{cm}^{-1},$$

and are a factor of four less than the corresponding frequencies of free internal rotation in view of Eqs. (3) and (4), which largely determines the form of the torsional potential $U(\theta)$.

2.2. Photoexcitation of rotational degrees of freedom of the molecule

In a field

$$\mathbf{E}(t)=\mathbf{E}_1u_1(t)\cos(\omega_1t+\varphi_1)$$

with envelope $u_1(t)$ and frequency $\omega_1\gg\omega_{n\to n+1}$, the quantum features of the resulting excitation are unimportant and the response is described by the classical equation

$$J\frac{d^2\tilde{\theta}}{dt^2}=\frac{\partial}{\partial\tilde{\theta}}\mathbf{E}(t)\mathbf{d},$$

where J is the moment of inertia of the molecule. To zeroth order in the deviation, in the expression for the force on the right-hand side of the equation and for pulses that are not too short, with durations $\tau_1\gg1/\omega_1$, the response in the form of the change in the angle $\tilde{\theta}$ at frequency ω_1 is given by

$$\Delta\tilde{\theta}\simeq-\frac{\partial}{\partial\tilde{\theta}}\frac{\mathbf{E}_1(t)\mathbf{d}}{J\omega_1^2}. \tag{5}$$

This change leads to a modulation at frequency ω_1 in the quadrupole Hamiltonian acting on the internal rotation, and therefore to partial orientation of the molecule, which shows up as a loss of symmetry in the quadrupole Hamiltonian. In particular, when the O–O axis of a molecule is oriented along the z axis, i.e., $\vartheta=0$, and the field $\mathbf{E}_1$ is oriented along the y axis, the rotational symmetry reduces to a uniform distribution over the angle $\tilde{\theta}$; the quadrupole moment has a dependence of the form $\sin\tilde{\theta}$. Here the orientation shows up in Eq. (5), which takes the form

$$\Delta\tilde{\theta}\simeq\frac{E_1d}{2J\omega_1^2}u_1(t)\cos\tilde{\theta}\cos\frac{\theta}{2}\cos(\omega_1t+\varphi_1), \tag{6}$$

with $\cos\tilde{\theta}$ having the same angular dependence as the derivative of the quadrupole Hamiltonian for the field $\mathbf{E}_2\|x$, so that

$$\Delta\hat{H}_Q=(\partial\hat{H}_Q/\partial\tilde{\theta})\Delta\tilde{\theta}\propto\cos^2\tilde{\theta}.$$

An order of magnitude estimate in the visible range is

$$\Delta\tilde{\theta}\sim10^{-10}\sqrt{I_1}, \tag{7}$$

where I_1 is the intensity of the field $\mathbf{E}_1$ in W/cm^2 and the angle is in radians. This estimate reaches order unity only in fields stronger than the intra-atomic fields.

The dynamics of the torsional vibrations corresponding to the Hamiltonian $\hat{H}_{\theta}$ are largely quantum mechanical owing to the relatively small mass of the proton compared to the heavier atoms. This shows up in the tunneling between the *d*- and *l*-states of the local potential minima (Fig. 1b), which leads to the formation of superposition eigenstates ψ_S and ψ_A that are split in energy and have equally represented *d*- and *l*-configurations, as well as to a nonrigidity of the molecular configuration owing to the quantum mechanical indeterminacy of the wave functions with respect to the torsion angle.

The form of the torsional potential of the H_2O_2 molecule and the eigenenergies of the torsional potential have been studied both by *ab initio* computational techniques and by analyzing experimental spectroscopic data. Figure 1b shows the potential function and structure of the eigenlevels.[14] An estimate of the uncertainty in the local states with respect to the torsion angle using the formula for position fluctuations in the ground state of a harmonic oscillator yields

$$\sigma_{\theta}\approx[\hbar/(m_{\mathrm{H}}r_{\mathrm{H}}^2\omega_0)]^{1/2}\approx20°.$$

The Hamiltonian $\hat{\mathscr{H}}_I$ for interaction with a laser field in the dipole approximation, allowing only for the displacement of the proton charges, is

$$\hat{H}_D=-E_Lea_{\mathrm{H}}\,\mathrm{Re}\,\mathbf{e}(\mathbf{e}_1+\mathbf{e}_2), \tag{8}$$

where a_{H} is the H–O distance, $\mathbf{e}_{1,2}$ are the corresponding unit vectors for the directions of the proton bonds, $\mathbf{e}$ is the polarization vector of the laser field, and e is the proton charge. The quadrupole component of the interaction Hamiltonian depends on the choice of the coordinate center. The displacement of the center leads to an additional dipole term which, however, is small and can be discarded, since it describes the same qualitative features of the interaction as the main dipole contribution. Thus, the coordinate center must be chosen as in Fig. 1 in order to obtain the simplest form of the Hamiltonian. The corresponding expression is

$$\hat{H}_Q = -\mathbf{E}_L \mathbf{d}_Q = -\frac{\mathbf{k}\mathbf{n}_O a_O}{2} E_L e a_H \mathrm{Re}[i\mathbf{e}(\mathbf{e}_2 - \mathbf{e}_1)], \qquad (9)$$

where a_O is the O–O distance. This choice of the coordinate center on the O–O line for light propagating along that line (i.e., $\mathbf{k} \| \mathbf{n}_O$) means that the dependence of the quadrupole Hamiltonian H_Q on the torsion angle is determined solely by the odd function $\sin(\theta/2)$.

To calculate Eqs. (8) and (9), one must represent the polarization vector of the field in the form $\mathbf{e}=(e_x, e_y, 0)$, where the coefficients e_x and e_y are, in general, complex numbers, and calculate the corresponding coefficients,

$$C_{H_k} = (\mathbf{n}_x \mathbf{n}_{H_k}), \quad S_{H_k} (\mathbf{n}_y \mathbf{n}_{H_k}),$$

where $\mathbf{n}_x$ and $\mathbf{n}_y$ are the unit vectors along the x and y axes and $\mathbf{n}_{H_k}$ are the unit vectors along the O–H_k bonds ($k=1,2$). In this notation, the Hamiltonians (8) and (9) take the form

$$\hat{H}_D = -E_L e a_H \mathrm{Re}\, e^{-i\varphi_L}[e_x(C_{H_1} + C_{H_2}) + e_y(S_{H_1} + S_{H_2})],$$

$$\hat{H}_Q = -\frac{k_L a_O}{2} E_L e a_H \mathrm{Re}\, i e^{-i\varphi_L}[e_x(C_{H_2} - C_{H_1}) + e_y(S_{H_2} - S_{H_1})], \qquad (10)$$

where k_L is the modulus of the wave vector of the laser field, φ_L is the phase of the laser field, and the rotation angles $\theta_{1,2}$ of the hydrogen bonds serve as coordinate operators. The projection coefficients of the proton dipole moments in the x or y directions of the laser polarization vector $\mathbf{e}$ are calculated according to the formulas

$$C_{H_k} = \mathbf{n}_x O(\mathbf{n}^{\perp}_{OOH_1}, \delta_H) O(\mathbf{n}_O, \theta_k) O(\mathbf{n}^{\perp}_{\mathbf{n}_x \mathbf{n}_O}, \pi/2 - \angle \mathbf{n}_x \mathbf{n}_O) \mathbf{n}_x,$$

$$S_{H_k} = \mathbf{n}_y O(\mathbf{n}^{\perp}_{OOH_2}, \delta_H) O(\mathbf{n}_O, \theta_k) O(\mathbf{n}^{\perp}_{\mathbf{n}_x \mathbf{n}_O}, \pi/2 - \angle \mathbf{n}_x \mathbf{n}_O) \mathbf{n}_x$$

as the scalar product of the vectors $\mathbf{n}_{x,y}$ and the vectors obtained (a) by rotating the vector $\mathbf{n}_x$ initially by an angle $\pi/2 - \angle \mathbf{n}_x \mathbf{n}_O$ in the plane of $\mathbf{n}_x \mathbf{n}_O$ until reaching a perpendicular to the $\mathbf{n}_O$ axis, which is used as a basis axis for reading the torsion angles, (b) by rotating $\mathbf{n}_x$ around the O—O axis by the torsion angle θ_k, and (c) by a subsequent rotation of $\mathbf{n}_x$ in the OOH_k plane by an angle $\delta_H = \theta_H - \pi/2$ until the unit vector along $\mathbf{n}_{H_k}$ is obtained.

Given the nonzero value of δ_H, the analytic expressions for these coefficients end up being very cumbersome, so here we give them only for the approximation $\delta_H = 0$, i.e., for directions of the hydrogen bonds orthogonal to the O–O axis.[1)] The resulting relative error is less than $\sim 10\%$ because $\delta_H \approx 10°$ is small. The three-dimensional rotation matrix for a rotation α about the $\mathbf{n}$ axis is

$$O(\mathbf{n}, \alpha) = \begin{pmatrix} n_x^2 + n_y^2 \cos\alpha + n_z^2 \cos\alpha & n_x n_y - n_x n_y \cos\alpha - n_z \sin\alpha & n_x n_z - n_x n_z \cos\alpha + n_y \sin\alpha \\ n_x n_y - n_x n_y \cos\alpha + n_z \sin\alpha & n_y^2 + n_x^2 \cos\alpha + n_z^2 \cos\alpha & n_y n_z - n_y n_z \cos\alpha - n_x \sin\alpha \\ n_x n_z - n_x n_z \cos\alpha - n_y \sin\alpha & n_y n_z - n_y n_z \cos\alpha + n_x \sin\alpha & n_z^2 + n_x^2 \cos\alpha + n_y^2 \cos\alpha \end{pmatrix}.$$

Hence, the coefficients can finally be written in the form

$$C_{H_1} = \sqrt{1 - \sin^2\vartheta \cos^2\varphi}\, \cos\theta_1,$$

$$C_{H_2} = \sqrt{1 - \sin^2\vartheta \cos^2\varphi}\, \cos\theta_2,$$

$$S_{H_1} = (16\sqrt{\cos^2\vartheta + \sin^2\varphi \sin^2\vartheta})^{-1}[-2\sin(2\varphi - \theta_1) + \sin(2\varphi - 2\vartheta - \theta_1) - 8\sin(\vartheta - \theta_1) + \sin(2\varphi + 2\vartheta - \theta_1) - 2\sin(2\varphi + \theta_1) + \sin(2\varphi - 2\vartheta + \theta_1) + 8\sin(\vartheta + \theta_1) + \sin(2\varphi + 2\vartheta + \theta_1)],$$

$$S_{H_2} = (16\sqrt{\cos^2\vartheta + \sin^2\varphi \sin^2\vartheta})^{-1}[-2\sin(2\varphi - \theta_2) + \sin(2\varphi - 2\vartheta - \theta_2) - 8\sin(\vartheta - \theta_2) + \sin(2\varphi + 2\vartheta - \theta_2) - 2\sin(2\varphi + \theta_2) + \sin(2\varphi - 2\vartheta + \theta_2) + 8\sin(\vartheta + \theta_2) + \sin(2\varphi + 2\vartheta + \theta_2)]. \qquad (11)$$

For the sum of the coefficients that determine the dipole potential we obtain

$$C_{H_1} + C_{H_2} = C_+ \cos(\theta/2), \quad S_{H_1} + S_{H_2} = S_+ \cos(\theta/2), \qquad (12)$$

$$C_+ = 2\sqrt{1 - \sin^2\vartheta \cos^2\varphi}\, \cos\tilde{\theta},$$

$$S_+ = 2(16\sqrt{\cos^2\vartheta + 2\sin^2\varphi \sin^2\vartheta})^{-1}[-2\sin(2\varphi - \tilde{\theta}) - 2\sin(2\varphi + \tilde{\theta}) + \sin(2\varphi - \tilde{\theta} - 2\vartheta) + \sin(2\varphi + \tilde{\theta} - 2\vartheta) + 8\sin(\tilde{\theta} - \vartheta) + 8\sin(\tilde{\theta} + \vartheta) + \sin(2\varphi - \tilde{\theta} + 2\vartheta) + \sin(2\varphi + \tilde{\theta} + 2\vartheta)],$$

which depend on the torsional angle via the even function $\cos(\theta/2)$. Here $\tilde{\theta} = (\theta_1 + \theta_2)/2$ is the average rotation angle of the hydrogen bonds. For the differences of the coefficients determining the quadrupole potential, we obtain an expression that depends on the torsion angle in terms of the odd function $\sin(\theta/2)$:

$$C_{H_2} - C_{H_1} = C_- \sin\frac{\theta}{2}, \quad S_{H_2} - S_{H_1} = S_- \sin\frac{\theta}{2},$$

$$C_- = -2\sqrt{1-\sin^2\vartheta\cos^2\varphi}\sin\tilde{\theta},$$

$$S_- = (16\sqrt{\cos^2\vartheta + 2\sin^2\varphi\sin^2\vartheta})^{-1}[(2\cos(2\varphi-\tilde{\theta}) - 2\cos(2\varphi+\tilde{\theta}) - \cos(2\varphi-\tilde{\theta}-2\vartheta) + \cos(2\varphi+\tilde{\theta} - 2\vartheta) + 8\cos(\tilde{\theta}-\vartheta) + 8\cos(\tilde{\theta}+\vartheta) - \cos(2\varphi-\tilde{\theta} + 2\vartheta) + \cos(2\varphi+\tilde{\theta}+2\vartheta). \quad (13)$$

For a simplified analysis of the interaction of the rotational motion of a molecule about the O–O axis and the torsional vibrations, the numerical parameters of the torsional potential $U(\theta)$ are important. The corresponding values of the frequency splitting of the levels,[14] $\Delta E_0 = 11.44$ cm^{-1}, $\Delta E_1 = 116.34$ cm^{-1}, and $\Delta E_2 = 206.57$ cm^{-1}, exceed the characteristic free rotation frequencies of the molecule as a whole for $n \geqslant 1$; this means that it is possible to study transitions between the eigenstates of the torsional Hamiltonian $\hat{H}_\theta$ directly in terms of the classical rotational coordinate $\tilde{\theta}$.

3. PHOTOEXCITATION DYNAMICS IN A TWO-LEVEL MODEL

If it is assumed that the frequencies of the transition between the eigenstates of the torsional Hamiltonian $\hat{H}_\theta$ greatly exceed the rotation frequency of the proton bonds, then laser excitation is possible for a negligibly small deviation of the orientation angles ϑ and φ of the molecule and a small deviation of the rotational angle $\tilde{\theta}$. Here it is possible to excite the symmetric/antisymmetric states ψ_S and ψ_A into a coherent superposition $C_A\psi_A + C_S\psi_S$. For $S-A$-transitions with high enough transition energy ΔE_n, which is nonzero because of tunneling between the *d*- and *l*-states, it can be estimated as

$$\Delta E \propto \exp(-2\sqrt{m_H a_H^2 \Delta V}\,\Delta\theta/\hbar),$$

where ΔV and $\Delta\theta$ are the characteristic height and width of the potential barrier. When the excitation frequency is chosen to be the resonant transition frequency $\omega_0 = \Delta E_n/\hbar$, only the resonant matrix elements will be significant in the Hamiltonians (10) and the torsional dynamics of the molecule can be examined in a two-level approximation.

Given the form (10) for the Hamiltonian, we have

$$\hat{H}_D \propto \cos\theta, \quad \hat{H}_Q \propto \sin\theta.$$

We show the form of the 2×2 matrix for the single-photon interaction Hamiltonian $\hat{H}_L$ for the total Hamiltonian $\hat{H}_I = \hat{H}_D + \hat{H}_Q$ for the case of linear polarization $e_x = 1$, $e_y = 0$. (Here there is no advantage in using circular polarization in the quadrupole approximation, because the polarization does not appear in the dependence of the quadrupole moment on the coordinate θ for this transition.) Given the representation of the interaction Hamiltonian in terms of the Pauli matrices for this transition, we obtain

$$\hat{H}_I \to V_{12}\hat{\sigma}^+ + V_{21}\hat{\sigma}^- = V_{12}(t)\hat{\sigma}_1(t) = V_{12}(0)\cos(\omega_L t + \tilde{\varphi}_L)(\cos(\omega_L t\hat{\sigma}_1) + \sin(\omega_L t\hat{\sigma}_2)),$$

where $V_{12} = V_{21}$ because the eigenfunctions ψ_k are real. After averaging over the oscillations in the field and the atomic polarization of the components at frequencies $2\omega_L$ in the rotating wave approximation,[15] we obtain

$$\hat{H}_L = \begin{pmatrix} 0 & \frac{1}{2}QS_{AS}e^{-i\tilde{\varphi}_L} \\ \frac{1}{2}QS_{AS}e^{i\tilde{\varphi}_L} & 0 \end{pmatrix}, \quad (14)$$

where, with Eqs. (10), (12), and (13), we obtain

$$Q = k_L a_O E_L e a_H \sqrt{1-\sin^2\vartheta\cos^2\varphi}\sin\tilde{\theta}, \quad (15)$$

where $\tilde{\varphi}_L$ is the initial phase of the laser field, which also includes a phase contribution, determined by Eq. (10), to the polarization of the field and the orientation of the molecule, and

$$S_{AS} = \int_{-\pi}^{\pi} \psi_A(\theta)\sin\frac{\theta}{2}\psi_S(\theta)\,d\theta \quad (16)$$

is a dimensionless matrix element describing the tunneling effect.

As a result, the Hamiltonian (14) is described by the matrix

$$\hat{H}_\Omega = \begin{pmatrix} -\frac{\hbar\delta}{2} & \frac{Q}{2}S_{AS}e^{-i\tilde{\varphi}_L} \\ \frac{Q}{2}S_{AS}e^{i\tilde{\varphi}_L} & \frac{\hbar\delta}{2} \end{pmatrix}, \quad (17)$$

where $\delta = \omega_L - \omega_{12}$ is the detuning of the laser field. This operator can be expressed in terms of the Pauli matrices:

$$\hat{H}_\Omega = \frac{\hbar}{2}(\mathbf{\Omega}\hat{\boldsymbol{\sigma}}), \quad \mathbf{\Omega} = (-\delta, QS_{AS}\cos\tilde{\varphi}_L, \sin\tilde{\varphi}_L). \quad (18)$$

The temporal evolution operator corresponding to the operator (17),

$$U(t) = \mathrm{T}\exp\left[(-i/\hbar)\int \hat{H}_\Omega dt\right]$$

can be calculated analytically for (a) a rectangular pulse $E_L = \text{const}$, or (b) zero offset $\delta = 0$, using the equations

a) $$U(t) = \cos\left(\frac{\Omega}{2}t\right)\hat{I} - i\sin\left(\frac{\Omega}{2}t\right)\left[-\frac{\delta}{\Omega}\hat{\sigma}_3 + \frac{QS_{AS}}{\Omega}(\hat{\sigma}_1\cos\tilde{\varphi}_L + \hat{\sigma}_2\sin\tilde{\varphi}_L)\right], \quad (19a)$$

b) $$U(t) = \cos\left(\frac{\Phi}{2}\right)\hat{I} - i\sin\left(\frac{\Phi}{2}\right)(\hat{\sigma}_1\cos\tilde{\varphi}_L + \hat{\sigma}_2\sin\tilde{\varphi}_L), \quad (19b)$$

where $\Omega = \sqrt{\Omega_0^2 + \delta^2}$ is the total Rabi frequency, $\Omega_0 = QS_{AS}$ is the Rabi frequency, and $\Phi = \int\Omega(t)dt$ is the angle of the laser pulse.

For $\delta = 0$, in accordance with Eq. (19b), the initial incoherent states, represented in the form

$$\hat{\rho}_0 = \hat{I}/2 + w\hat{\sigma}_3/2 \quad (-1 \leqslant w \leqslant 1),$$

transform into the states

$$\hat{\rho}_t=\frac{1}{2}[\hat{I}+w(\hat{\sigma}_3\cos\Phi-\hat{\sigma}_1\sin\Phi\sin(\omega_L t+\tilde{\varphi}_L)$$

$$-\hat{\sigma}_2\sin\Phi\cos(\omega_L t+\tilde{\varphi}_L))], \qquad (20)$$

where, besides the transformation (19), we have included free precession at the laser frequency, which is applicable in the standard form for the theory of the interaction[16] and the rotating wave approximation to the operators for the physical quantities. The components with $\hat{\sigma}_1$ and $\hat{\sigma}_2$ describe the contribution of the coherent superposition of the states ψ_S and ψ_A. In particular, for the lower initial state, which corresponds to $w=1$ for $\omega_L t+\tilde{\varphi}_L=\pi/2$, a $\pi/2$-pulse, for which $\Phi=\pm\pi/2$ and $\cos\Phi=\pm1$ in Eq. (20) and $\Phi/2=\pm\pi/4$ and $\cos(\Phi/2)=\pm1/\sqrt{2}$ in Eq. (19b), produces a jump from the initial ψ_S state into the chiral states $\psi_{1,2}=(\psi_S\pm\psi_A)/\sqrt{2}$, which correspond to the density matrices $(\hat{I}\mp\hat{\sigma}_1)/2$. Thus, for fixed angles

$$\Theta=(\vartheta,\varphi,\tilde{\theta})$$

it is possible to bring a molecule into a *d*- or *l*-state with 100% probability by an appropriate choice of parameters. In general, however, the resulting transformation of the state must be averaged over the angles Θ.

This averaging can be accomplished using a standard superoperator technique.[17] For the result (20) of the excitation of an incoherent initial state, upon averaging only over the sign of the parameter Q, which depends on the orientation angle $\tilde{\theta}$ of the hydrogen bonds along the x axis, which enters into the factor $\sin\Phi$ in Eq. (20), we obtain the following structure for the density matrix:

$$\hat{\rho}_t=(\hat{I}-w\hat{\sigma}_3\cos\Phi)/2.$$

This means that when there is no preferred orientation of the molecules in $\tilde{\theta}$, the density matrix transforms incoherently, i.e., diagonal density matrices transform into diagonal matrices. As a result of this transformation, the symmetry of the state, *S* or *A*, does not change, but the squared modulus of the wave function preserves reflection symmetry under the transformation $\theta\rightarrow-\theta$. Therefore, in order for reflection symmetry to be disrupted, the molecules must be oriented beforehand.

3.1. Excitation of initially oriented molecules

If there is an inhomogeneity in the distribution of the angles $\tilde{\theta}$ owing to the existence of a preferred orientation of the molecules, then the excitation of incoherent states can contain a coherent component associated with the excitation of chiral states that differ from ψ_A and ψ_S. The scalar characteristic of the degree of chirality is the average

$$\chi=2\left(\langle\psi_l|\hat{\rho}|\psi_l\rangle-\frac{1}{2}\right)=-2\left(\langle\psi_d|\hat{\rho}|\psi_d\rangle-\frac{1}{2}\right), \qquad (21)$$

where

$$\psi_{l,d}=\frac{1}{2}[\psi_S\pm\psi_A]$$

describe, respectively, left- and right-handed chiral states, which correspond to degrees of chirality $\chi=\pm1$ for $\hat{\rho}=|\psi_{l,d}\rangle\langle\psi_{l,d}|$. For a state $\hat{\rho}(t)$ excited by a rectangular laser pulse of duration τ_p with phase $\Phi=\Omega\tau_p$, the corresponding dependence of the degree of chirality of the excited state on the detuning δ and phases Φ, $\tilde{\varphi}_L$ can be calculated analytically.

For zero frequency offset, the analytic dependence of the degree of chirality on the angle and phase of the laser pulse has the form

$$\chi=-\sin\Phi\sin\tilde{\varphi}_L.$$

The qualitative dependence for nonzero offset is shown in Fig. 2a. The important point is the dependence of the effect on the phase of the laser field.

4. EXPERIMENTAL SCHEME FOR OBSERVING PHOTOINDUCED CHIRALITY IN H_2O_2

Given the features of the hydrogen peroxide molecule analyzed above, we propose the following experiment for observing photoexcited optical activity as an indicator of the breakdown of reflection symmetry.

As a working two-level system, the most appropriate is the $S-A$-transition with $n=1$, which corresponds to a wavelength $\lambda\simeq86$ μm and frequency $\omega_{12}=116.34$ cm^{-1}, which is substantially greater than the corresponding frequency for $n=0$ (11.44 cm^{-1}). To create a population in this transition it is easiest to use Raman excitation by two lasers tuned to the frequency of the $S-S$-transition $n=0\rightarrow n=1$. Since only dipole-active transitions are used here, we can obtain an essentially complete saturation regime in the active volume of the medium, so that in this stage there are no important limitations and we can start with $n_S\sim1$ for estimating the initial population on this transition.

In view of the oscillations in the degree of chirality excited in this scheme, it seems appropriate to use a laser with frequency ω_1 for probing and another frequency ω_2, which differs by the frequency $\omega_1-\omega_2=\omega_{12}$ of the precession in the degree of chirality, for detection. This can be done efficiently using two crossed polarizers. Detection of nonzero polarization of the output field perpendicular to the probe field corresponds to detection of the rotation effect. Here the probe field can be simultaneously employed to create an effect analogous to the orientation by an electrostatic field described in Sec. 1. According to Eq. (6), it leads to an angular displacement by this frequency. Therefore, the quadrupole Hamiltonian corresponding to excitation by the other laser field at frequency ω_2 and with a polarization direction along the x axis receives a correction at the resonant frequency ω_{12},

$$\Delta Q=\frac{E_1E_2e^2a_{\mathrm{H}}^2}{8J\omega_1^2}k_La_{\mathrm{O}}\sqrt{1-\sin^2\vartheta\cos^2\varphi}$$

$$\times\cos^2\tilde{\theta}u_1(t)u_2(t). \qquad (22)$$

Here an additional dependence on the torsional angle of the form $\cos(\theta/2)$ should be included in the matrix element (16),

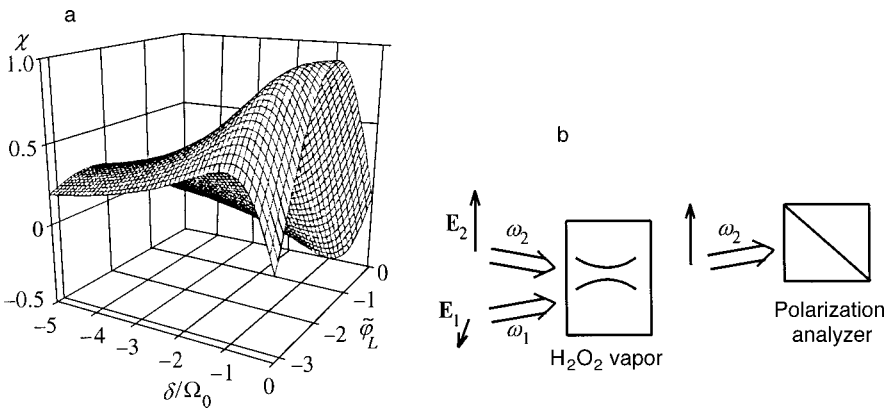


FIG. 2. (a) Degree of chirality χ of the excited state as a function of the relative detuning δ/Ω_0 and phase $\tilde{\varphi}_L$ of the laser field. (b) Layout of an experiment with two-frequency excitation. Radiation from one of the lasers is also used as the probe radiation.

which reduces to the substitution $\sin(\theta/2)\rightarrow(\sin\theta)/2$; this matrix element is also nonzero, while the additional time dependence of the form $\cos(\omega_1 t+\varphi_1)$ reduces to replacing the laser frequency, $\omega_L\rightarrow\omega_1-\omega_2$, and the phase, $\tilde{\varphi}_L\rightarrow\varphi_1-\varphi_2$. Equation (22) yields the order of magnitude estimate

$$\Delta Q\tau_p/\hbar\sim 10^{-4}\sqrt{I_1 I_2}\,\tau_p,$$

where the pulse duration τ_p is in seconds and the intensities I_1 and I_2 are in W/cm^2. Thus, in order to obtain the effect of a $\pi/2$-pulse for pulse lengths ~ 1 ns, the average geometric intensities of the pulses employed must be of order $I_0=10^4/\tau_p\sim 10^{13}$ W/cm^2. This quantity only yields an upper bound on the laser power to be used, which may actually be bounded by the substantially lower intensities corresponding to small $\Delta Q\tau_p/\hbar$.

The minimum measured rotation angle $\varphi_{\min}$, in conjunction with the expected angular rotation $\alpha=\Delta\varphi/\Delta L$, determines the length L of the active region. For α, we can proceed from the estimate $\alpha=k_L^2 a_O(\varepsilon-1)$, where $\varepsilon-1\sim 10^{-4}$ is the characteristic dispersion in the visible for H_2O_2 vapor at atmospheric pressure produced by suitable heating of the cell. This estimate is based on the assumption that for the chiral state of the molecule, the order of the specific rotation compared to the linear polarization effect contains a small parameter $k_L a_O\sim 10^{-3}$. The corresponding minimum length of the active region is $L=\varphi_{\min}/\alpha$, which for $\varphi_{\min}$ of order 1 arcsec yields $L\sim 10^{-2}$ cm. It is important to note that this is precisely of the same order as the wavelength corresponding to free precession at this transition. In this way, propagation effects on the frequency of the transition are relatively unimportant. The corresponding minimum laser beam waist in the active region[18] $w_0^2=\lambda_L L/\pi$ corresponds to pulsed laser powers

$$W_L=I_0 w_0^2,$$

which for these parameters is of order 10^7 W.

The geometry of a two-frequency experimental layout is shown in Fig. 2b. Light at either of the two frequencies $\omega_{1,2}$ can be used to detect the optical rotation effect. In this scheme, averaging over orientation shows up only as uncertainty in the Rabi frequency owing to the $\cos^2\tilde{\theta}$ dependence, which, as opposed to the electrostatic orientation method, does not lead to a drop in the response of the same order of magnitude as the effect itself. In order for the exciting field at frequency ω_2 not to interfere with detection of the effect at the same frequency, the beams must be slightly skewed so that this does not have a significant effect on the field distribution in the active region.

5. CONCLUSION

The above analysis indicates the feasibility of controlled excitation and detection of molecular states with disrupted reflection symmetry by optical methods. Introducing the techniques of nonlinear optics into the study of chiral states would signify the emergence of an efficient source of new information on the dynamic parameters of molecules which determine the conformational properties of chiral states.

This work was initiated by Prof. N. I. Koroteev, who devoted the last years of his short life to studying the puzzle of chiral purity in nature. One of his major ideas was to apply the methods of nonlinear optics to research on this problem. He hoped that in this way, not only could new highly sensitive methods for nonlinear optical diagnostics of media with chiral symmetry be developed, but also the mechanisms for controlled regulation of chirality might be understood. His untimely passing precluded his doing so, but his ideas continue to live on in the work of his successors and students. We respectfully dedicate this paper to the memory of N. I. Koroteev.

This work was partly supported by the Russian Fund for Fundamental Research, Grant No. 96-15-96460, for the ''S. A. Akhmanov and R. V. Khokhlov School on Nonlinear Optics and Laser Physics.'' We also thank V. I. Tyulin for providing the latest data on the potential for internal rotation of the hydrogen peroxide molecule, and A. Yu. Chikishev for discussing the design of a possible experiment.

*)E-mail: grishan@comsim1.ilc.msu.su

1)As necessary, the numerical calculations can be carried out using exact formulas.

[1] V. A. Avetisov and V. I. Goldanskiĭ, Usp. Fiz. Nauk **166**, 874 (1996).
[2] A. I. Kitaĭgorodskiĭ, *Introduction to Physics* [in Russian], Nauka, Moscow (1973).
[3] D. C. Walker, ed., *Origins of Optical Activity in Nature*, Elsevier, Amsterdam (1979).
[4] W. Thiemann, ed., *Generation and Amplification of Chirality in Chemical Systems*, Reidel, Dordrecht (1981).

[5] B. Ya. Zel'dovich, D. B. Saakyan, and I. I. Sobel'man, JETP Lett. **25**, 94 (1977).
[6] R. A. Hegstrom, D. W. Rein, and P. G. C. Sandars, Chem. Phys. **73**, 2329 (1980).
[7] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
[8] A. Salam, in *Proc. of the Eighth Nobel Symposium* (1968), p. 367.
[9] S. A. Akhmanov and N. I. Koroteev, *Methods of Nonlinear Optics in Light Scattering Spectroscopy* [in Russian], Nauka, Moscow (1981).
[10] Y. Amako and P. A. Goguère, Can. J. Chem. **40**, 765 (1962).
[11] R. L. Redington, W. B. Olson, and P. C. Gross, J. Chem. Phys. **36**, 1311 (1962).
[12] W. R. Busing and H. A. Levy, J. Chem. Phys. **42**, 3054 (1965).
[13] V. I. Tyulin, P. A. L. Bachi-Tom, and V. K. Matveev, Vestn. Mosk. Univ., Ser. 2, Khimiya **39**(2), 75 (1998).
[14] P. A. L. Bachi-Tom, Candidate's Dissertation in Chemical Sciences [in Russian], Moscow State Univ., Moscow (1998).
[15] A. Allen and J. Eberly, *Optical Resonance and Two-level Atoms*, Wiley, New York (1975).
[16] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, 3rd ed., Pergamon, New York (1977).
[17] B. A. Grishanin, *Quantum Electrodynamics for Electrical Engineers* [in Russian], Izd. MGU, Moscow (1981).
[18] W. Demtroeder, *Laser Spectroscopy: Basic Concepts and Instrumentation*, Springer-Verlag, New York (1981).

Translated by D. H. McNeill

Negative magnetic viscosity in two dimensions

A. V. Chechkin*)

National Academy of Science of Ukraine, Institute for Single Crystals, 310001 Kharkov, Ukraine
(Submitted 12 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1264–1286 (October 1999)

The occurrence of ''negative viscosities'' is studied within the framework of two-dimensional magnetohydrodynamics MHD. We use assumptions which are typical when studying the effects of smaller-scale fields on larger-scale ones, namely, the small-scale MHD fields are assumed to be sufficiently weak, jointly stationary, homogeneous, and maintained by external sources. The criteria of large-scale field generation due to negative viscosities are derived for various special forms of isotropic small-scale fields as well as anisotropic ones; the latter can be regarded as MHD stochastic analogs of the known Kolmogorov flow.
 [S1063-7761(99)01010-0]

1. INTRODUCTION

Problems of pattern formation and self-organization have been extensively studied in various hydrodynamic models of fluids and plasmas for at least three decades. Effects called ''negative viscosity'' belong to this wide class of the phenomena. This term was introduced by geophysicists and specialists in hydrodynamics in the 1950s, when analyzing large-scale geophysical processes.[1,2] However, it is still not well known, and even seems paradoxical. Therefore, let us discuss its origin in more detail.

We first consider the classical theory of gases. In this framework, starting with the Boltzmann equation, one can derive hydrodynamic equations. The equation for momentum transport has the form

$$\frac{\partial v_i}{\partial t}+(\mathbf{v}\nabla)v_i=-\frac{1}{\rho}\frac{\partial p}{\partial x_i}-\frac{1}{\rho}\frac{\partial \pi_{ij}}{\partial x_j}. \tag{1.1}$$

Here ρ, $\rho\mathbf{v}$ are the mass and momentum densities, respectively, $x_i=x,y,z$; p is pressure (here it is the ideal gas pressure), and π_{ij} is the viscosity tensor which describes irreversible ''viscous'' momentum transport. An explicit form of this tensor is obtained by invoking the smallness of the Knudsen number $\mathrm{Kn}=l_{mfp}/L$, where l_{mfp} is the mean free path and L is an external scale (scale of inhomogeneity). In the first order of Kn, that is, in the 13-moment approximation of the Grad method,[3] one has

$$\pi_{ij}=-\eta\left(\frac{\partial v_i}{\partial x_j}+\frac{\partial v_j}{\partial x_i}-\frac{2}{3}\delta_{ij}\frac{\partial v_k}{\partial x_k}\right), \tag{1.2}$$

where η is the dynamical viscosity, which is of the order of $\rho v_{th} l_{mfp}$, v_{th} is the thermal velocity, and η is always positive; this is in agreement with the intuitively obvious fact that irreversible momentum transport proceeds from regions of higher velocity to those of lower velocity.

We now proceed with the thermodynamics of irreversible processes. In this framework the equation for momentum transport retains the form (1.1) (where the pressure p is determined by the motion of particles and their interaction). The form of the viscous tensor, however, cannot be determined from the kinetic equation, but instead has to be established from general principles.[4] Specifically, (i) viscous momentum transfer appears when different parts of the fluid move with different velocities, so π_{ij} must depend on the space derivatives of the velocity $\mathbf{v}$; (ii) it is assumed that these derivatives are not too large; thus, viscous momentum transfer depends on the first derivatives only; and (iii) π_{ij} must tend to zero when the fluid rotates uniformly as a whole at some angular frequency Ω. Since linear combinations of the type

$$\frac{\partial v_i}{\partial x_j}+\frac{\partial v_j}{\partial x_i}$$

tend to zero when $\mathbf{v}$ is equal to $[\Omega\times x]$, only those linear combinations are contained in π_{ij}. Finally, below we assume an incompressible fluid, $\rho=$const, and thus

$$\frac{\partial v_k}{\partial x_k}=0.$$

Therefore, the viscous tensor retains the form (1.2) (without the last term in brackets), where η, of course, does not obey the simple gas law, but is instead a function of pressure and temperature. Assuming that the change in viscosity along the fluid is negligible, one can replace the term $-(1/\rho)\partial\pi_{ij}/\partial x_j$ in Eq. (1.1) with $\nu\Delta v_i$, where $\nu=\eta/\rho$ is the kinematic viscosity and Δ is the Laplacian. The positivity of η stems from the second law of thermodynamics for irreversible processes: the entropy S of a closed (isolated) system cannot decrease, $dS\geqslant 0$, whereas entropy production in a local equilibrium approximation has the same sign as η; see the detailed derivation in Ref. 3.

We now turn to turbulent processes that occur in open hydrodynamic systems. It is assumed for the turbulence of a fluid that Eqs. (1.1) and (1.2) are also valid for the stochastic velocity and pressure fields. Thus, we can define mean and fluctuating quantities

$$\mathbf{v}=\overline{\mathbf{v}}+\mathbf{v}^T,\quad p=\overline{p}+p^T,$$

where the bar denotes statistical averaging, and a superscript T denotes fluctuating (i.e., turbulent) components. In this paper we also use angle brackets $\langle ... \rangle$ to indicate statistical averaging in correlation functions. For the mean fields one has the equation

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x_j} \overline{v}_j \right) \overline{v}_i = -\frac{1}{\rho} \frac{\partial \overline{p}}{\partial x_i} + \nu \Delta \overline{v}_i - \frac{1}{\rho} \frac{\partial}{\partial x_j} \pi_{ij}^{\text{turb}}, \quad (1.3)$$

where

$$\pi_{ij}^{\text{turb}} = \rho \langle v_i^T v_j^T \rangle$$

are the Reynolds stresses.

Based on the analogy with the kinetic theory of gases, Boussinesq proposed to approximate the Reynolds stresses as

$$\langle v_i^T v_j^T \rangle = -\nu_{\text{turb}} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right), \quad (1.4)$$

where ν_{turb} is the turbulent kinematic viscosity. In contrast to the kinematic viscosity ν, which describes the physical properties of the gas (fluid), the turbulent kinematic viscosity describes the statistical properties of turbulent flows. An important fact is that ν_{turb} need not be positive, because there is no thermodynamic basis for its positivity. Indeed, we can speculate that when an external force (source) is added to the right hand side of Eq. (1.1), such that momentum (energy) is pumped into the small-scale component of the turbulence from this source (not from the mean flow), the momentum (energy) of fluctuations will be transferred to the mean flow. Assuming that ν_{turb} does not depend on x, Eq. (1.3) can be rewritten as

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x_j} \overline{v}_j \right) \overline{v}_i = -\frac{1}{\rho} \frac{\partial \overline{p}}{\partial x_i} + (\nu + \nu_{\text{turb}}) \Delta \overline{v}_i, \quad (1.5)$$

where the sum $\nu + \nu_{\text{turb}}$ can be negative, and thus large-scale instability occurs, accompanied by an increase in the mean flow.

Approximation (1.3) assumes the locality of the turbulent mechanism of momentum transport. It implies a small ratio of the characteristic scale of turbulent fluctuations (vortices) to the scale of the mean flow (by analogy with the Knudsen number). Experimental data[5] suggest that in many cases the scale of turbulent fluctuations is of the order of that of the mean flow, and the gradient approximation (1.4) becomes inadequate. In such cases, the contragradient transport is also observed frequently,[6] but this case is much more involved from a theoretical standpoint. Thus, in this paper we restrict ourselves to the case in which the scales of fluctuation are much smaller than those of the mean quantities. We also note that in Ref. 7 the peculiarities of numerical simulation of nonlocal momentum and thermal turbulent transport are discussed in detail.

In Ref. 2 a qualitative analysis (with an accent on empirical data analysis) is carried out on a set of geophysical and astrophysical processes and laboratory experiments, in which negative viscosity effects have already been detected, or, can at least be suspected:

1) differential rotation of the Earth's atmosphere;
2) differential rotation of the Sun photosphere;
3) differential rotation of disks of spiral galaxies;
4) flows in some laboratory experiments;
5) Gulf Stream near American coast.

One can conclude that negative viscosity phenomena are ubiquitous; they occur in systems possessing various physical and geometrical properties, and on a very different range of scales, ranging from laboratory flows a few centimeters in diameter to galaxies that are kiloparsecs in diameter. It is worthwhile to note that a similar problem also occurs in a tokamak plasma, where the peaked profiles in H-mode regime can be attributed to negative turbulent dissipative coefficients and contragradient transport.

A review of the examples above suggests a necessary condition for maintenance of a (quasi)stationary regime with negative viscosity effects prevailing: turbulence must not be ''passive''[2] in the sense that it is not fed by the kinetic energy of the mean flow, but instead has another source of energy, e.g., a thermal source, as occurs in Earth's atmosphere. Therefore, in papers dealing with a quantitative description of negative viscosity, the problem is formulated as follows. Suppose that we have a source giving rise to deterministic flows or stationary turbulence in a hydrodynamic medium. In the former case, the explicit form of the deterministic flows (or, at least, their symmetry properties) is assumed to be known, whereas in the latter case their statistical properties are known. The characteristic scale(s) of the given motion is assumed to be much smaller than the outer scale of the system of interest. The question is: can these small-scale motions act as a negative turbulent viscosity upon large-scale ones? From a theoretical standpoint, this implies the appearance of a negative dissipative factor in the equation for the mean flow, and its growth is understood as a manifestation of long-wavelength instability in a system of small-scale flows or vortices, the energy of small-scale motion being constant (it is mathematically convenient to treat the small-scale motion as being generated by an external source). This formulation is an example of inverse cascade problems in hydrodynamic systems. A similar (but not identical) example is related to the description of anomalous flows of the turbulent kinetic energy through the spectrum toward small wave numbers via local interaction between turbulent modes, and to the formation of stationary turbulent spectra.[8,9]

A number of analytic studies of the effects (characterized by the effective viscosity) of smaller-scale motion on larger-scale motion have begun since the late 1950s, when A. Kolmogorov proposed to study stability of a plane periodic flow sustained by a one-dimensional space-periodic external source in a viscous incompressible fluid.[10] This problem was first considered in Ref. 11, where the criterion of large-scale instability of one-dimensional space-periodic flow was found. Many subsequent papers are devoted to various aspects of the theory of Kolmogorov flow and its ''relatives'' in fluids; see, e.g., Refs. 12–15 and references therein. Being regarded as an ''elementary object'' of realistic turbulence with many degrees of freedom, Kolmogorov flow appears to be very useful in systematic studies of the peculiarities of the transition to turbulence, of the inverse cascade process in

two-dimensional (2D) turbulent flow, and in coherent structure generation. Experimentally, a 2D flow subject to periodic forcing was studied in a thin layer of an electrolyte.[16] Two-dimensional hydrodynamic flows also attract considerable attention because the direct numerical solution of 2D fluid equations is a simpler problem than the solution of 3D problems.[17]

For plasmas an analogous problem was considered in Ref. 18 where the stability of a gradient-drift wave was studied and coherent nonlinear structures formed as the result of instability were found.

Kolmogorov flow instability can be regarded as a simple manifestation of the negative viscosity effect when the small-scale basic flow is one-dimensional and space-periodic. Other basic forms of small-scale motion in fluids (isotropic time-independent[14] and δ-correlated in time[19]) have also been considered. A possible occurrence of the negative viscosity effect was studied in Ref. 20 for coherent wave motion, as well as for the small-scale Rossby turbulence and gradient drift-wave turbulence. We also mention the emergence of negative viscosity in a ferrofluid in an alternating magnetic field.[21,22]

Electrically conducting fluids exhibit a wide variety of turbulent phenomena. Here the concept of negative viscosities (both kinematic and magnetic) can also be useful for understanding the peculiarities of self-organizing processes. In Refs. 23 and 24 it was pointed out that in a low-β plasma such as in a tokamak, small-scale magnetic turbulence acts as a negative effective magnetic viscosity on large-scale magnetic field perturbations. This leads to amplification of the large-scale field, and is a very likely mechanism in explosive magnetic phenomena, such as disruptions in tokamaks and solar flares. The reduced MHD equations[25] were taken as a starting point in these papers. It was found that the turbulent magnetic viscosity becomes negative if the magnetic energy of small-scale turbulence exceeds the kinetic energy, whereas the turbulent kinematic viscosity is positive. This problem was reconsidered in Ref. 19. It was found that in 2D MHD, the conclusions are the same as in Refs. 23 and 24, but for reduced MHD the results are inconsistent with those obtained in these articles. Furthermore, the role of cross-correlations and anisotropy of fluctuations remains unclear.

A more recent review,[26] as well as numerical simulations[27,28] (performed for freely decaying, not forced, turbulence), do not shed light on the problem of interest. Thus, it seems reasonable to study negative viscosity effects in MHD in more detail. In this paper we consider the problem in the context of 2D MHD, and based on the formulation of negative viscosities presented above.

This paper is organized as follows. In Sec. 2 we obtain general equations governing the evolution of large-scale MHD fields, and demonstrate explicitly how the negative magnetic viscosity term can appear in the equation for the mean magnetic potential. In Secs. 3–5, we study the influence (characterized by the turbulent viscosities) of various small-scale fields on large-scale fields. Specifically, in Sec. 3, small-scale turbulence is generated by a stationary white noise source; the transition to the results of Ref. 19 is demonstrated. In Sec. 4, more general forms of isotropic small-scale turbulence with correlation times spanning a broad range are used, and the transition to the results of Refs. 14 and 24 is shown. Finally, in Sec. 5 we consider stochastic analogs of Kolmogorov flow for magnetohydrodynamics. In Secs. 3–5 the criteria for large-scale field growth are also derived. The results are summarized in Sec. 6. A detailed derivation of the equations governing the evolution of large-scale MHD fields is presented in the Appendix.

2. EQUATIONS FOR LARGE-SCALE FIELD EVOLUTION AND THE ORIGIN OF THE NEGATIVE MAGNETIC VISCOSITY TERM

We study a 2D incompressible conducting fluid with velocity field $\mathbf{v}(\mathbf{x},t)=[\mathbf{e}_z\nabla\psi]_z$ and magnetic field $\mathbf{B}(\mathbf{x},t)=[\mathbf{e}_z\nabla a]_z$, both in the xy plane, $\mathbf{x}=(x,y)$; ψ is the stream function, a is the magnetic potential, $\nabla\equiv\mathbf{e}_x\partial/\partial x+\mathbf{e}_y\partial/\partial y$, $[\ldots]_z$ implies z-projection of the vector product. The 2D MHD equations can be written as[26]

$$\frac{\partial}{\partial t}W+\mathbf{v}\nabla W=\mathbf{B}\nabla j+\nu\Delta W,$$

$$\frac{\partial}{\partial t}a+\mathbf{v}\nabla a=\mathbf{B}\nabla j+\eta\Delta a. \tag{2.1}$$

Here $W=\nabla^2\psi$ is the vorticity, $j=\nabla^2 a$ is the current, $\nabla^2\equiv\Delta=\partial^2/\partial x^2+\partial^2/\partial y^2$, and ν, η are the kinematic and the magnetic viscosities, respectively. In (2.1) the density of the fluid is assumed to be unity, and the magnetic field has the dimension of velocity. Furthermore, it is normally assumed that the kinematic and magnetic viscosities are of the same order. For our purposes it is convenient to rewrite (2.1) as

$$\frac{\partial\Delta\psi}{\partial t}+[\nabla\psi\times\nabla\Delta\psi]_z=[\nabla a\times\nabla\Delta a]_z+\nu\Delta^2\psi,$$

$$\frac{\partial a}{\partial t}+[\nabla\psi\times\nabla a]_z=\eta\Delta a. \tag{2.2}$$

Now we divide ψ and a into mean and fluctuating (turbulent) components:

$$\psi=\overline{\psi}+\psi^T,\quad a=\overline{a}+a^T,$$

where the bar denotes statistical averaging, and T signifies ''turbulent.'' Below we also use angle brackets $\langle\ldots\rangle$ for the statistical averaging.

From Eqs. (2.2) we obtain

$$\left(\frac{\partial}{\partial t}-\nu\Delta\right)\Delta\overline{\psi}=-\langle[\nabla\psi^T\times\nabla\Delta\psi^T]_z\rangle+\langle[\nabla a^T\times\nabla\Delta a^T]_z\rangle-[\nabla\overline{\psi}\times\nabla\Delta\overline{\psi}]_z+[\nabla\overline{a}\times\nabla\Delta\overline{a}]_z, \tag{2.3}$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta\right)\overline{a}=-\langle[\nabla\psi^T\times\nabla a^T]_z\rangle-[\nabla\overline{\psi}\times\nabla\overline{a}]_z, \tag{2.4}$$

whereas for the fluctuating components we obtain, by subtracting Eqs. (2.3) and (2.4) from (2.2),

$$\left(\frac{\partial}{\partial t}-\nu\Delta\right)\Delta\psi^T+[\nabla\psi^T\times\nabla\Delta\overline{\psi}]_z+[\nabla\overline{\psi}\times\nabla\Delta\psi^T]_z$$
$$+([\nabla\psi^T\times\nabla\Delta\psi^T]_z-\langle[\nabla\psi^T\times\nabla\Delta\overline{\psi}]_z\rangle)$$
$$=[\nabla a^T\times\nabla\Delta\overline{a}]_z+[\nabla\overline{a}\times\nabla\Delta a^T]_z+([\nabla a^T$$
$$\times\nabla\Delta a^T]_z-\langle[\nabla a^T\times\nabla\Delta a^T]_z\rangle), \tag{2.5}$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta\right)a^T+[\nabla\psi^T\times\nabla\overline{a}]_z+[\nabla\overline{\psi}\times\nabla a^T]_z$$
$$+([\nabla\psi^T\nabla a^T]_z-\langle[\nabla\psi^T\times\nabla a^T]_z\rangle)=0. \tag{2.6}$$

To derive a closed set of equations for $\overline{\psi}$, $\overline{a}$ it is necessary to express the quantities

$$Q_1=-\langle[\nabla\psi^T\times\nabla\Delta\psi^T]_z\rangle, \tag{2.7}$$

$$Q_2=\langle[\nabla a^T\times\nabla\Delta a^T]_z\rangle, \tag{2.8}$$

$$Q_3=-\langle[\nabla\psi^T\times\nabla a^T]_z\rangle \tag{2.9}$$

in terms of $\overline{\psi}$, $\overline{a}$ [see Eqs. (2.3) and (2.4)]. Since we are interested in calculating turbulent viscosity (but not in the problems related to the evaluation of the turbulent spectrum), we use an approach developed in Ref. 20 to study negative viscosity in Rossby and drift-wave turbulence, and resembling an approach previously used in the dynamo problem.[29,30] It also resembles the ''quasilinear approximation'' frequently employed to calculate the turbulent transport coefficients in magnetized inhomogeneous plasmas.[31] In so doing, we assume (in accordance with the discussion in Sec. 1) that the mean quantities vary on spatial and temporal scales that are larger than the characteristic scales of the fluctuating fields; that the statistical properties of the small-scale fields are known; and that quadratic terms in ψ^T and a^T can be neglected in Eqs. (2.5) and (2.6). Thus, instead of Eqs. (2.5) and (2.6) we obtain

$$\left(\frac{\partial}{\partial t}-\nu\Delta\right)\Delta\psi^T+[\nabla\psi^T\times\nabla\Delta\overline{\psi}]_z+[\nabla\overline{\psi}\times\nabla\Delta\psi^T]_z$$
$$=[\nabla a^T\times\nabla\Delta\overline{a}]_z+[\nabla\overline{a}\times\nabla\Delta a^T]_z+F_\psi, \tag{2.10}$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta\right)a^T+[\nabla\psi^T\times\nabla\overline{a}]_z+[\nabla\overline{\psi}\times\nabla a^T]_z=F_a, \tag{2.11}$$

where the stochastic sources $F_\psi(\mathbf{x},t)$ and $F_a(\mathbf{x},t)$ which maintain the stationary level of MHD fluctuations, are formally added to the right-hand side of the equations. We then solve Eqs. (2.10) and (2.11), insert the solutions into Eqs. (2.7)–(2.9) and then into Eqs. (2.3) and (2.4), and obtain the evolution equations for the mean fields $\overline{\psi}$, $\overline{a}$. Because we invoke a two-scale approximation for our problem, we introduce the ''slow'' variable $\mathbf{X}$ and the ''fast'' variable $\mathbf{x}$. The mean quantities depend on the slow variable only, whereas the fluctuating components depend on both fast and slow variables. We also have

$$\left|\frac{\partial}{\partial\mathbf{X}}\right|\approx|\mathbf{K}|\ll\left|\frac{\partial}{\partial\mathbf{x}}\right|\approx|\mathbf{k}|,$$

where $\mathbf{K}$ and $\mathbf{k}$ are large-scale and small-scale wave vectors, respectively.

We seek solutions of Eqs. (2.10) and (2.11) in powers of K, that is,

$$\psi^T=\psi^{(0)}(\mathbf{x},t)+\psi^{(1)}(\mathbf{x},\mathbf{X},t)+\ldots,$$

$$a^T=a^{(0)}(\mathbf{x},t)+a^{(1)}(\mathbf{x},\mathbf{X},t)+\ldots,$$

where the terms $\psi^{(0)}$, $a^{(0)}$ are sustained by external sources, whereas $\psi^{(1)}$, $\psi^{(2)}$, etc. appear because of the interaction between small- and large-scale components. Assuming that the small-scale fluctuations are jointly stationary and homogeneous (which, in turn, implies that both fields are stationary and homogeneous), we introduce their correlation and cross-correlation functions $C_{\psi\psi}$, C_{aa}, $C_{\psi a}$ as well as their space-time spectral functions $\hat{C}_{\psi\psi}$, $\hat{C}_{aa}$, $\hat{C}_{\psi a}$, e.g., as

$$\langle\psi^{(0)}(\mathbf{x},t)a^{(0)}(\mathbf{x}',t')\rangle=C_{\psi a}(\mathbf{x}-\mathbf{x}',t-t')$$
$$=\int\frac{d\mathbf{k}\,d\omega}{(2\pi)^3}\hat{C}_{\psi a}(\mathbf{k},\omega)$$
$$\times\exp[-i\omega(t-t')+i\mathbf{k}(\mathbf{x}-\mathbf{x}')].$$

The properties of the spectral functions of zeroth-order quantities are listed in Appendix, along with a detailed derivation of $Q_{1,2,3}$. Here we present the final result. For the evolution of the mean quantities $\overline{\psi}$, $\overline{a}$ instead of Eqs. (2.3) and (2.4), we have

$$\left(\frac{\partial}{\partial t}-\nu\Delta_s\right)\Delta_s\overline{\psi}=\varepsilon_{kl}\varepsilon_{mn}\left\{\delta^{(1)}_{lnp}\frac{\partial^3\overline{\psi}}{\partial X_k\partial X_m\partial X_p}+\delta^{(2)}_{lnp}\right.$$
$$\times\frac{\partial^3\overline{a}}{\partial X_k\partial X_m\partial X_p}+\nu^{(1)}_{ln}\frac{\partial^2\Delta_s\overline{\psi}}{\partial X_k\partial X_m}$$
$$+\nu^{(2)}_{ln}\frac{\partial^2\Delta_s\overline{a}}{\partial X_k\partial X_m}+\nu^{(1)}_{lnpr}\frac{\partial^4\overline{\psi}}{\partial X_k\partial X_m\partial X_p\partial X_r}$$
$$\left.+\nu^{(2)}_{lnpr}\frac{\partial^4\overline{a}}{\partial X_k\partial X_m\partial X_p\partial X_r}\right\}, \tag{2.12}$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta_s\right)\overline{a}=\varepsilon_{kl}\varepsilon_{mn}\left\{\eta^{(1)}_{ln}\frac{\partial^2\overline{a}}{\partial X_k\partial X_m}+\eta^{(2)}_{ln}\frac{\partial^2\overline{\psi}}{\partial X_k\partial X_m}\right\}, \tag{2.13}$$

where $\Delta_s\equiv\partial^2/\partial X^2+\partial^2/\partial Y^2$, ε_{mn} is the unit antisymmetric tensor of the second rank, and

$$\delta^{(1)}_{lnp}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}2ik_lk_nk_p\left\{\frac{\hat{C}_{aa}}{-i\omega+\eta k^2}-\frac{\hat{C}_{\psi\psi}}{-i\omega+\nu k^2}\right\},$$

$$\delta^{(2)}_{lnp}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}2ik_lk_nk_p\left\{\frac{\hat{C}_{\psi a}}{-i\omega+\eta k^2}-\frac{\hat{C}^*_{\psi a}}{-i\omega+\nu k^2}\right\},$$

$$\nu^{(1)}_{ln}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}k_lk_n\left\{\frac{\hat{C}_{aa}}{-i\omega+\eta k^2}-\frac{\hat{C}_{\psi\psi}}{-i\omega+\nu k^2}\right\},$$

$$\nu_{ln}^{(2)}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}k_lk_n\left\{\frac{\hat{C}_{\psi a}}{-i\omega+\eta k^2}-\frac{\hat{C}_{\psi a}^*}{-i\omega+\nu k^2}\right\},$$

$$\nu_{lnpr}^{(1)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\left\{\frac{\hat{C}_{\psi\psi}}{-i\omega+\nu k^2}+\frac{\nu k^2}{(-i\omega+\nu k^2)^2}\hat{C}_{\psi\psi}-\frac{\eta k^2}{(-i\omega+\eta k^2)^2}\hat{C}_{aa}\right\},$$

$$\nu_{lnpr}^{(2)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\left\{\frac{\eta k^2}{(-i\omega+\eta k^2)^2}\hat{C}_{\psi a}-\frac{\hat{C}_{\psi a}^*}{-i\omega+\nu k^2}-\frac{\nu k^2}{(-i\omega+\nu k^2)^2}\hat{C}_{\psi a}^*\right\},$$

$$\eta_{ln}^{(1)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}k_lk_n\left\{\frac{\hat{C}_{\psi\psi}}{-i\omega+\eta k^2}-\frac{\hat{C}_{aa}}{-i\omega+\nu k^2}\right\},$$

$$\eta_{ln}^{(2)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}k_lk_n\left\{\frac{\hat{C}_{\psi a}}{-i\omega+\nu k^2}-\frac{\hat{C}_{\psi a}^*}{-i\omega+\eta k^2}\right\}.$$

The reality of all coefficients in Eqs. (2.12) and (2.13) is easily demonstrated using the properties (A9) of the space-time spectral functions.

The resulting equations enable us to study the influence (characterized by the turbulent viscosities) of various forms of small-scale fields on large-scale fields. However, in order to make the results more transparent, in this Section we also demonstrate explicitly how negative magnetic viscosity originates in Eqs. (2.4) and (2.10). For this purpose it is convenient to assume that the field a^T is given. Then the smallness of large-scale gradients enables us to derive from Eq. (2.10) the following relation between the small-scale Fourier-components $\hat{\psi}^T(\mathbf{k},\omega)$, $\hat{a}^T(\mathbf{k},\omega)$:

$$\hat{\psi}^T(\mathbf{k},\omega)=\varepsilon_{mn}\frac{ik_n}{-i\omega+\nu k^2}\frac{\partial\overline{a}}{\partial X_m}\hat{a}^T+(\text{other terms}). \tag{2.14}$$

Here we explicitly write the term that derives from $[\nabla\overline{a}\times\nabla\Delta a^T]_z$ in Eq. (2.10). Further, we note that the terms entering into $Q_3=-\langle[\nabla\psi^T\times\nabla a^T]_z\rangle$ include

$$-\left\langle\left[\frac{\partial\psi^T}{\partial\mathbf{X}}\times\frac{\partial a^T}{\partial\mathbf{x}}\right]\right\rangle,$$

which, after inserting $\hat{\psi}^T$ from Eq. (2.14), gives rise to the term

$$-\varepsilon_{kl}\varepsilon_{mn}\frac{\partial^2\overline{a}}{\partial X_k\partial X_m}\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{k_lk_n}{-i\omega+\nu k^2}\langle(\hat{a}^T)^2\rangle_{\mathbf{k},\omega} \tag{2.15}$$

on the right-hand side of Eq. (2.4); here $\langle(\hat{a}^T)^2\rangle_{\mathbf{k},\omega}$ is the spectrum of the small-scale magnetic potential. Assuming isotropy and integrating over azimuthal angle in $\mathbf{k}$-space, one can easily show that the term (2.15) can be rewritten as

$$-\nu^T\Delta_s\overline{a}, \tag{2.16}$$

where

$$\nu^T=\int\frac{kdk\,d\omega}{8\pi^2}\frac{\nu k^4}{\omega^2+\nu^2k^4}\langle(\hat{a}^T)^2\rangle_{k,\omega}. \tag{2.17}$$

Equations (2.16) and (2.17) reveal a term of negative viscosity type in the equation for the mean component of the magnetic potential.

3. SMALL-SCALE ISOTROPIC TURBULENCE GENERATED BY STATIONARY WHITE NOISE SOURCE

It is mathematically convenient to treat small-scale turbulence as being generated by a stationary white noise source, to keep the statistical properties as simple as possible. This kind of the source, possessing zero cross-correlations, was considered in Ref. 19, so we are able not only to compare the results but also to clarify the role of cross-correlation terms. We define the properties of the sources F_ψ, F_a in Eqs. (2.10) and (2.11) as follows:

$$\begin{aligned}&\langle F_\psi\rangle=\langle F_a\rangle=0,\\&\langle F_\psi(\mathbf{x},t)F_\psi(\mathbf{x}',t')\rangle=\Psi(\mathbf{x}-\mathbf{x}')\delta(t-t'),\\&\langle F_a(\mathbf{x},t)F_a(\mathbf{x}',t')\rangle=A(\mathbf{x}-\mathbf{x}')\delta(t-t'),\\&\langle F_\psi(\mathbf{x},t)F_a(\mathbf{x}',t')\rangle=H(\mathbf{x}-\mathbf{x}')\delta(t-t').\end{aligned} \tag{3.1}$$

Since fluctuations $\psi^{(0)}$, $a^{(0)}$ are related to the sources F_ψ, F_a by Eqs. (A4), we find for the space-time spectral functions of the small-scale turbulence

$$\begin{aligned}&\hat{C}_{\psi\psi}(\mathbf{k},\omega)=\frac{\hat{\Psi}(\mathbf{k})}{k^4(\omega^2+\nu^2k^4)},\\&\hat{C}_{aa}(\mathbf{k},\omega)=\frac{\hat{A}(\mathbf{k})}{\omega^2+\eta^2k^4},\\&\hat{C}_{\psi a}(\mathbf{k},\omega)=-\frac{\hat{H}(\mathbf{k})}{k^2(i\omega+\nu k^2)(i\omega+\eta k^2)},\end{aligned} \tag{3.2}$$

where $\hat{\Psi}$, $\hat{A}$, $\hat{H}$ are the spatial Fourier transforms of Ψ, A, H, respectively, e.g.,

$$\hat{\Psi}(\mathbf{k})=\int d\boldsymbol{\kappa}\Psi(\boldsymbol{\kappa})\,\exp(-i\mathbf{k}\boldsymbol{\kappa}).$$

It is also useful to express space-time spectral functions (3.2) in terms of spatial functions,

$$\hat{C}_{\psi a}(\mathbf{k})=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\hat{C}_{\psi a}(\mathbf{k},\omega).$$

Integrating over ω in Eqs. (3.2), we express the sources in terms of spatial spectral functions and obtain

$$\begin{aligned}&\hat{C}_{\psi\psi}(\mathbf{k},\omega)=\frac{2\nu k^2}{\omega^2+\nu^2k^4}\hat{C}_{\psi\psi}(\mathbf{k}),\\&\hat{C}_{aa}(\mathbf{k},\omega)=\frac{2\eta k^2}{\omega^2+\eta^2k^4}\hat{C}_{aa}(\mathbf{k}),\\&\hat{C}_{\psi a}(\mathbf{k},\omega)=\frac{(\nu+\eta)k^2}{(-i\omega+\nu k^2)(i\omega+\eta k^2)}\hat{C}_{\psi a}(\mathbf{k}).\end{aligned} \tag{3.3}$$

We then insert Eqs. (3.3) into the coefficients of Eqs. (2.12) and (2.13) and integrate over ω:

$$\delta_{lnp}^{(1)}=0,$$

$$\delta_{lnp}^{(2)}=-\int\frac{d\mathbf{k}}{4\pi^2}\,i\,\frac{k_l k_n k_p}{k^2}\left(\frac{\hat{C}_{\psi a}(\mathbf{k})}{\eta}-\frac{\hat{C}^*_{\psi a}(\mathbf{k})}{\nu}\right),$$

$$\nu_{ln}^{(1)}=\int\frac{d\mathbf{k}}{4\pi^2}\,i\,\frac{k_l k_n}{2k^2}\left(\frac{\hat{C}_{aa}(\mathbf{k})}{\eta}-\frac{\hat{C}_{\psi\psi}(\mathbf{k})}{\nu}\right),$$

$$\nu_{lnpr}^{(1)}=\int\frac{d\mathbf{k}}{4\pi^2}\,i\,\frac{k_l k_n k_p k_r}{k^4}\left(\frac{3\hat{C}_{\psi\psi}(\mathbf{k})}{\nu}-\frac{\hat{C}_{aa}(\mathbf{k})}{\eta}\right),$$

$$\nu_{ln}^{(2)}=-\int\frac{d\mathbf{k}}{4\pi^2}\,\frac{k_l k_n}{2k^2}\left(\frac{\hat{C}_{\psi a}(\mathbf{k})}{\eta}-\frac{\hat{C}^*_{\psi a}(\mathbf{k})}{\nu}\right), \tag{3.4}$$

$$\nu_{lnpr}^{(1)}=\int\frac{d\mathbf{k}}{4\pi^2}\,i\,\frac{k_l k_n k_p k_r}{k^4}\left(\frac{\hat{C}_{\psi a}(\mathbf{k})}{\eta}-\frac{3\hat{C}^*_{\psi a}(\mathbf{k})}{\nu}\right),$$

$$\eta_{ln}^{(1)}=\int\frac{d\mathbf{k}}{4\pi^2}\,i\,\frac{k_l k_n}{k^2(\nu+\eta)}(\hat{C}_{\psi\psi}(\mathbf{k})-\hat{C}_{aa}(\mathbf{k})),$$

$$\eta_{ln}^{(2)}=0.$$

We note that if we set $\hat{C}_{\psi a}=0$, then Eqs. (2.12) and (2.13) together with Eqs. (3.4) appear to be in complete agreement with Eqs. (2.15) and (3.12)–(3.15) of Ref. 19 after some easy transforms. In this Section we assume isotropy of the small-scale spectra; the spatial spectral functions in Eqs. (3.4) therefore depend on $k\equiv|\mathbf{k}|$. Integrating over the azimuthal angle φ in $\mathbf{k}$-space in Eqs. (3.4) using the subsidiary integrals

$$\int_0^{2\pi}d\varphi k_m k_n=\pi k^2\delta_{mn},$$

$$\int_0^{2\pi}d\varphi k_k k_l k_m k_n=\frac{\pi k^4}{4}(\delta_{kl}\delta_{mn}+\delta_{kn}\delta_{lm}+\delta_{km}\delta_{ln}),$$

Eqs. (2.11) and (2.12) take the form

$$\left(\frac{\partial}{\partial t}-\nu\Delta_s\right)\Delta_s\overline{\psi}=\nu_{\rm turb}\Delta_s^2\overline{\psi}+\delta\nu_{\rm turb}\Delta_s^2\overline{a}, \tag{3.5}$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta_s\right)\overline{a}=\eta_{\rm turb}\Delta_s\overline{a},$$

where

$$\nu_{\rm turb}=\frac{1}{8}\left(\frac{\langle(\psi^{(0)})^2\rangle}{\nu}+\frac{\langle(a^{(0)})^2\rangle}{\eta}\right),$$

$$\delta\nu_{\rm turb}=-\left(\frac{1}{\nu}+\frac{1}{\eta}\right)\frac{\langle(\psi^{(0)}a^{(0)})\rangle}{8}, \tag{3.6}$$

$$\eta_{\rm turb}=\frac{1}{2(\nu+\eta)}(\langle(\psi^{(0)})^2\rangle-\langle(a^{(0)})^2\rangle),$$

$\langle(\psi^{(0)})^2\rangle$, $\langle(a^{(0)})^2\rangle$ are the stream-function and magnetic-potential variances of the small-scale fields, respectively:

$$\langle(\psi^{(0)})^2\rangle=\int\frac{d\mathbf{k}}{4\pi^2}\hat{C}_{\psi\psi}(k),$$

and

$$\langle(\psi^{(0)}a^{(0)})\rangle=\int\frac{d\mathbf{k}}{4\pi^2}\hat{C}_{\psi a}(k).$$

It follows from Eqs. (3.5) and (3.6) that for the case of small-scale turbulence generated by a stationary white noise source, the turbulent viscosity is always positive, the turbulent magnetic viscosity is negative if

$$\langle(a^{(0)})^2\rangle>\langle(\psi^{(0)})^2\rangle, \tag{3.7}$$

and if cross-correlations vanish, the large-scale magnetic field grows if $|\eta_{\rm turb}|$ is large enough that $(\eta+\eta_{\rm turb})$ becomes negative, while the large-scale velocity field does not grow. The existence of nonzero cross-correlations leads to amplification of both large-scale fields when $(\eta+\eta_{\rm turb})$ becomes negative; the latter conclusion is unaffected by the sign of cross-correlations $\langle(\psi^{(0)}a^{(0)})\rangle$.

4. MORE GENERAL FORMS OF ISOTROPIC SMALL-SCALE TURBULENCE

To provide a more general treatment of the isotropic case with zero cross-correlations, we define space-time spectral functions as follows:

$$\hat{C}_{\psi\psi}(k,\omega)=\frac{2\gamma_{1k}}{\omega^2+\gamma_{1k}^2}\hat{C}_{\psi\psi}(k),$$

$$\hat{C}_{aa}(k,\omega)=\frac{2\gamma_{2k}}{\omega^2+\gamma_{2k}^2}\hat{C}_{aa}(k), \tag{4.1}$$

$$\hat{C}_{\psi a}(k,\omega)=0.$$

The ω-dependent part of the spectrum is taken in the Lorentzian form here. This form is frequently used in the literature; however, we choose it for convenience only. It can be easily verified that the results are changed only by a factor of order unity if one chooses other shapes, for example, the Gaussian shape or the rectangular one. Inserting Eqs. (4.1) into Eqs. (2.12) and (2.13) and integrating over ω and φ in the coefficients, we arrive at Eqs. (3.5), where

$$\nu_{\rm turb}=\int\frac{dk\;k^3}{4\pi}\left\{\frac{\nu k^2}{(\gamma_{1k}+\nu k^2)^2}\hat{C}_{\psi\psi}(k)+\frac{\gamma_{2k}}{(\gamma_{2k}+\eta k^2)^2}\hat{C}_{aa}(k)\right\},$$

$$\eta_{\rm turb}=\int\frac{dk\;k^3}{4\pi}\left\{\frac{\hat{C}_{\psi\psi}(k)}{\gamma_{1k}+\eta k^2}-\frac{\hat{C}_{aa}(k)}{\gamma_{2k}+\nu k^2}\right\}, \tag{4.2}$$

whereas $\delta\nu_{\rm turb}=0$. For $\gamma_{1k}=\nu k^2$, $\gamma_{2k}=\eta k^2$ we naturally obtain $\nu_{\rm turb}$, $\eta_{\rm turb}$ of Sec. 3; see Eqs. (3.6). Here we consider two special cases.

1. Long correlation times of the small-scale fluctuations, $\gamma_{1k},\gamma_{2k}\ll\nu k^2,\eta k^2$. In this limit

$$\nu_{\rm turb}=\frac{\langle(\psi^{(0)})^2\rangle}{2\nu},$$

$$\eta_{\rm turb}=\frac{\langle(\psi^{(0)})^2\rangle}{2\eta}-\frac{\langle(a^{(0)})^2\rangle}{2\nu}, \tag{4.3}$$

and ν_{turb} coincides with that obtained in Ref. 14 for time-independent random basic flow (which, in fact, corresponds to $\gamma_{1k}\to 0$) in the ordinary fluid, $B=0$. In the case considered, the turbulent magnetic viscosity is negative if

$$\langle (a^{(0)})^2\rangle > \frac{\nu}{\eta}\langle (\psi^{(0)})^2\rangle, \tag{4.4}$$

whereas ν_{turb} is always positive.

2. Short correlation times, $\gamma_{1k}, \gamma_{2k} \gg \nu k^2, \eta k^2$.

Assuming that

$$\gamma_{1k} \approx \gamma_{2k} \approx 1/\tau_c,$$

where τ_c is the correlation time, which is independent of k, we obtain from Eqs. (4.2)

$$\nu_{\text{turb}} = \frac{\tau_c}{2}\langle (B^{(0)})^2\rangle,$$

$$\eta_{\text{turb}} = \frac{\tau_c}{2}(\langle (\mathbf{v}^{(0)})^2\rangle - \langle (\mathbf{B}^{(0)})^2\rangle), \tag{4.5}$$

where $\langle (\mathbf{v}^{(0)})^2\rangle$ and $\langle (\mathbf{B}^{(0)})^2\rangle$ are the velocity and magnetic-field variances, respectively. In this limit the turbulent viscosities are the same as those obtained in Ref. 24. From the second equation, it follows that the turbulent magnetic viscosity is negative if

$$\langle (\mathbf{B}^{(0)})^2\rangle > \langle (\mathbf{v}^{(0)})^2\rangle, \tag{4.6}$$

whereas ν_{turb} is again always positive.

Now we consider the role of cross-correlations. Whereas the general form (4.1) of the space-time spectral functions $\hat{C}_{\psi\psi}$, $\hat{C}_{aa}$ is natural and widely used, it is not so easy, in the author's opinion, to choose an analogous general form of the cross-correlation spectrum. We therefore restrict attention to time-independent isotropic fluctuations (special case 1; see above) of $\psi^{(0)}$, $a^{(0)}$:

$$\begin{aligned}
\hat{C}_{\psi\psi}(k,\omega) &= 2\pi\delta(\omega)\hat{C}_{\psi\psi}(k),\\
\hat{C}_{aa}(k,\omega) &= 2\pi\delta(\omega)\hat{C}_{aa}(k),\\
\hat{C}_{\psi a}(k,\omega) &= 2\pi\delta(\omega)\hat{C}_{\psi a}(k).
\end{aligned} \tag{4.7}$$

Inserting Eqs. (4.7) into Eqs. (2.12) and (2.13), we have

$$\begin{aligned}
\left(\frac{\partial}{\partial t} - \nu\Delta_s\right)\Delta_s\overline{\psi} &= \nu_{\text{turb}}\Delta_s^2\overline{\psi} + \delta\nu_{\text{turb}}\Delta_s^2\overline{a},\\
\left(\frac{\partial}{\partial t} - \eta\Delta_s\right)\overline{a} &= \eta_{\text{turb}}\Delta_s\overline{a} + \delta\eta_{\text{turb}}\Delta_s\overline{\psi},
\end{aligned} \tag{4.8}$$

where ν_{turb} and η_{turb} are given by Eqs. (4.3), whereas

$$\begin{aligned}
\delta\nu_{\text{turb}} &= -\frac{\langle (\psi^{(0)}a^{(0)})\rangle}{2\nu},\\
\delta\eta_{\text{turb}} &= \frac{\langle (\psi^{(0)}a^{(0)})\rangle}{2}\left(\frac{1}{\nu} - \frac{1}{\eta}\right).
\end{aligned} \tag{4.9}$$

Since we want to illustrate the role of cross-correlations, we consider the simplest case $\nu=\eta$. This is the usual assumption in numerical simulations; see, e.g., Ref. 28. Equations (4.8) then take the form of Eqs. (3.5). In this case, both large-scale fields grow if $\eta+\eta_{\text{turb}}<0$, regardless of the sign of the cross-correlations $\langle (\psi^{(0)}a^{(0)})\rangle$, just as in the case of δ-correlated sources; see Sec. 3.

5. STOCHASTIC ANALOGS OF KOLMOGOROV FLOW FOR MAGNETOHYDRODYNAMICS

In this Section we consider specific examples of anisotropic time-independent fields, which can be regarded as magnetohydrodynamic stochastic analogs of Kolmogorov flow. In particular, we choose $\psi^{(0)}$ and $a^{(0)}$ to be

$$\begin{aligned}
\psi^{(0)}(\mathbf{x},t) &= A_1\cos(\mathbf{k}_0\mathbf{x}+\alpha),\\
a^{(0)}(\mathbf{x},t) &= A_2\cos(\mathbf{k}_0\mathbf{x}+\varphi+\alpha),
\end{aligned} \tag{5.1}$$

where A_1 and A_2 are the constant amplitudes of the zeroth-order stream function and magnetic potential, respectively; α is a random phase, uniformly distributed in $[0;2\pi]$; and φ is a constant phase. The spectral functions are

$$\begin{aligned}
\hat{C}_{\psi\psi}(\mathbf{k},\omega) &= 2\pi^3A_1^2\delta(\omega)[\delta(\mathbf{k}+\mathbf{k}_0)+\delta(\mathbf{k}-\mathbf{k}_0)],\\
\hat{C}_{aa}(\mathbf{k},\omega) &= 2\pi^3A_2^2\delta(\omega)[\delta(\mathbf{k}+\mathbf{k}_0)+\delta(\mathbf{k}-\mathbf{k}_0)],\\
\hat{C}_{\psi a}(\mathbf{k},\omega) &= 2\pi^3A_1A_2\delta(\omega)\{\cos\varphi[\delta(\mathbf{k}+\mathbf{k}_0)+\delta(\mathbf{k}\\
&\quad -\mathbf{k}_0)] + i\sin\varphi[\delta(\mathbf{k}+\mathbf{k}_0)-\delta(\mathbf{k}-\mathbf{k}_0)]\}.
\end{aligned} \tag{5.2}$$

It follows from Eqs. (5.1) and (5.2) that the zeroth-order fields so chosen are jointly homogeneous; that one is able to consider various forms of cross-correlations by varying φ; and that if there are different random phases (say, α and β) in Eqs. (5.1), then there are no cross-correlations. Thus, Eqs. (5.1) enable one to study a set of interesting consequences. Here we consider only the simple case,

$$\varphi=0, \quad \mathbf{k}_0 = k_0\mathbf{e}_x. \tag{5.3}$$

Inserting Eqs. (5.2) and (5.3) into Eqs. (2.12) and (2.13), we obtain

$$\begin{aligned}
\left(\frac{\partial}{\partial t} - \nu\Delta_s\right)\Delta_s\overline{\psi} &= -\left(\frac{A_1^2}{2\nu} - \frac{A_2^2}{2\eta}\right)\frac{\partial^2}{\partial Y^2}\Delta_s\overline{\psi}\\
&\quad + \left(\frac{4A_1^2}{\nu} - \frac{2A_2^2}{\eta}\right)\frac{\partial^4\overline{\psi}}{\partial X^2\partial Y^2} - A_1A_2\\
&\quad \times\left(\frac{4}{\nu} - \frac{2}{\eta}\right)\frac{\partial^4\overline{a}}{\partial X^2\partial Y^2} + \frac{A_1A_2}{2}\\
&\quad \times\left(\frac{1}{\nu} - \frac{1}{\eta}\right)\frac{\partial^2}{\partial Y^2}\Delta_s\overline{a},\\
\left(\frac{\partial}{\partial t} - \eta\Delta_s\right)\overline{a} &= \frac{A_1A_2}{2}\left(\frac{1}{\nu} - \frac{1}{\eta}\right)\frac{\partial^2\overline{\psi}}{\partial Y^2} + \left(\frac{A_1^2}{2\eta} - \frac{A_2^2}{2\nu}\right)\frac{\partial^2\overline{a}}{\partial Y^2}.
\end{aligned} \tag{5.4}$$

Examples resulting from subsequent simplifications are as follows.

1. $A_2=0$. There is no small-scale magnetic field. Equations (5.4) reduce to

$$\left(\frac{\partial}{\partial t} - \nu\Delta_s\right)\Delta_s\overline{\psi} = -\frac{A_1^2}{2\nu}\frac{\partial^2}{\partial Y^2}\Delta_s\overline{\psi} + \frac{4A_1^2}{\nu}\frac{\partial^4\overline{\psi}}{\partial X^2\partial Y^2},$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta_s\right)\overline{a}=\frac{A_1^2}{2\eta}\frac{\partial^2\overline{a}}{\partial Y^2}. \tag{5.5}$$

It follows from Eqs. (5.5) that the ''turbulent'' magnetic viscosity is always positive, and thus no large-scale magnetic field is generated. The growth rate for large-scale perturbations of the stream function is

$$\gamma=-\nu K^2+\frac{A_1^2}{2\nu}K_y^2-\frac{4A_1^2}{\nu}\frac{K_x^2K_y^2}{K^2}, \tag{5.6}$$

where **K** is the wave vector of large-scale perturbations, $K^2=K_x^2+K_y^2$. The growth rate (5.6) naturally coincides with that of Kolmogorov flow in an ordinary fluid; the latter has been calculated via the multi-scales expansion technique of Ref. 14. The most ''dangerous'' (that is, those that appear first when the amplitude of the small-scale field increases) are the large-scale perturbations with a wave vector, perpendicular to that of the small-scale field, i.e., $K_x=0$. Such perturbations grow if

$$A_1>A_{1\min}=\sqrt{2}\,\nu. \tag{5.7}$$

The criterion (5.7) is derived for an ordinary fluid in Ref. 11 for a regular small-scale velocity field.

2. $A_1=0$. There is no small-scale velocity field. Equations (5.4) take the form

$$\left(\frac{\partial}{\partial t}-\nu\Delta_s\right)\Delta_s\overline{\psi}=-\frac{2A_2^2}{\eta}\frac{\partial^4\overline{\psi}}{\partial X^2\partial Y^2}+\frac{A_2^2}{2\eta}\frac{\partial^2}{\partial Y^2}\Delta_s\overline{\psi},$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta_s\right)\overline{a}=-\frac{A_2^2}{2\nu}\frac{\partial^2\overline{a}}{\partial Y^2}. \tag{5.8}$$

The second equation signals the onset of negative magnetic viscosity. The most dangerous are large-scale magnetic field perturbations with a wave vector perpendicular to that of the small-scale field, $K_x=0$. The instability criterion is then

$$A_2>A_{2\min}=\sqrt{2\nu\eta}. \tag{5.9}$$

A large-scale velocity field with only a K_y-component does not grow. However, it can be shown that $\overline{\psi}$ grows if its wave vector makes an acute angle with the x axis. This special case is studied in detail in Ref. 32 using multiple-scale methods.

3. $A_1=A_2=A$. We note that equal amplitudes (or energies) of the magnetic and velocity fields are frequently chosen at $t=0$ in numerical simulations of freely decaying magnetohydrodynamic turbulence; see Refs. 27 and 28. Taking $\overline{\psi}$, $\overline{a}$ in the form

$$\overline{\psi}=R\exp(\gamma t+iK_xX+iK_yY),$$

$$\overline{a}=P\exp(\gamma t+iK_xX+iK_yY), \tag{5.10}$$

and inserting Eqs. (5.10) into Eqs. (5.4), we obtain a linear system with unknown P and R. The equation for the growth rate then follows by equating the determinant to zero:

$$a\gamma^2+b\gamma+c=0, \tag{5.11}$$

where

$$a=K^2,$$

$$b=(\nu+\eta)K^4-2\Lambda K^2K_y^2+\Lambda_1K_x^2K_y^2,$$

$$c=\nu\eta K^6-\Lambda(\nu+\eta)K^4K_y^2+\Lambda_1\eta K^2K_x^2K_y^2,$$

$$\Lambda=\frac{A^2}{2}\left(\frac{1}{\nu}-\frac{1}{\eta}\right),\quad \Lambda_1=2A^2\left(\frac{2}{\nu}-\frac{1}{\eta}\right).$$

Introducing polar coordinates

$$K_x=K\cos\theta,\quad K_y=K\sin\theta$$

we obtain an equation for the neutral curve $\gamma=0$ in terms of A, θ:

$$\frac{A^2}{2\nu^2}[1-(\mathrm{Pr}_m)^2-4(2-\mathrm{Pr}_m)\cos^2\theta]\sin^2\theta=1, \tag{5.12}$$

where $\mathrm{Pr}_m=\nu/\eta$ is the magnetic Prandtl number. Instability is possible for

$$\mathrm{Pr}_m<1$$

and

$$|\cos\theta|<\frac{1}{2}\left[\frac{1-(\mathrm{Pr}_m)^2}{2-\mathrm{Pr}_m}\right]^{1/2}.$$

The most dangerous perturbations are those with $K_x=0$ [as can be seen by comparing $A^2(\theta)$ with $A^2(\pi/2)$ estimated from Eq. (5.12)]. In this case, the instability criterion is

$$A>A_{\min}=\frac{\sqrt{2}\nu}{\sqrt{1-(\mathrm{Pr}_m)^2}}, \tag{5.13}$$

whereas the growth rate takes the form

$$\gamma=\left\{\Lambda-\frac{\nu+\eta}{2}+\sqrt{\Lambda^2+\frac{(\nu-\eta)^2}{4}}\right\}K^2. \tag{5.14}$$

Because both large-scale fields increase, this case is of interest for subsequent nonlinear analysis and numerical simulation, which will be the subject of future research.

6. RESULTS

In this paper, in the framework of 2D magnetohydrodynamics, we have studied the possible occurrence of large-scale mean velocity and magnetic fields generated by small-scale random fields. The latter are assumed jointly stationary, homogeneous, and maintained by an external source.

The random fields lead to negative dissipative factors in the equations for the mean fields, which is why the term ''negative viscosity'' is used. Viscous damping of large-scale fields is thus replaced by growth, which is limited due to nonlinear effects in the amplitudes of the large-scale fields. This picture, being simplified, though, is fruitful for studying the effects of smaller-scale fields on large-scale ones.

Our results are as follows.

1. Using a two-scale expansion, we obtain equations, that describe the evolution of the mean stream function and the mean magnetic potential in the presence of small-scale MHD fluctuations. These expressions enable us to study the

evolution of large-scale MHD perturbations with the assumption that the statistical properties of the small-scale fields are known.

2. With our approach we easily demonstrate how a negative magnetic viscosity term can appear in the equation for the mean magnetic potential.

3. The general expressions also enable us to recover previous results on the eddy (turbulent) viscosity of the ordinary fluid and on turbulent viscosities in the presence of small-scale MHD fluctuations.

4. For isotropic small-scale fluctuations, we estimate turbulent viscosities and find criteria for the onset of negative magnetic viscosity, as well as for the growth of large-scale MHD fields, in three cases:

(i) fluctuations are generated by white noise source;

(ii) fluctuations possess long correlation times;

(iii) fluctuations possess short correlation times (in comparison with characteristic dissipation times associated with molecular kinematic and magnetic viscosities).

In particular, it is shown when the cross-correlations among small-scale fields vanish, the turbulent viscosity is always positive, whereas the turbulent magnetic viscosity can be negative, thus giving rise to the growth of large-scale magnetic perturbations. When cross-correlations are nonvanishing, both large-scale fields can be amplified.

5. We also consider how large-scale fields are influenced by anisotropic small-scale random fields, which can be regarded as stochastic analogs of Kolmogorov flow. We find that

(i) if there is only a small-scale velocity field, the growth rate of the large-scale velocity field corresponds to that of a Kolmogorov flow, whereas no magnetic field is generated;

(ii) if there is a small-scale magnetic field only, then the large-scale field increases fastest for perturbations transverse to the small-scale anisotropic ones;

(iii) finally, if the random anisotropic fields are of equal amplitude, then both large-scale fields grow; again, the growth rate is greatest for large-scale perturbations transverse to the small-scale ones.

This paper was supported by the ''Chaos-2'' Project of the National Academy of Sciences of Ukraine, and by Project INTAS 93-1194. Information support provided by Project INTAS LA-96-09 is also acknowledged.

APPENDIX

Derivation of Q_1, Q_2, Q_3

In this Appendix, the terms Q_1, Q_2, Q_3 [see Eqs. (2.7)–(2.9)] are expressed in terms of the mean components $\overline{\psi}$, $\overline{a}$ and space-time spectral functions of the fluctuating components ψ^T, a^T obtained from Eqs. (2.10) and (2.11). Taking the approach outlined in Sec. 2, we introduce slow and fast spatial variables $\mathbf{X}$ and $\mathbf{x}$. The spatial operators are then written in the form

$$\nabla \to \frac{\partial}{\partial \mathbf{x}} + \frac{\partial}{\partial \mathbf{X}}, \quad \Delta \to \Delta + 2\frac{\partial^2}{\partial x_p \partial X_p} + \Delta_s,$$

$$\Delta^2 \to \Delta^2 + 2\Delta\Delta_s + 4\frac{\partial^2}{\partial x_p \partial X_p}\Delta + 4\frac{\partial^2}{\partial x_p \partial X_p}\Delta_s + 4\frac{\partial^4}{\partial x_p \partial X_p \partial x_r \partial X_r}. \tag{A1}$$

We start by deriving Q_3 since it has a simpler form than Q_1 and Q_2. According to Eqs. (2.9) and (A1), this term can be written

$$Q_3 = -\langle [\nabla \psi^T \times \nabla a^T]_z \rangle = Q_3^{(00)} + Q_3^{(01)} + Q_3^{(10)} + Q_3^{(11)} + Q_3^{(02)} + Q_3^{(20)} + O(K^3, K^4 \ldots), \tag{A2}$$

where

$$Q_3^{(00)} = -\varepsilon_{mn}\left\langle \frac{\partial \psi^{(0)}}{\partial x_m}\frac{\partial a^{(0)}}{\partial x_n}\right\rangle = 0$$

due to homogeneity of the turbulence,

$$Q_3^{(01)} = -\varepsilon_{mn}\left\langle \frac{\partial \psi^{(0)}}{\partial x_m}\left(\frac{\partial}{\partial x_n} + \frac{\partial}{\partial X_n}\right) a^{(1)}\right\rangle,$$

$$Q_3^{(10)} = -\varepsilon_{mn}\left\langle \left(\frac{\partial}{\partial x_m} + \frac{\partial}{\partial X_m}\right)\psi^{(1)}\frac{\partial a^{(0)}}{\partial x_n}\right\rangle, \tag{A3}$$

and the remaining terms in Eq. (A2) have a similar structure, which is now obvious. We retain only those terms in Eq. (A2), that are of order K, K^2. As will be seen below, it is just these terms that give rise to negative magnetic viscosity.

To calculate the terms in Eq. (A2), it is necessary to derive expressions for $\psi^{(i)}$ and $a^{(i)}$, $i=0,1,2$. In the zeroth approximation, Eqs. (2.10) and (2.11) yield

$$\left(\frac{\partial}{\partial t} - \nu\Delta\right)\Delta\psi^{(0)} = F_\psi(\mathbf{x},t),$$

$$\left(\frac{\partial}{\partial t} - \eta\Delta\right)a^{(0)} = F_a(\mathbf{x},t). \tag{A4}$$

Before solving the equations in various orders of approximation, we introduce correlation functions and Fourier spectra of the zero-order fields. Since we assume joint homogeneity and stationarity of the small-scale fields, we have

$$C_{\psi\psi}(\mathbf{x}-\mathbf{x}', t-t') \equiv \langle \psi^{(0)}(\mathbf{x},t)\psi^{(0)}(\mathbf{x}',t')\rangle,$$

$$C_{aa}(\mathbf{x}-\mathbf{x}', t-t') \equiv \langle a^{(0)}(\mathbf{x},t)a^{(0)}(\mathbf{x}',t')\rangle, \tag{A5}$$

$$C_{\psi a}(\mathbf{x}-\mathbf{x}', t-t') \equiv \langle \psi^{(0)}(\mathbf{x},t)a^{(0)}(\mathbf{x}',t')\rangle.$$

Using the Fourier transform over the fast variables $\mathbf{x}$, t,

$$\psi^{(0)}(\mathbf{x},t) = \int \frac{d\mathbf{k}d\omega}{(2\pi)^3}\hat{\psi}^{(0)}(\mathbf{k},\omega)\ \exp(-i\omega t + i\mathbf{k}\mathbf{x}), \tag{A6}$$

the corresponding space-time spectral functions are defined as

$$C_{\psi a}(\boldsymbol{\kappa},\tau) = \int \frac{d\mathbf{k}d\omega}{(2\pi)^3}\hat{C}^{(0)}_{\psi a}(\mathbf{k},\omega)\ \exp(-i\omega\tau + i\mathbf{k}\boldsymbol{\kappa}). \tag{A7}$$

Since the small-scale fluctuations are stationary and homogeneous,

$$\langle \hat{\psi}^{(0)}(\mathbf{k},\omega)\hat{a}^{(0)}(\mathbf{k}',\omega')\rangle = (2\pi)^3\delta(\omega+\omega')\,\delta(\mathbf{k}+\mathbf{k}')\hat{C}_{\psi a}(\mathbf{k},\omega). \tag{A8}$$

The properties of the spectral functions are

$$\text{(i)}\quad \hat{C}^*_{\psi\psi}(\mathbf{k},\omega)=\hat{C}_{\psi\psi}(\mathbf{k},\omega)=\hat{C}_{\psi\psi}(-\mathbf{k},-\omega),$$

$$\text{(ii)}\quad \hat{C}^*_{\psi a}(\mathbf{k},\omega)=\hat{C}_{\psi a}(-\mathbf{k},-\omega)=\hat{C}_{a\psi}(\mathbf{k},\omega), \tag{A9}$$

where the asterisk denotes the complex conjugate. The properties (ii) stem from the reality of $C_{\psi a}(\boldsymbol{\kappa},\tau)$ and the condition

$$C_{\psi a}(\boldsymbol{\kappa},\tau)=C_{a\psi}(-\boldsymbol{\kappa},-\tau).$$

It is also useful to introduce the spectra of the fields $\mathbf{v}^{(0)}(x,t)$ and $\mathbf{B}^{(0)}(x,t)$ for the Fourier components

$$\hat{v}_i(\mathbf{k},\omega)=-i\varepsilon_{ij}k_j\hat{\psi}(\mathbf{k},\omega),\quad \hat{B}_i(\mathbf{k},\omega)=-i\varepsilon_{ij}k_j\hat{a}(\mathbf{k},\omega), \tag{A10}$$

where ε_{ij} is the unit antisymmetric tensor of the second rank. Since

$$\langle \hat{v}_i^{(0)}(\mathbf{k},\omega)\hat{v}_j^{(0)}(\mathbf{k}',\omega')\rangle = (2\pi)^3\delta(\omega+\omega')\,\delta(\mathbf{k}+\mathbf{k}')\langle v_i^{(0)}v_j^{(0)}\rangle_{\mathbf{k},\omega}, \tag{A11}$$

where $\langle v_i^{(0)}v_j^{(0)}\rangle_{\mathbf{k},\omega}$ is the space-time spectral tensor of the zero-order velocity field, we obtain with the help of Eqs. (A10) and (A11)

$$\langle(\mathbf{v}_0)^2\rangle_{\mathbf{k},\omega}=k^2\hat{C}_{\psi\psi}(\mathbf{k},\omega),$$

$$\langle(\mathbf{B}^{(0)})^2\rangle_{\mathbf{k},\omega}=k^2\hat{C}_{aa}(\mathbf{k},\omega),$$

$$\langle(\mathbf{v}^{(0)}\mathbf{B}^{(0)})\rangle_{\mathbf{k},\omega}=k^2\hat{C}_{\psi a}(\mathbf{k},\omega). \tag{A12}$$

Now we return to Eqs. (2.10) and (2.11). In the first approximation we have

$$\left(\frac{\partial}{\partial t}-\nu\Delta\right)\Delta\psi^{(1)}+\varepsilon_{mn}\frac{\partial\overline{\psi}}{\partial X_m}\frac{\partial}{\partial x_n}\Delta\psi^{(0)}=\varepsilon_{mn}\frac{\partial\overline{a}}{\partial X_m}\frac{\partial}{\partial x_n}\Delta a^{(0)},$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta\right)a^{(1)}+\varepsilon_{mn}\frac{\partial\psi}{\partial x_m}\frac{\partial\overline{a}}{\partial X_n}+\varepsilon_{mn}\frac{\partial\overline{\psi}}{\partial X_m}\frac{\partial a^{(0)}}{\partial x_n}=0. \tag{A13}$$

Taking the Fourier transform, we obtain

$$\hat{\psi}^{(1)}(\mathbf{k},\omega)=\varepsilon_{mn}\frac{ik_n}{-i\omega+\nu k^2}\times\left(\frac{\partial\overline{a}}{\partial X_m}\hat{a}^{(0)}(\mathbf{k},\omega)-\frac{\partial\overline{\psi}}{\partial X_m}\hat{\psi}^{(0)}(\mathbf{k},\omega)\right),$$

$$\hat{a}^{(1)}(\mathbf{k},\omega)=\varepsilon_{mn}\frac{ik_n}{-i\omega+\eta k^2}\times\left(\frac{\partial\overline{a}}{\partial X_m}\hat{\psi}^{(0)}(\mathbf{k},\omega)-\frac{\partial\overline{\psi}}{\partial X_m}\hat{a}^{(0)}(\mathbf{k},\omega)\right). \tag{A14}$$

In the second approximation we have from Eqs. (2.10) and (2.11)

$$\left(\frac{\partial}{\partial t}-\nu\Delta\right)\Delta\psi^{(2)}+2\frac{\partial^3\psi^{(1)}}{\partial t\partial x_p\partial X_p}-4\nu\frac{\partial^2}{\partial x_p\partial X_p}\Delta\psi^{(1)}+\varepsilon_{mn}\frac{\partial\overline{\psi}}{\partial X_m}\frac{\partial}{\partial x_n}\Delta\psi^{(1)}=\varepsilon_{mn}\frac{\partial\overline{a}}{\partial X_m}\frac{\partial}{\partial x_n}\Delta a^{(1)},$$

$$\left(\frac{\partial}{\partial t}-\eta\Delta\right)a^{(2)}-2\eta\frac{\partial^2a^{(1)}}{\partial x_p\partial X_p}+\varepsilon_{mn}\frac{\partial\psi^{(1)}}{\partial x_m}\frac{\partial\overline{a}}{\partial X_n}+\varepsilon_{mn}\frac{\partial\overline{\psi}}{\partial X_m}\frac{\partial a^{(1)}}{\partial x_n}=0. \tag{A15}$$

Taking the Fourier transform, we obtain

$$\hat{\psi}^{(2)}(\mathbf{k},\omega)=2\frac{\omega k_p}{k^2}\frac{1}{-i\omega+\nu k^2}\frac{\partial\hat{\psi}^{(1)}}{\partial X_p}+\frac{4i\nu k_p}{-i\omega+\nu k^2}\frac{\partial\hat{\psi}^{(1)}}{\partial X_p}-\varepsilon_{mn}\frac{ik_n}{-i\omega+\nu k^2}\left(\frac{\partial\overline{\psi}}{\partial X_m}\hat{\psi}^{(1)}-\frac{\partial\overline{a}}{\partial X_m}\hat{a}^{(1)}\right),$$

$$\hat{a}^{(2)}(\mathbf{k},\omega)=\frac{2i\eta k_p}{-i\omega+\eta k^2}\frac{\partial\hat{a}^{(1)}}{\partial X_p}+\varepsilon_{mn}\frac{ik_m}{-i\omega+\eta k^2}\times\left(\frac{\partial\overline{\psi}}{\partial X_n}\hat{a}^{(1)}-\frac{\partial\overline{a}}{\partial X_n}\hat{\psi}^{(1)}\right). \tag{A16}$$

We do not calculate terms of third and fourth order, since they do not contribute to $Q_{1,2,3}$, as will be seen below. Furthermore, since we are interested in negative viscosity effects (linear in the mean quantities $\overline{\psi},\overline{a}$), we neglect terms nonlinear in $\overline{\psi},\overline{a}$ in $\hat{\psi}^{(2)},\hat{a}^{(2)}$, namely, the last terms on the right-hand side of Eqs. (A16). These can be taken into account in the same manner as the linear terms, and this was done for the more straightforward case in Ref. 20. However, in this paper we do not consider nonlinear effects in the mean quantities.

Now we are ready to calculate all terms in Eq. (A2). Note that the term $Q_3^{(11)}$ need not be taken into account because it is nonlinear in $\overline{\psi}$ and $\overline{a}$. Then, it can easily be seen by explicitly writing the terms $Q_3^{(20)}$ and $Q_3^{(02)}$ that they yield zero to the second order inclusive. Therefore, only the sum $Q_3^{(01)}+Q_3^{(10)}$ has to be evaluated; see Eqs. (A3). Using Eqs. (A14) and the properties (A5)–(A9) of the zero-order spectral functions, we obtain to order K^2:

$$Q_3^{(01)}=-\varepsilon_{kl}\varepsilon_{mn}\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{k_lk_m}{-i\omega+\eta k^2}\times\left\{\hat{C}_{\psi\psi}\frac{\partial^2\overline{a}}{\partial X_k\partial X_n}-\hat{C}_{a\psi}\frac{\partial^2\overline{\psi}}{\partial X_k\partial X_n}\right\},$$

$$Q_3^{(10)}=-\varepsilon_{kl}\varepsilon_{mn}\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{k_lk_n}{-i\omega+\nu k^2}\times\left\{\hat{C}_{aa}\frac{\partial^2\overline{a}}{\partial X_k\partial X_m}-\hat{C}_{\psi a}\frac{\partial^2\overline{\psi}}{\partial X_k\partial X_m}\right\},$$

and then obtain Eq. (2.13) with the right hand side being the sum $Q_3^{(01)}+Q_3^{(10)}$.

Now we calculate

$$Q_1=-\langle[\nabla\psi^T\times\nabla\Delta\psi^T]_z\rangle=Q_1^{(00)}+Q_1^{(01)}+Q_1^{(10)}+Q_1^{(02)}+Q_1^{(20)}+Q_1^{(03)}+Q_1^{(30)}+Q_1^{(04)}+Q_1^{(40)}+Q_1^{(11)}+Q_1^{(12)}+Q_1^{(21)}+Q_1^{(13)}+Q_1^{(31)}+O(K^5,K^6,\ldots), \tag{A17}$$

where

$$Q_1^{(00)}=-\varepsilon_{mn}\left\langle\frac{\partial\psi^{(0)}}{\partial x_m}\frac{\partial}{\partial x_n}\Delta\psi^{(0)}\right\rangle=0$$

due to homogeneity of the turbulence, and

$$Q_1^{(01)}=-\varepsilon_{mn}\left\langle\frac{\partial\psi^{(0)}}{\partial x_m}\left(\frac{\partial}{\partial x_n}+\frac{\partial}{\partial X_m}\right)\times\left(\Delta+2\frac{\partial^2}{\partial x_p\partial X_p}+\Delta_s\right)\psi^{(1)}\right\rangle,$$

$$Q_1^{(10)}=-\varepsilon_{mn}\left\langle\left(\frac{\partial}{\partial x_n}+\frac{\partial}{\partial X_m}\right)\psi^{(1)}\frac{\partial}{\partial x_n}\Delta\psi^{(0)}\right\rangle. \tag{A18}$$

The remaining terms in Eq. (A17) have similar structure. We retain only terms of order $K^1,\ldots,K^4$. We naturally expect that among these are terms that give rise to the effective viscosity term in the equation for the mean flow.

As in the case of Q_3, we omit all nonlinear terms in $\overline{\psi},\overline{a}$, which can be taken into account on the same basis as the linear terms. Then, starting with expressions similar to those in Eq. (A18) it can be easily verified that $Q_1^{(03)}+Q_1^{(30)}=0$ to order K^4; the same conclusion holds $Q_1^{(04)}+Q_1^{(40)}$. Thus, we only need to calculate

$$Q_1^{(01)}+Q_1^{(10)}+Q_1^{(02)}+Q_1^{(20)}.$$

It is convenient to divide this sum into two terms, $Q_1^{(01)}+Q_1^{(10)}$ and $Q_1^{(02)}+Q_1^{(20)}$, and evaluate them separately. Using Eqs. (A14) and the properties (A5)–(A9), we have for the first sum

$$Q_1^{(01)}+Q_1^{(10)}=\varepsilon_{kl}\varepsilon_{mn}\left\{\delta_{1lnp}^{(1)}\frac{\partial^3\overline{\psi}}{\partial X_k\partial X_m\partial X_p}+\delta_{1lnp}^{(2)}\frac{\partial^3\overline{a}}{\partial X_k\partial X_m\partial X_p}+\nu_{1ln}^{(1)}\frac{\partial^2}{\partial X_k\partial X_m}\Delta_s\overline{\psi}+\nu_{1ln}^{(2)}\frac{\partial^2}{\partial X_k\partial X_m}\Delta_s\overline{a}\right\}, \tag{A19}$$

where

$$\delta_{1lnp}^{(1)}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{2ik_lk_nk_p}{-i\omega+\nu k^2}\hat{C}_{\psi\psi},$$

$$\delta_{1lnp}^{(2)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{2ik_lk_nk_p}{-i\omega+\nu k^2}\hat{C}_{a\psi},$$

$$\nu_{1ln}^{(1)}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{k_lk_n}{-i\omega+\nu k^2}\hat{C}_{\psi\psi},$$

$$\nu_{1ln}^{(2)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{k_lk_n}{-i\omega+\nu k^2}\hat{C}_{a\psi}.$$

Using Eqs. (A16) and the properties of Eqs. (A5)–(A9), we also obtain

$$Q_1^{(02)}+Q_1^{(20)}=\varepsilon_{kl}\varepsilon_{mn}\left\{\nu_{1lnpr}^{(1)}\frac{\partial^4\overline{\psi}}{\partial X_k\partial X_m\partial X_p\partial X_r}+\nu_{1lnpr}^{(2)}\frac{\partial^4\overline{a}}{\partial X_k\partial X_m\partial X_p\partial X_r}\right\}, \tag{A20}$$

where

$$\nu_{1lnpr}^{(1)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\left\{\frac{1}{-i\omega+\nu k^2}+\frac{\nu k^2}{(-i\omega+\nu k^2)^2}\right\}\hat{C}_{\psi\psi},$$

$$\nu_{1lnpr}^{(2)}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\left\{\frac{1}{-i\omega+\nu k^2}+\frac{\nu k^2}{(-i\omega+\nu k^2)^2}\right\}\hat{C}_{a\psi}.$$

Q_2 is evaluated in exactly the same way as Q_1. As a result, terms with $\delta_{2lnp}^{(1)}$, $\delta_{2lnp}^{(2)}$, $\nu_{2ln}^{(1)}$, $\nu_{2ln}^{(2)}$ appear in the sum $Q_2^{(01)}+Q_2^{(10)}$, which differ from their counterparts in Eq. (A19) by the interchange of ν and η, $\hat{C}_{\psi\psi}$ and $-\hat{C}_{aa}$, and $-\hat{C}_{a\psi}$ and $-\hat{C}_{\psi a}$. Then, for the sum $Q_2^{(02)}+Q_2^{(20)}$ we obtain

$$Q_2^{(02)}+Q_2^{(20)}=\varepsilon_{kl}\varepsilon_{mn}\left\{\nu_{2lnpr}^{(1)}\frac{\partial^4\overline{\psi}}{\partial X_k\partial X_m\partial X_p\partial X_r}+\nu_{2lnpr}^{(2)}\frac{\partial^4\overline{a}}{\partial X_k\partial X_m\partial X_p\partial X_r}\right\}, \tag{A21}$$

where

$$\nu_{2lnpr}^{(1)}=-\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\frac{\eta k^2}{(-i\omega+\eta k^2)^2}\hat{C}_{aa},$$

$$\nu_{2lnpr}^{(2)}=\int\frac{d\mathbf{k}d\omega}{(2\pi)^3}\frac{4k_lk_nk_pk_r}{k^2}\frac{\eta k^2}{(-i\omega+\eta k^2)^2}\hat{C}_{\psi a}.$$

Finally, summing the terms $Q_1^{(01)}+Q_1^{(10)}$, $Q_2^{(01)}+Q_2^{(10)}$, $Q_1^{(02)}+Q_1^{(20)}$, and $Q_2^{(02)}+Q_2^{(20)}$, and introducing $\delta_{lnp}^{(1)}=\delta_{1lnp}^{(1)}+\delta_{2lnp}^{(1)}$, $\delta_{lnp}^{(2)}=\delta_{1lnp}^{(2)}+\delta_{2lnp}^{(2)}$, $\nu_{ln}^{(1)}=\nu_{1ln}^{(1)}+\nu_{2ln}^{(1)}$, etc., we obtain the final expression for the right-hand side of Eq. (2.12).

*)E-mail: chechkin@ipp.kharkov.ua

[1] E. N. Lorenz, *The Nature and Theory of the General Circulation of the Atmosphere*, World Meterological Organization, Geneva (1967).
[2] V. P. Starr, *Physics of Negative Viscosity Phenomena*, McGraw Hill, New York (1968).
[3] Yu. L. Klimontovich, *Statistical Physics*, Nauka, Moscow (1982) [Engl. transl.: Harwood Academic Publishers, New York (1986)].
[4] L. D. Landau and E. M. Lifshitz, *Hydrodynamics* [in Russian], Nauka, Moscow (1986).

[5] S. Corrsin, in *Turbulent Diffusion in Environmental Pollution*, Pergamon Press, New York (1974), vol. 18A, p. 25.
[6] S. Eskinazi and F. F. Erian, Phys. Fluids **12**, 1988 (1969).
[7] A. F. Kurbatskii, *Modelling of Nonlocal Momentum and Heat Transport*, Nauka, Novosibirsk (1988).
[8] R. H. Kraichnan, J. Atmos. Sci. **33**, 1521 (1976).
[9] A. Pouquet, J. Fluid Mech. **88**, 1 (1978).
[10] V. I. Arnold and L. D. Meshalkin, Usp. Mat. Nauk **15**, 247 (1960).
[11] L. D. Meshalkin and Ya. G. Sinai, Prikl. Mat. Mekh. **25**, 1140 (1961).
[12] A. Nepomnyaschii, Ibid. **40**, 886 (1976).
[13] G. I. Sivashinsky, Physica D **17**, 243 (1985).
[14] B. Dubrulle and U. Frisch, Phys. Rev. A **43**, 5355 (1991).
[15] E. Weimann and C.-W. Shu, Phys. Fluids A **5**, 998 (1993).
[16] N. F. Bondarenko, M. Z. Gak, and F. V. Dolzhaskii, Atmos. Oceanic Phys. **15**, 711 (1979).
[17] J. C. Williams, J. Fluid Mech. **146**, 21 (1984).
[18] A. V. Chechkin, A. V. Tur, and V. V. Yanovsky, Phys. Fluids B **4**, 3513 (1992).
[19] D. Montgomery and T. Hatory, Plasma Phys. Controlled Fusion **26**, 717 (1984).
[20] A. V. Chechkin, M. I. Kopp, A. V. Tur, and V. V. Yanovsky, Zh. Éksp. Teor. Fiz. **113**, 646 (1998) [JETP **86**, 357 (1998)].
[21] M. I. Shliomis and K. Morozov, Phys. Fluids **6**, 2855 (1984).
[22] J.-C. Bacri, R. Perzynski, M. I. Shliomis, and G. I. Burde, Phys. Rev. Lett. **75**, 21128 (1995).
[23] D. Biskamp and H. Welter, Phys. Lett. A **96**, 25 (1983).
[24] D. Biskamp, Plasma Phys. Controlled Fusion **26**, 311 (1984).
[25] H. R. Strauss, Phys. Fluids **19**, 134 (1976).
[26] D. Biskamp, Phys. Rep. **237**, 179 (1994).
[27] D. Biskamp and H. Welter, Phys. Fluids B **1**, 1964 (1989).
[28] R. Kinney and J. C. McWilliams, Phys. Plasmas **2**, 3623 (1995).
[29] M. Steenbeck, F. Krause, and K.-H. Radler, Z. Naturforsch. A **21**, 369 (1966).
[30] H. K. Moffat, *Magnetic Field Generation in Electrically Conducting Fluids*, Cambridge Univ. Press, Cambridge (1978).
[31] W. Horton, in *Handbook of Plasma Physics*, A. A. Galeev and R. N. Sudan (eds.), North-Holland, Amsterdam, Vol. II, p. 384 (1984).
[32] A. V. Chechkin, Ukr. J. Phys. No. 5 (1999).

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

Anomalous burn-through of thin foils by high-intensity laser radiation

V. V. Ivanov, A. V. Kutsenko, I. G. Lebo, A. A. Matsveĭko, Yu. A. Mikhaĭlov,*)
V. P. Osetrov, A. I. Popov, V. B. Rozanov, G. V. Sklizkov, and A. N. Starodub,

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924, Moscow, Russia

V. V. Nikishin and V. F. Tishkin

Institute of Mathematical Modeling, Russian Academy of Sciences, 125047 Moscow, Russia
(Submitted 1 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1287–1299 (October 1999)

The paper presents results of experiments performed on the Pico facility in which foils were heated by laser radiation, and anomalously fast burn-through of foils by a structured laser beam was detected. Comparison with two-dimensional calculations has allowed us to suggest a tentative mechanism for the effect under investigation. The targets in the experiments were thin aluminum foils of thickness 3 to 40 μm. The flux density of laser radiation on the target surface varied between 10^{13} and 10^{14} W/cm^2. We detected a strong dependence of the transmitted energy on the foil thickness and the shortening of the transmitted laser pulse. Penetration of laser radiation through foils with thicknesses considerably larger than 3 μm has been observed, although it was stated in earlier publications [V. V. Ivanov, A. K. Knyazev, A. V. Kutsenko, *et al.*, Kratk. Soobshch. Fiz. FIAN No. 7–8, 37 (1997)]; A. É. Bugrov, I. N. Burdonskii, V. V. Gol'tsov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 903 (1997) [JETP **84**, 903 (1997)] that, at the laser radiation parameters used in our experiment, the evaporated layer of the foil could not be thicker than 2 μm. Two-dimensional calculations have allowed us to interpret this effect in terms of local ''piercing'' of the target at spots on the target surface where the radiation intensity has its peaks. The possibility of reducing these peaks by using a symmetrizing prepulse is discussed in the paper. © *1999 American Institute of Physics.* [S1063-7761(99)01110-5]

1. INTRODUCTION

An important part of research into the interaction between laser radiation and matter is experiments in which a laser beam of high intensity is focused on a thin-foil target which can be fabricated from materials of differing densities and atomic numbers.

The primary goal of such experiments is investigation of heating and stability of compression of shells to be used in laser-driven thermonuclear fusion. Such experiments are often performed not only on spherical thin shells, but also on plane foils with a view to simulating various processes that are of great importance in thin-shell targets. It is remarkable that such experiments can be performed on facilities generating radiation of energy considerably lower than that required for spherically symmetrical heating of shell targets. Thus, one can study such complex and important processes as absorption and reflection of laser light by an ablator shell, simulated by a foil or a combination of foils of different compositions and thicknesses, burn-through of a target wall and development of instabilities, generation of x-rays and fast particles in targets, and, finally, acceleration of the target shell.[1,2]

An important advantage of relatively simple one-beam laser facilities in simulation of processes in thermonuclear targets is the possibility of controlling the variation of the laser spot pattern on the target, which is almost impossible in experiments in facilities utilizing multiple laser beams, where the pattern of laser radiation intensity on the target is largely determined by the interference between the laser beams, which usually have a high degree of coherence. In particular, this interference between coherent beams generates small-scale peaks of radiation intensity on target surfaces, which can cause catastrophically fast development of hydrodynamic instabilities disrupting the process of plasma compression. Single-beam experiments with foils simulating shell targets, in which the laser spot pattern is controllable, allow one to study effects of inhomogeneities on the heating and acceleration of foils and model processes in the development of instabilities in plasmas, and also to investigate techniques suggested for suppressing instabilities not only through variations in the spatial distribution of the laser field intensity, which is very difficult, even though it is possible, in the case of spherically symmetrical irradiation of a target, but mostly through changes in the time distribution of the laser radiation intensity or target configuration.

An important component of such research is the study of energy balance in a target irradiated by laser radiation, i.e., the relationships among the radiation incident on the foil surface, the radiation scattered by the target, and the transmitted radiation downstream of the target. This paper reports on experiments conducted in the Pico facility[1,3,4] designed for heating thin foils by laser radiation with an inhomogeneous far-field pattern (on the target surface), which allows one to detect anomalously fast burn-through of foils. After that, comparison between experimental data and theoretical

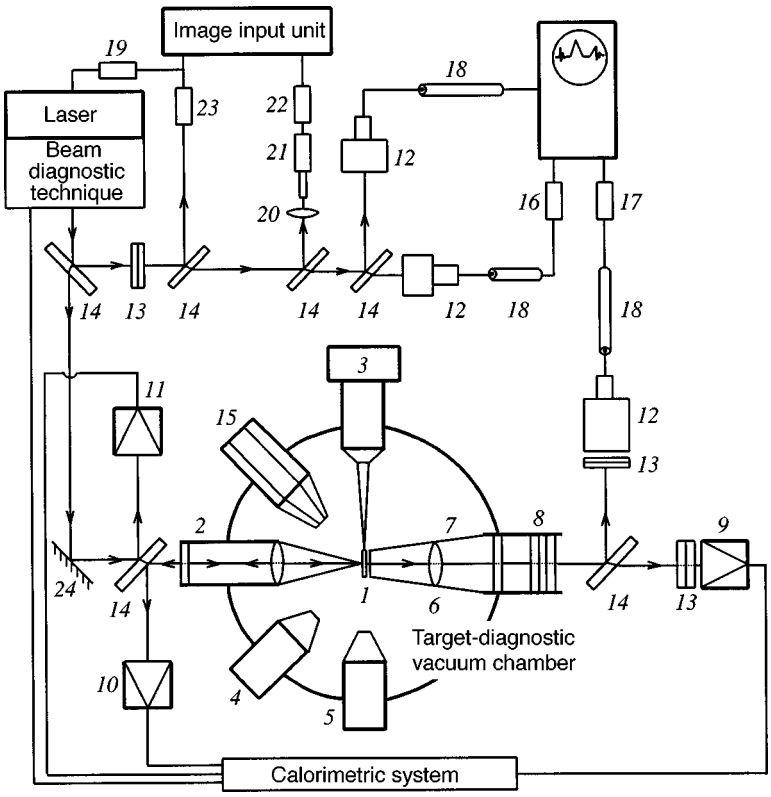


FIG. 1. Diagram of experimental facility: (*1*) target; (*2*) focusing system; (*3*) stage of target alignment; (*4*), (*5*) system for monitoring target position at the focal point; (*6*), (*7*), and (*8*) system conducting transmitted radiation; (*9*), (*10*), and (*11*) calorimeters; (*12*) coaxial photocells; (*13*) optical filters; (*14*) beam splitters; (*15*) multichannel X-ray microscope; (*16*) and (*17*) fast attenuators; (*18*) delay lines for matching pulses from photodiodes; (*19*) frame synchronization unit; (*20*) objective lens forming far-field image; (*21*) microscope; (*22*) and (*23*) CCD videocameras; (*24*) package of full-reflection mirrors.

calculations enables one to identify plausible mechanisms of the observed effects and to suggest methods for limiting the impact of inhomogeneities in the radiation intensity on the process of burn-through detected in experiments.

2. EXPERIMENTAL FACILITY

A schematic of the experimental facility is shown in Fig. 1. The targets exposed to laser radiation are fabricated from aluminum foils. The foil thickness varies between 3 to 40 μm. The targets are irradiated by a beam from a Nd glass laser generating nanosecond pulses. The pressure of the residual gas (air) on the target ranges between 10^{-2} and $6\cdot 10^{-6}$ Torr. The FWHM of the laser pulse is 2 ns. The output laser pulse energy varies between 2–20 J, which allows us to obtain a power density on the target surface from 10^{13} to 10^{14} W/cm^2. The laser beam divergence is $2\alpha=(5-8)\cdot 10^{-4}$ rad, and the energy contrast ratio is $K_E=10^4-10^5$. The FWHM of the spectral line is $\delta\lambda=30$ Å .

For these experiments on the Pico facility designed for simulation of processes in thermonuclear targets, we have developed a special technique for studying laser plasma generated in heating foils. The facility includes a system for accumulation and processing of data on the laser energy distribution in space and time in the zone of interaction. The system is based on a set of calorimetric devices and coaxial photoelectric cells and an automatic system monitoring the pattern of the laser spot on the target surface. The diagnostic technique allows us to monitor concurrently the energy balance in a target and the process of foil burn-through with coordinate and time resolution, along with the pattern of the laser spot on the target surface. This combination of diagnostic devices ensures correct interpretation of experimental data.

The calorimetry employs a multichannel system of the Pico facility monitoring the laser radiation energy[5,6] and based on using the long thermal time constant of calorimetric sensors for intermediate storage of detected signals and subsequent readout by a common measuring device. The laser radiation detectors are five calorimeters VChD-2, VChD-3, VChD-5, VKDS, and KDS (developed by the Design Bureau of the Lebedev Institute of Physics) operating in the energy range 10^{-5} to 10^2 J with an uncertainty in absolute measurements of 6–8%. The total number of calorimetric channels used in this technique is 18 (eight on the path of the laser beam and ten in the vacuum chamber of the tested target). The calorimeters placed in the chamber (in particular, those labeled by *9*, *10*, and *11* in Fig. 1) are used for measuring the energy distribution in the far-field pattern of the laser beam making the plasma and for measuring the intensity of incident radiation passing through the plasma in the case of burn-through (bleaching) of the target and of radiation scattered (reflected) by the plasma in different directions. In measuring the energy transmitted through a burnt-through foil, we use a special system for screening light detectors from spurious background radiation (that generated in the plasma and scattered laser light), which includes lens *6*, light-screening tube *7*, and a narrow-band ($\delta\lambda\approx 5$ Å for $\lambda=1.06$ μm) optical interference filter. The minimum energy detected by the calorimetric system is 0.1 mJ, the uncertainty of absolute measurements is 7–9%, the uncertainty of relative measurements is 1.5–3% at a resolution of the ADC

circuit of 2.5 μV/bit. Fast measurement of the intensities of incident and transmitted energy with a time resolution $\tau \approx 0.2$ ns are performed using coaxial photocells.

The pattern of the laser spot focused on the target in the near and far fields is imaged by a dedicated automatic system for remotely controlled measurements developed for diagnostic devices of the Pico facility. A schematic of the system for measuring laser radiation parameters is given in Fig. 1. The beam pattern is monitored prior to entering the target chamber close to its input window. Far-field measurements are performed by a DDC-4M camera *22* built around a CCD detector array, and in the near field by camera *23* built around an LI 427 vidicon with a memory time of ~ 5 min. The intensity of measured laser beams is attenuated to the required level by optical filters *13*, including interference filters placed in front of the objective lenses of the cameras.

The relative averaged values for the distributions of power density in the laser beam are related to the measurements of laser power and energy averaged over one pulse using calorimeter *11*. A videosignal from each camera is fed to a dedicated input card for image processing. The two cards are installed in one 486DX2-66 personal computer in 16-bit slots of the system bus and have different addresses for interrupt processing.

Images generated by camera *22* are input via and processed by an Intel Smart Videorecorder Pro card at a rate of 25 frames per second in the Intel Indeo Video 3.2 format with 320×240 pixels and 255 grades of gray. The laser pulse is matched to the frame sync pulse from the video camera using synchronization system *19* built around TTL components. The videosignal from camera *23* is fed to a BW-03 Candella image input card, and the format of the output frame is 256×256 pixels with 64 grades of gray. Since this camera has its own memory unit, it does not require a circuit for synchronization with the laser because it is synchronized with the laser by the software when a signal is detected in the background operation mode of Windows 3.1. Recorded images of the laser beam are stored in the form of BMP files and processed using Mathcad 6.0 PLUS.

3. RESULTS OF EXPERIMENTS

By processing far-field images of the laser beam, we could record a specific far-field pattern of the laser radiation, i.e., in the focusing plane on the target surface. The laser field on the target surface does not have a uniform distribution over the beam cross section, but has areas of higher energy concentration with characteristic sizes of less than 10 μm. The number of such spots (''speckles'') with higher energy densities on the target surface was ten to twelve and varied from measurement to measurement. Measurements of density with a high spatial resolution gave peak power densities in these regions about a factor of 10–20 higher than the average power density in the focused laser spot. A plausible cause of this pattern is the specific properties of the laser field in high-power multistage laser systems due to the high degree of coherence of the output radiation.

With a view to modeling processes in thin-shell thermonuclear laser-driven targets, in these experiments we measured the fraction of incident laser energy scattered by a target and transmitted through the target using the calorimetric system and coaxial photocells. Such measurements have allowed us to study processes of foil explosion excited by the laser, along with propagation of shocks and thermal waves in targets.

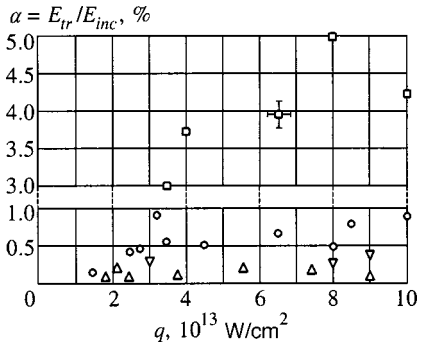


FIG. 2. Fraction α of laser energy transmitted through foils of different thicknesses as a function of power density q on the target. (□) α_3; (○) α_6; (△) α_{40}; (▽) α_{12}.

Figure 2 gives the measurement data for the fraction α of the laser energy transmitted through the target ($E_{\rm tr}$) with respect to the energy $E_{\rm inc}$ incident on the target as a function of the power density q for foils of different thicknesses. Figure 3 shows measurements of the fraction α of transmitted energy versus the foil thickness δ. The numbers next to the experimental points indicate the power density on the target surface expressed as $q\cdot10^{13}$ W/cm^2. It was found that for aluminum foils of thicknesses larger than 12 μm the fraction of transmitted energy is less than 0.4%, whereas for foils of a thickness of 3 μm this fraction is up to 5% of the incident energy. A strong dependence of the transmitted energy on the foil thickness and a relatively small narrowing of the laser pulse (10–15%) have been detected in this region of small fractions of transmitted energy. At larger foil thicknesses, the onset of the transmitted pulse is delayed with respect to the incident pulse because a fraction of the latter is absorbed.

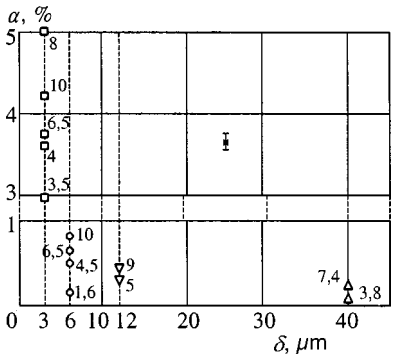


FIG. 3. Fraction of laser energy transmitted through a foil as a function of the foil thickness. Numbers near the points denote the power density on the target surface expressed as $q\cdot10^{-13}$ W/cm^2.

4. THEORETICAL MODEL AND DISCUSSION OF EXPERIMENTAL RESULTS

As follows from previously published results,[7] the evaporation rate dm/dt of the material per unit of the surface area is a function of the power density q and wavelength λ of the laser radiation is

$$dm/dt\,[\mathrm{g/cm^2\cdot s}] \approx 1.43\cdot 10^5 \times \left(10^{-14} q[\mathrm{W/cm^2}]\right)^{1/3} (\lambda\,[\mu\mathrm{m}])^{-4/3};$$

it follows that, for $q \leqslant 10^{14}$ W/cm^2 and $\lambda = 1.06$ μm an aluminum layer of thickness 1.5 μm can evaporate in a time $t = 3$ ns.

Using the one-dimensional Diana code,[8] which incorporates complex equations of state[9] and ionization kinetics,[10] we numerically simulated the heating and acceleration of aluminum foils under the conditions of the experiments performed on the Pico facility and described in the previous section. The calculations were performed in Cartesian coordinates. The real shape of the laser pulse was replaced with an isosceles triangle with a width of 4 ns. We varied the maximum power density q_m on the target surface and the foil thickness δ. More than fifty versions of calculations have been performed. The power density was varied over the range $q_m = 10^{13}$–10^{15} W/cm^2, and the foil thickness was 1–20 μm.

Our one-dimensional calculations show that a foil can be ''bleached'' owing to evaporation of its external layers only if its thickness is 2–4 μm at power densities $q_m = 10^{13}$–10^{14} W/cm^2, which is in accordance with the formula given above. Nonetheless, in experiments we observed bleaching of foils with considerably larger thicknesses. A plausible mechanism of this local bleaching of a target can be gas-dynamic puncturing of solid fractions of a target in the region of ''hot spots'' with subsequent disruption of the target.[11]

The next series of calculations simulating inhomogeneous laser heating and acceleration of foils was performed using the NATSI code[12] based on the Euler equation and numerically solving the two-dimensional gas-dynamic and heat equations in cylindrical coordinates r, z, and t. Incident laser radiation propagates along the z axis and is absorbed by the inverse bremsstrahlung mechanism. The radiation flux penetrating to the critical surface is fully absorbed in this region. Since we considered laser beams with power densities of 10^{13}–10^{14} W/cm^2, the calculations were based on Spitzer's classical model of thermal conductivity. Generation of spontaneous magnetic fields was neglected. The flux of laser power propagating from the right-hand side (in the negative z direction) was determined by the formula

$$q_1(r,t) = q_1(t) q_2(r),$$

where $q_1(t)$ is a pulse shaped like an isosceles triangle with a base of 4 ns and

$$q_2(r) = \exp[-(r/R_f)^2]/C_1,$$

i.e., the power distribution in the transverse direction is Gaussian. Here C_1 is a normalizing constant such that

$$\int_0^{R_0} q_2 r\, dr = 1,$$

$R_0 = 200$ μm is the transverse size of the calculation area and $R_f = 70$ μm is the effective radius of the laser beam. The laser pulse energy was set at 10 J and the aluminum foil thickness (whose initial density is 2.7 g/cm^3 and the charge of emitted ions is 12) was varied.

Figure 4a shows our two-dimensional calculations and compares them to the experimental data. It is clear that burn-through of a foil under a smooth radiation front with a Gaussian distribution is possible only when the foil thickness $\delta < \delta^* \approx 3$ μm.

In order to determine the foil burning rate as a function of the incident laser energy, we have also performed calculations at laser pulse energies $E_L = 20$ J and $E_L = 50$ J at fixed $R_f = 70$ μm. It was found that the density on the beam axis was close to the critical value at the termination of the laser pulse only at a pulse energy as high as 50 J, i.e., the laser beam could penetrate to the back surface of the foil (Fig. 4c). Figure 4b shows distributions of plasma density on the ($r = 0$) axis at the conclusion of the laser pulse ($t = 4$ ns) calculated at pulse energies of 10 and 50 J. One can clearly see that the plasma density drops to approximately the critical value only in the second case. Figure 4c shows also that by this time the density peak has traveled through a distance $L_z = 500$ μm.

Thus, in the case of a small Gaussian intensity distribution in the focusing plane, a ''threshold'' effect should have been detected in investigating the burning rates of foils with different thicknesses, namely, for $\delta < \delta^*$ the foil burns through and for $\delta > \delta^*$ no laser light should have been transmitted to the back surface of the foil (certainly, the parameter δ^* should be different for different laser facilities, and in this specific case δ^* should be approximately 3 μm).

In our experiments only a small fraction of laser energy was transmitted for foil thicknesses substantially larger than 3 μm. In order to explain this phenomenon, we have to take account of nonuniform hydrodynamic effects. In the region where laser radiation is absorbed, a cone-shaped shock front is generated, and the latter propagates both along the beam axis and in the transverse direction. This shock front separates the plasma layers, which allows the radiation to penetrate into deeper layers of the target. A similar mechanism of laser piercing or ''boring'' of aluminum foils with thicknesses of 200–300 μm by a KrF laser was investigated in the case of a single Gaussian beam of a variable radius R_f (Ref. 11).

If a laser spot on a surface contains speckles, such speckles can penetrate through thicker layers than an averaged laser beam with a smooth intensity distribution over its cross section because, at a fixed intensity of absorbed laser radiation, a larger force driving the plasma is generated in a speckle of a smaller transverse dimension.

The dashed vertical line in Fig. 4a is an approximate boundary between two regions of two different regimes of radiation transmission through foils. The burning (evaporation) regime is realized on the left of the dashed line, and the microscopic piercing regime on the right of the dashed line.

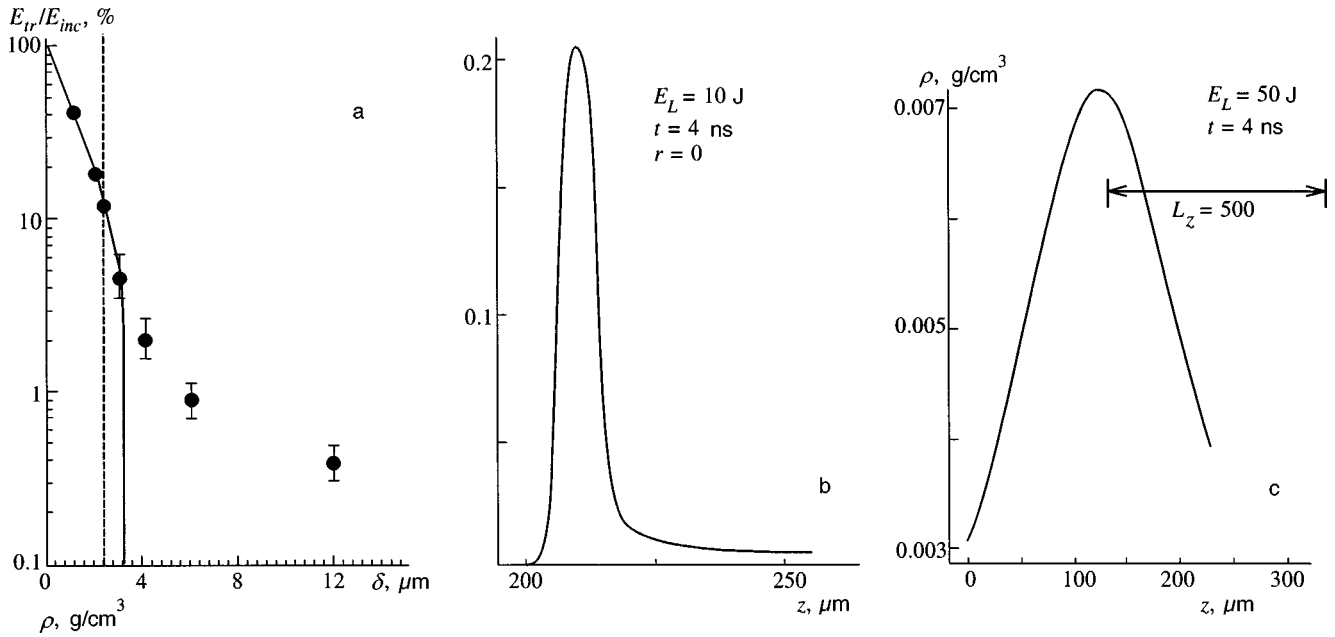


FIG. 4. (a) Fraction of energy transmitted through a foil as a function of foil thickness δ ($E_{\rm tr}$ and $E_{\rm inc}$ are the transmitted and incident energies generated by the laser). Comparison between experimental data (points) and calculations (solid line). The vertical dashed line divides regions of burn-through (on the left-hand side) and piercing (on the right-hand side) regimes. The plasma density distributions along the optical axis are two-dimensional calculations at pulse energies (b) 10 J and (c) 50 J.

In order to numerically simulate the microscopic piercing regime in a foil by speckles in a laser spot, we performed a series of two-dimensional calculations. The foil thickness was set to 4 μm, the laser radiation distribution over the beam cross section was assumed to be Gaussian, but R_f and the pulse energy were varied so that the laser beam intensity was constant and equal to the radiation intensity in the experiments averaged over the laser spot (i.e., the pulse energy in the calculations was determined by the formula $E_L = 10(R_f/70)^2$, where the radius is expressed in micrometers and the pulse energy in joules).

Calculations were performed for $R_f = 5$, 10, and 20 μm. We modeled the impact of an isolated speckle on the burning of a foil with a fixed thickness. Figure 5 shows these calculations. The upper set of curves plasma density shows contours when the foil is burnt through on the beam axis (at $r = 0$). For $R_f = 20$ μm the aluminum layer has not been bleached at the end of the pulse; for $R_f = 10$ μm bleaching is

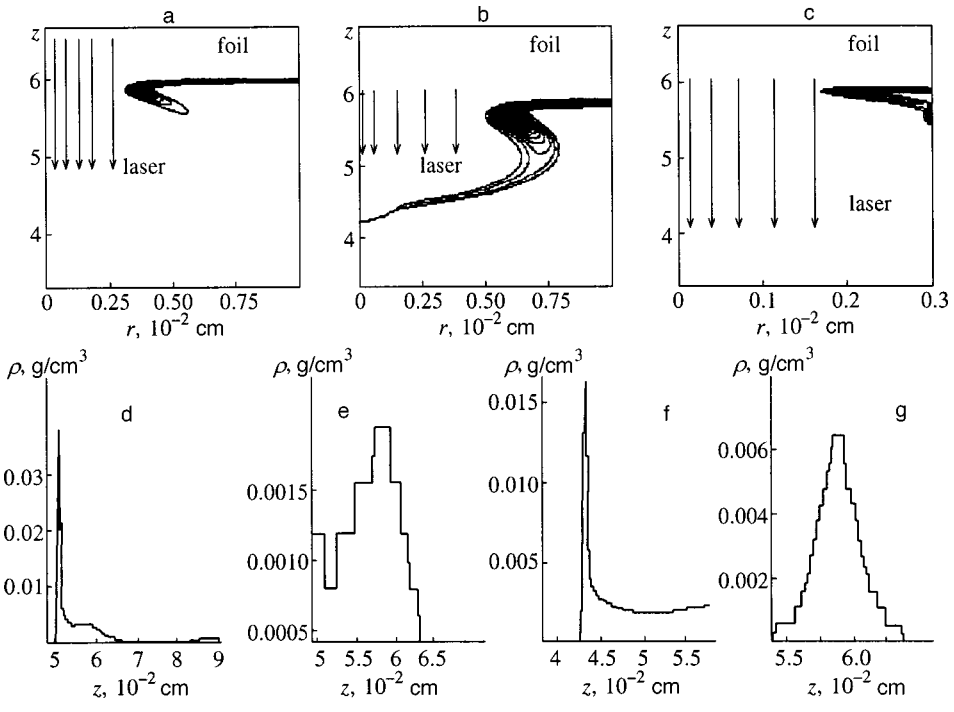


FIG. 5. Two-dimensional calculations for speckles with effective radii $R_f = 10$, 20, and 5 μm at the time of burn-through (when the plasma density on the axis becomes equal to the critical density). The upper part of the graph plots lines of equal density for three times: (a) 3.5 ns, $R_f = 10$ μm; (b) 4 ns, $R_f = 20$ μm (in this case, the target has not been pierced since the density is higher than the critical value); (c) 2.5 ns, $R_f = 5$ μm. Plasma density as a function of distance along the axis ($r = 0$) at times (d) $t = 3.0$ ns, $R_f = 10$ μm; (e) $t = 3.5$ ns, $R_f = 10$ μm; (f) $t = 4$ ns, $R_f = 20$ μm; (g) $t = 2.5$ ns, $R_f = 5$ μm is plotted in the lower part of the graph.

observed at $t=3-3.5$ ns. The graph shows density distributions at the time 3 ns, when the density is higher than the critical value, and at $t=3.5$ ns, when the density has already dropped below the critical value. In the case of $R_f=5$ μm, the layer has been bleached by the time $t=2.5$ ns. Note that by this time the density peak has traveled through a distance approximately equal to ten radii R_f of the focused laser spot.

Two-dimensional calculations are difficult, therefore, a large set of one-dimensional calculations by the Diana code has been performed for different δ and q_m. The calculations are plotted in Fig. 6, where one can see the distance L_z through which the density peak travels as a function of δ and q_m. The ''dips'' on the calculated surface arise because thin foils are fully heated owing to their thermal conductivity and ''explode,'' as a result of which the density peak moves over a smaller distance then in the case of a piece of solid shell with a larger initial thickness. Assuming that a foil is disrupted if the density peak moves through a distance $(10-20)R_f$, where R_f is the speckle radius, we can estimate the parameters of speckles that should bleach a foil with a given thickness.

A distribution of laser radiation can be smoothed on a target surface if, for example, a symmetrizing prepulse is used.[13–15] In order to demonstrate the symmetrizing prepulse effect, we have performed the following calculations with the NATSI code: the laser pulse consisted of two parts with similar shapes $q_1(t)$, i.e., each had a shape of an isosceles triangle with a base width of 4 ns, but shifted with respect to one another by 3.6 ns (i.e., the peak of the first pulse was set at $t=2$ ns and that of the second at $t=5.6$ ns). In the first

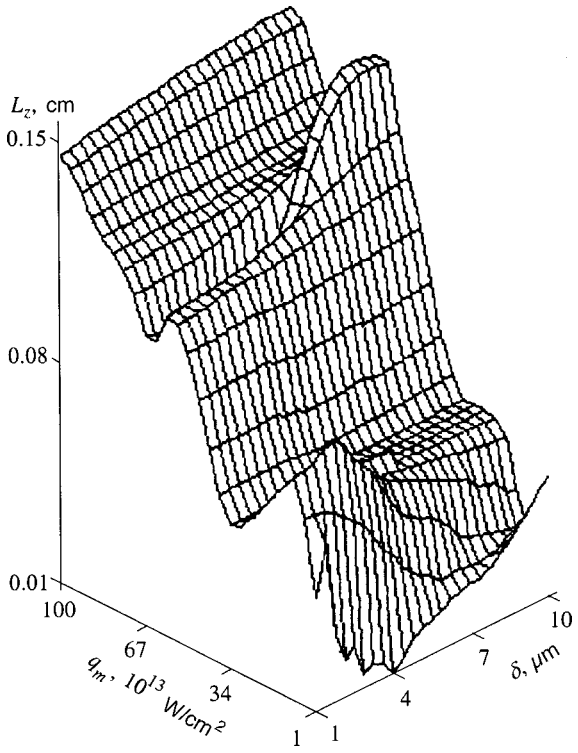


FIG. 6. Distance L_z through which foil has traveled by the end of the laser pulse as a function of the power density q_m and foil thickness δ.

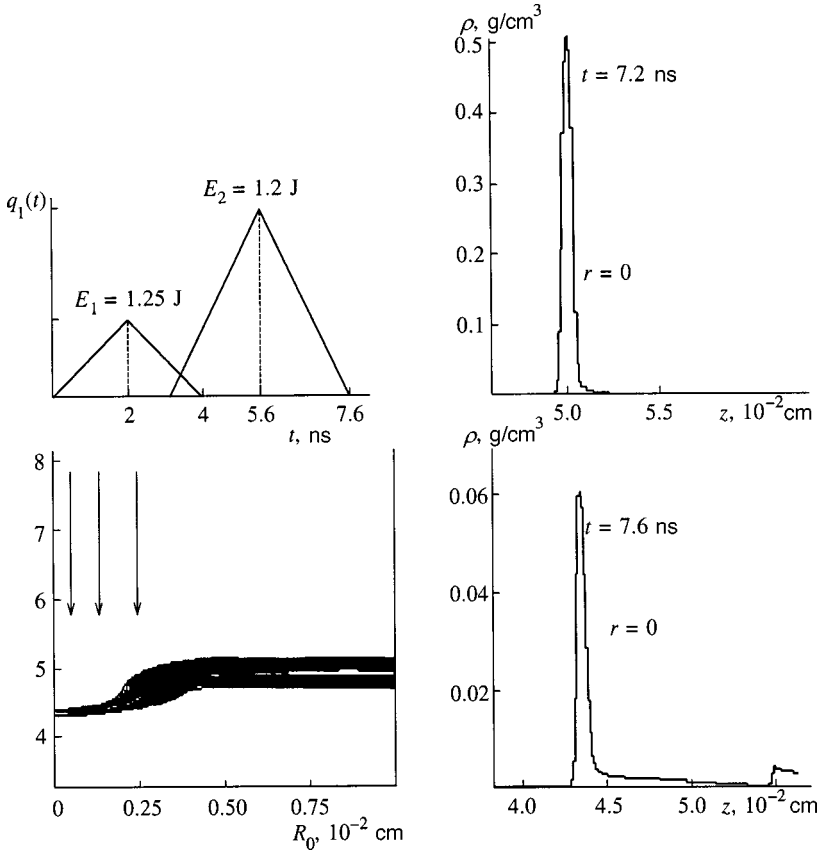


FIG. 7. Calculations for burn-through of a foil in the presence of a laser prepulse. The graph illustrates the case of a speckle with $R_f=10$ μm. The plasma density distributions along the axis are given at times 7.2 and 7.6 ns (the pulse termination). The laser power density is defined by the formula $q(t,r)=q_1(t)q_2(r)$, where $q_1(t)$ is shown in the picture and $q_2=\{^{1/C_2,\quad t<3.6\text{ ns}}_{(1/C_1)\exp[-(r/R_f)^2],\quad t>3.6\text{ ns}}$, where C_1 and C_2 are normalizing constants determined by the condition $\int_0^{R_0} q_2 r dr=1$.

pulse, the value of q_2 was constant, and that in the second was a Gaussian with $R_f = 10$ μm. The energy of the first pulse was $E_1 = 1.25$ J and distributed uniformly over a surface with size $R_0 = 100$ μm, and the energy of the second pulse was $E_2 = 0.2$ J, as in the previous case of a single speckle discussed in this section. Thus, the power density in the first pulse was a factor of about 16 lower than in the focused spot of the speckle ($r < R_f = 10$ μm). Thus, the effect of a defocused prepulse on the foil piercing by a speckle, with an effective radius of 10 μm was simulated. Figure 7 shows lines of equal density and its distribution at times 7.2 and 7.6 ns. The time 7.2 ns corresponds to the time 3.5 ns in the calculation without a prepulse (Fig. 5a). One can clearly see that with a prepulse the foil had not bleached even at the end of the laser pulse, whereas in the absence of a prepulse it had been pierced by the time 3.5 ns.

Such experiments are presently in the schedule of work to be performed on the Pico facility.

In conclusion, note that our calculations have not yet taken into account generation of spontaneous magnetic fields.[16–18] This effect can give rise to larger inhomogeneities in the target heating because magnetization impedes thermal conductivity in the transverse direction.

5. CONCLUSIONS

The analysis of experiments on burn-through of thin foils on the Pico facility based on one- and two-dimensional computer simulations has allowed us to investigate anomalously deep penetration of laser radiation into foils of relatively large thicknesses (i.e., when the mass of evaporated material per unit area is much larger than expected for such foils). The cause of this phenomenon is in the complex pattern of the laser spot, which consists of speckles with different intensities, and the effect of these speckles is microscopic piercing of the metal layer.

Our investigations have demonstrated that, using relatively simple calorimetric measurements of the fraction of energy transmitted through foils of various thicknesses and comparing them to numerical calculations, one can analyze the effect of inhomogeneities in the laser spot on local penetration of laser radiation through a target and even on partial destruction of the latter. This analysis seems to be of considerable importance for research into laser-driven thermonuclear fusion performed by facilities with multiple laser beams. The point is that, in addition to inhomogeneities in isolated laser beams, which can be monitored in principle at the output, the interference pattern generated by several beams on the target surface can be fairly complicated. The heating inhomogeneity is very difficult to monitor in this region. Owing to the effect of microscopic piercing described in this paper, additional undesirable entropy can be fed into the target, which can result in preheating of thermonuclear fuel, prevent high compression of fuel materials, and disrupt the internal structure of the target (at present the target for laser-driven thermonuclear fusion is a multilayered shell whose inside surface is coated with deuterium-tritium ice). With the help of the techniques reported in this paper, one can investigate penetration of laser radiation into a target, which gives rise to inhomogeneities in the power density and heating conditions.

In addition, using two-dimensional numerical calculations we have shown that partial symmetrization of the ablation pressure is possible when a laser prepulse is used.

*)E-mail: mikh@lpi.ac.ru

[1] V. V. Ivanov, A. K. Knyazev, A. V. Kutsenko *et al.*, Kratkie Soobshcheniya po Fizike FIAN No. 7–8, 37 (1997).
[2] A. É. Bugrov, I. N. Burdonskii, V. V. Gol'tsov *et al.*, Zh. Éksp. Teor. Fiz. **111**, 903 (1997) [JETP **84**, 903 (1997)].
[3] V. V. Ivanov, A. K. Knyazev, A. V. Kutsenko *et al.*, Zh. Éksp. Teor. Fiz. **109**, 1257 (1996) [JETP **82**, 677 (1996)].
[4] V. V. Ivanov, A. K. Knyazev, N. E. Korneev *et al.*, Prib. Tekh. Éksp. No. 4, 112 (1995).
[5] V. V. Ivanov, A. K. Knyazev, A. V. Kutsenko *et al.*, Prib. Tekh. Éksp. No. 4, 123 (1996).
[6] V. V. Ivanov, A. K. Knyazev, A. V. Kutsenko *et al.*, Prib. Tekh. Éksp. No. 4, 168 (1997).
[7] F. Dahmani and T. Kerdja, Phys. Rev. A **44**, 267 (1991).
[8] N. V. Zmitrenko, V. Ya. Karpov, A. P. Fadeev *et al.*, in *Topics in Atomic Science and Technology. Methods and Programs for Numerical Solution of Problems of Mathematical Physics* , No. 2, 38 (1982).
[9] V. Ya. Karpov, A. P. Fadeev, and G. V. Shpatkovskaya, IPM Preprint, USSR Academy of Sciences, No. 147 (1982).
[10] Z. Amad, Yu. A. Zakharenkov, I. G. Lebo *et al.*, Zh. Éksp. Teor. Fiz. **100**, 1140 (1991) [Sov. Phys. JETP **73**, 629 (1991)].
[11] V. G. Bakaev, V. Yu. Korol', I. G. Lebo *et al.*, Preprint FIAN No. 6 (1997).
[12] I. G. Lebo, V. B. Rozanov, V. F. Tishkin *et al.*, in *Proc. of 6-th Intern. Workshop on fhe Physics of Compressible Turbulent Mixing*, Marseille, France (1997), ed. by Institut Universitaire des Systèmes Thermiques Industriels (1997), p. 312.
[13] N. I. Bokov, A. A. Bunatyan, A. A. Lykov *et al.*, Problemy Matematiki i Teoreticheskoi Fiziki, No. 4, 20 (1982).
[14] E. G. Gamaly, A. O. Fedyanin, I. G. Lebo *et al.*, Laser Part. Beams **8**, 398 (1990).
[15] A. B. Iskakov, I. G. Lebo, I. Limpoukh *et al.*, Preprint FIAN No. 22 (1998); submitted to Phys. Rev. A.
[16] R. S. Craxton and M. G. Hams, Phys. Rev. Lett. **35**, 133 (1975).
[17] Yu. V. Afanas'ev, E. G. Gamalii, I. G. Lebo *et al.*, Zh. Éksp. Teor. Fiz. **74**, 516 (1978) [Sov. Phys. JETP **47**, 271 (1978)].
[18] L. A. Bol'shov, A. M. Dykhne, V. P. Liselev *et al.*, Zh. Éksp. Teor. Fiz. **84**, 921 (1983) [Sov. Phys. JETP **57**, 535 (1983)].

Translation provided by the Russian Editorial office.

Two-dimensional mesoscopic dusty plasma clusters: structure and phase transitions

G. E. Astrakharchik, A. I. Belousov, and Yu. E. Lozovik*)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 6 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1300–1312 (October 1999)

A two-dimensional mesoscopic cluster of ''dusty plasma'' particles, which can be interpreted as a system of microparticles in an rf gas discharge, is investigated. The ground-state configurations and corresponding eigenfrequencies and eigenvectors are found for clusters of $N=2-40$ particles in a harmonic confining potential. It is shown that a change in the Debye screening length R of the particle charge in the plasma can cause structural transformations of the ground state of the system, manifested as first-order or second-order phase transitions with respect to the parameter R. The disorder (''melting'') of the clusters is analyzed in detail by Monte Carlo simulation and molecular dynamics. By varying the characteristic range of particle interaction in a cluster, it is possible to modulate its thermodynamic properties and the character of the phase transitions, thereby causing a controlled transition of the system into the fully ordered, orientationally disordered, or fully disordered state. The possibility of dusty plasma clusters coexisting in different states is discussed. [S1063-7761(99)01210-X]

1. INTRODUCTION

Small charged ''dust'' particles in a neutralized plasma constitute a highly prevalent system, which can be observed in a variety of scales and media, e.g., dust clusters in interstellar space and in the upper layers of the atmosphere; ordered structures in a gas discharge of the kind used in the industrial treatment of various materials, etc.; the list of such systems goes on indefinitely. Lately a great deal of attention has been devoted to the experimental study of ''dusty plasma,'' a system of carbon, silicon, or polymer microparticles in an rf gas discharge,[1–4] a laminar, weakly ionized thermal plasma jet,[3,5] and even in a microgravity environment without the use of electric traps for particle confinement.[4] One of the reasons for the interest in such artificially produced objects is the possibility of making direct observations, by laser interferometry for example, of the types and dynamics of formation of ordered structures of ''dust'' particles. Worldwide *in situ* investigations of dusty plasma crystals and liquids in a number of laboratories[1–3] not only have important bearing on our understanding of plasma physics, but are also a powerful tool for studying the processes of melting, annealing, and formation of various kinds of defects in the crystalline phase. In addition, the last few years have witnessed great scientific and practical interest in the study of microclusters, which are intriguing by virtue of their strong structural sensitivity to the number of particles, extraordinary structural transformations as the temperature increases, etc.[6–11] The fascination with systems containing a small number of particles is further reinforced by a wealth of experimental and theoretical information indicating that the clusters can preserve their individuality inside a massive body and thereby influence its properties.

The objective of the present study is to investigate the static and thermodynamic properties of small ''dust'' clusters in a plasma. In ongoing experiments small particles immersed in a plasma acquire a substantial, usually negative charge $-Ze$, $Z\sim 10^3$ because of the higher mobility and temperature of plasma electrons. Debye screening of the particles' charge modifies their Coulomb interaction, and the interaction between particles can be described quite accurately (see Refs. 12 and 13 for a discussion of this problem) by the Yukawa pair potential.[1)] Using the parameters of experiments[1,2] with polymer particles having characteristic dimensions 3×10^{-4} cm in a low-pressure (~ 1 torr) argon rf discharge (~ 10 MHz) plasma, estimating the electron and ion temperatures of the plasma as $T_e \approx 1700$ K and $T_i \approx 300$ K, respectively, and assuming approximately equal densities of plasma electrons and ions $n_e \approx n_i \sim 10^9$ cm^{-3}, we can obtain the following expression for estimating the Debye screening length:

$$R=\left(\frac{4\pi e^2 n_i}{k_b T_i}+\frac{4\pi e^2 n_e}{k_b T_e}\right)^{-1/2},$$

i.e., we have

$$R\approx \sqrt{k_b T_i/4\pi e^2 n_i}\approx 10^{-2}\ \text{cm},$$

which is a quantity of the same order as the average distance between the dust particles. For a few-particle cluster all the particles are situated in a single layer (the electrode sheath), and we have a two-dimensional (2D) cluster of N dust particles in the plasma; an expression for the energy of the cluster can be written in the form

$$E=(Ze)^2\sum_{i<j}\frac{\exp(-|\mathbf{r}_{ij}|/R)}{|\mathbf{r}_{ij}|}+\alpha\sum_{i=1}^{N}|\mathbf{r}_i|^2. \qquad (1)$$

We assume here that the particles are confined by the quadratic external force potential α.

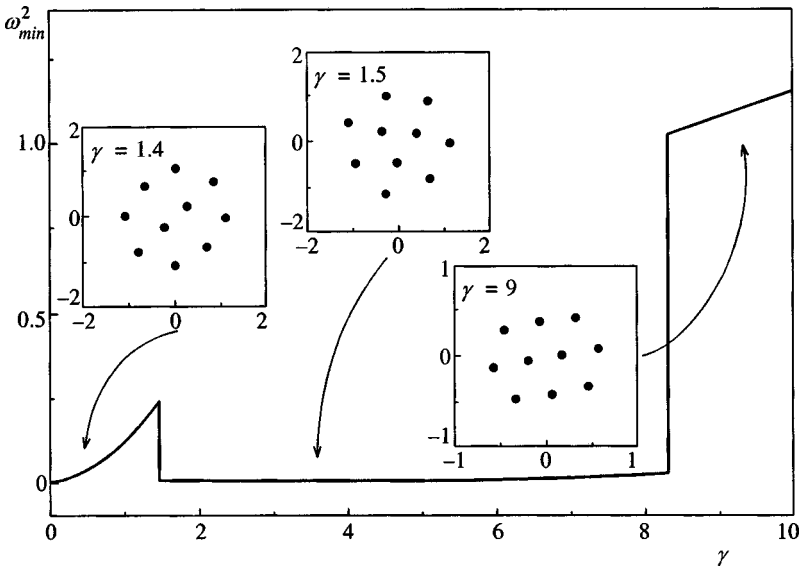


FIG. 1. Lowest nonzero eigenfrequency $\omega_{\min}$ of a dusty plasma cluster of $N=10$ particles. Insets: ground configurations of the system in three different ranges of the control parameter γ.

We note that in recent experiments aimed at the observation of small dust clusters[2] the investigated system has been essentially quasi-2D, consisting of vertical chains of dust particles. Chains of this kind can form[14–16] not only by virtue of the Yukawa potential, but also as a result of dipole–dipole interaction between particles, presumably caused by the negative charge of a dust particle focusing the stream of positive ions incident on the electrode from the plasma discharge cloud.[2,12,13] These short (20 particles or less) chains are capable of moving only in the horizontal plane, forming a cluster with a characteristic shell structure, which manifests orientational and total melting effects. We also note that the direct experimental determination of the interaction potential between two particles has yielded only Yukawa interaction, i.e., dipole–dipole interaction at the investigated distances has provided a smaller contribution.[17] Clearly, in the 2D case discussed here, the addition of dipole–dipole interaction does not produce any new qualitative phenomena.

Introducing dimensionless units for the distance $r_0=(Ze)^{2/3}/\alpha^{1/3}$ and the energy $E_0=\alpha r_0^2$, we can express the cluster energy in the form

$$E=\sum_{i<j}\frac{\exp(-\gamma|\mathbf{r}_{ij}|)}{|\mathbf{r}_{ij}|}+\sum_{i=1}^{N}|\mathbf{r}_i|^2. \tag{2}$$

Consequently, the thermodynamic state of a cluster of a given number of particles is governed by two dimensionless parameters: the reciprocal screening length $\gamma=r_0/R$ and the dimensionless temperature of the system $\Theta=k_bT/E_0$. The characteristic particle interaction range $1/\gamma$ can be controlled by varying the Debye screening length R, which is a function of the density and temperature of the plasma. The translational particle temperature[2)] can be measured on the basis of the virial theorem from observations of the particle vibrations (or from the ''Doppler'' contour of scattered radiation).

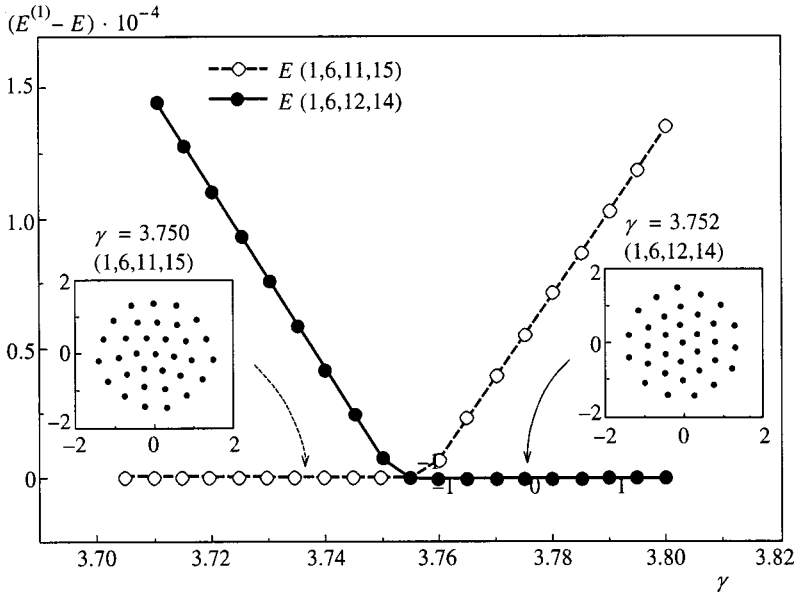


FIG. 2. System of $N=33$ particles: energies and configurations of the lowest local minimum $E^{(1)}$, measured from the global minimum energy E, for a cluster in the vicinity of phase transition.

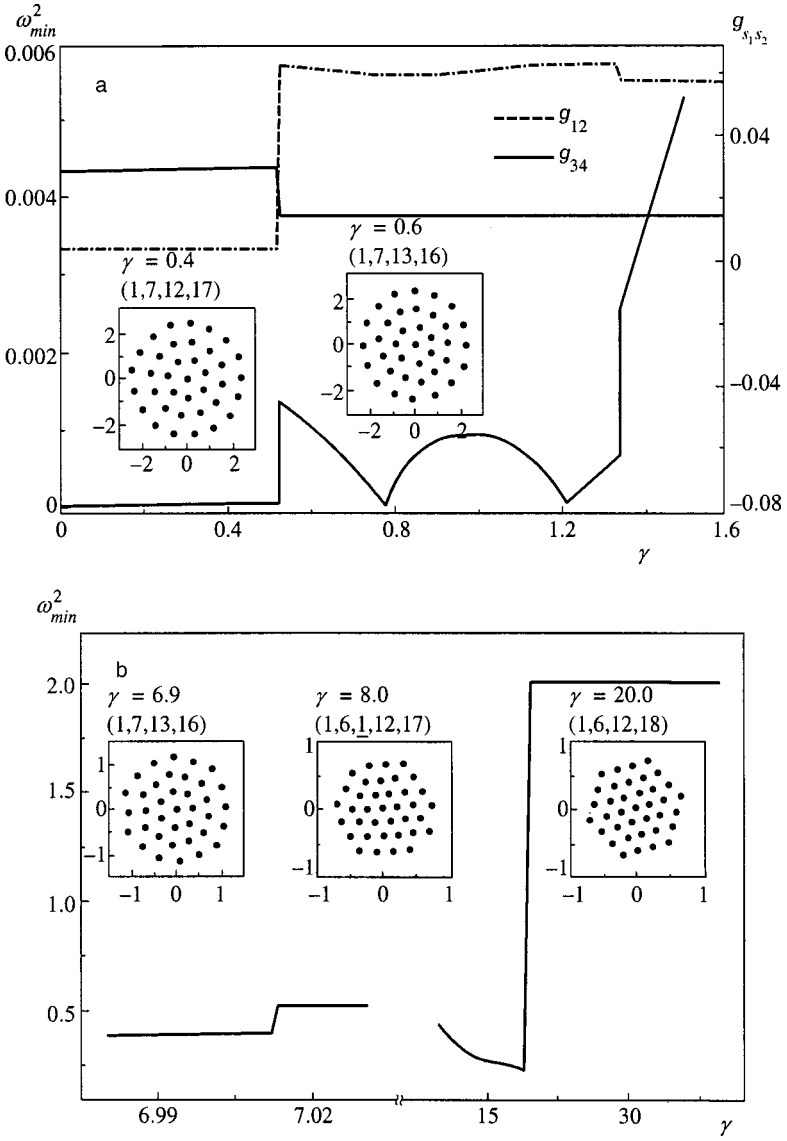


FIG. 3. System of $N=37$ particles: a) lowest nonzero eigenfrequency $\omega_{\min}$ and mutual orientational order parameter of different shells of the cluster as a function of γ; b) region of strong screening of the particle charge in the plasma. The eigenfrequencies and lowest nonzero frequency for $\gamma>19$ correspond to motion of the cluster as a whole in a harmonic potential.

In this paper we investigate the properties of 2D dusty plasma clusters (2) as functions of the control parameter γ and the temperature Θ. For a series of clusters containing $N<40$ particles we determine the structures of the dust clusters, the spectra of harmonic vibrations, and the ground-state energies (Sec. 2). The variation of the Debye screening length (the parameter γ) induces structural transformations of the ground state at certain values of the parameter γ^*; this effect can be regarded as the occurrence of various types of phase transitions with respect to the parameter γ. In the examples of clusters containing $N=10$, 33, and 37 particles we show that the investigated systems can undergo a series of first-order and second-order phase transitions over a wide range of $\gamma\in[0,30]$. We use molecular-dynamics and Monte Carlo techniques in a canonical ensemble to analyze the thermodynamic properties of small dusty plasma systems and various types of disorder (melting) phenomena (Sec. 3). We show that the thermodynamic state of small clusters of dust particles can be controlled not only by the temperature of the cluster, but also by the Debye screening length, i.e., by the density and temperature of the plasma in which the particles are immersed. Slight variations in the parameters of the experiment can give rise to major changes in the cluster structure and the ''phase transition'' temperatures and can even cause certain types of disorder effects to disappear.

2. CONFIGURATIONS FOR GLOBAL MINIMA

We use a modified Newton method[7] and a combined ''random search+gradient descent'' method[8] to find configurations that produce a global minimum in the system (2). The global minimum configurations shown below (see Figs. 1–3 and Table I) have been found by the two methods independently, improving the reliability of the results. Of course, none of the existing methods used to seek the minimum of a function of many variables is in a position to guarantee that the resulting configuration will correspond to a global minimum. To cope with this difficulty, we have analyzed up to 200 arbitrarily distributed configurations. The same approach can also be used to investigate local minima and the regions of convergence to them (relative weights of the local minima).

TABLE I. Distribution of particles among the shells of $\{N_1, N_2, \ldots\}$ ground states of clusters of N particles in a harmonic confining potential with dipole, Coulomb, and logarithmic interactions.

N	$1/r^3$	$1/r$	$-\ln r$
9	2,7	2,7	1,8
10	3,7	2,8	2,8
11	3,8	3,8	2,9
-	-	-	-
32	1,6,12,13	1,5,11,15	4,11,17
33	1,6,12,14	1,6,11,15	5,11,17
34	1,6,12,15	1,6,12,15	1,5,11,17
-	-	-	-
36	1,6,12,17	1,6,12,17	1,6,12,17
37	1,6,1,13,16	1,7,12,17	1,6,12,18
38	2,8,13,15	1,7,13,17	1,6,12,19

In the limit of weak screening of particle charges in the plasma, $\gamma \ll 1$, the model (2) describes a Coulomb cluster in a harmonic confining potential, a system that has been heavily investigated both experimentally[18,19] and by numerical simulation methods.[6,7,20] In particular, previous Coulomb cluster calculations have shown that finite systems of not too many particles are conveniently classified according to their shell structure (see Table I). Depending on how the concentric shells are filled, a system is assigned to one of the periods in a table of the same sort as Mendeleev's periodic table.

The presence of the parameter γ characterizing the range of the particle interaction potential in a cluster can be exploited to investigate the short-range influence of the particle interaction potential on the structure and properties of the ground states of clusters. The fact that the cluster structure depends on the parameters of the interaction potential becomes obvious from an inspection of the table, which gives certain ground-state configurations of 2D clusters in a harmonic trap (in the order of increasing long-range influence of the pair interaction potential): dipole clusters, Coulomb clusters, and logarithmic clusters. In the investigated case of ''dust clusters'' the ground state of the system undergoes a structural transformation as the parameter γ (or Debye screening length $1/\gamma$) is varied, and each point γ^* at which some type of structural change occurs can be regarded as a point of a particular order of phase transition.

The transition order can be determined from a plot of the energy of the system as a function of the control parameter γ: A discontinuity of the nth derivative of the ground-state energy $E(\gamma)$ with respect to γ at a point γ^* corresponds formally to an nth-order phase transition at the point γ^*. The order of phase transition at a point where the system configuration changes can also be deduced by investigating a set of $2N$ eigenfrequencies of normal modes of the cluster $\omega_i(\gamma)$, $i=\overline{1,2N}$, in the vicinity of the global minimum of $E(\gamma)$; specifically, a discontinuity of one of those frequencies implies a first-order phase transition, whereas zero frequency of a particular motion (mode ''softening'' of the corresponding motion) can be interpreted as a second-order transition at the given point.[9]

Figure 1 shows the results of calculations of the normal frequencies of the dynamical matrix of a cluster of $N=10$ particles for various values of the screening parameter $\gamma \in [0,10]$. The normal-mode spectrum and the corresponding eigenvectors are found by means of the Householder algorithm. The lowest nonzero frequency clearly exhibits discontinuous behavior at the points $\gamma \approx 1.4$ and $\gamma \approx 8.2$. An examination of the cluster ground-state configurations shows (see Fig. 1) that as the short-range influence of the interaction potential increases with increasing γ, the distribution of the particles among the shells initially (at $\gamma \approx 1.4$) undergoes the change typically associated with the transition from Coulomb interaction to dipole interaction ($\{2,8\} \rightarrow \{3,7\}$). The subsequent decrease in the screening length takes the cluster (at $\gamma \approx 8.2$) into a more closely packed state (with configuration $\{2,8\}$), which is typical of a system of rigid spheres.

Figure 2 shows the results of calculations of the ground configurations of a cluster of $N=33$ particles. At the point $\gamma^* \approx 3.751$ the occupation numbers of the two outer shells change as the short-range influence of the potential increases: $\{1,6,11,15\} \rightarrow \{1,6,12,14\}$. The first-order phase transition point γ^* can be determined as the point at which the energies of the ground and lowest excited (metastable) states become equal. This approach is illustrated in Fig. 2, which shows plots of the ground-state energy $E(\gamma)$ and the energy of the lowest local minimum $E^{(1)}(\gamma)$ in the vicinity of a phase transition of a cluster of $N=33$ particles. It is evident from the figure that the configurations producing the system with a global minimum in the system before transition (say, $\{1,6,11,15\}$ for $\gamma < \gamma^*$) form a metastable branch of the system in the immediate vicinity of the point γ^* after transition.

A very interesting pattern of structural transformations with variation of the screening length was expected on the part of a cluster of $N=37$ particles. In the corresponding dipole system one particle is situated between the second and third shells, forming an ''interstitial defect'' (analogous to a Frenkel defect in a crystal), and the partition of the ground configuration into shells is not unique.[8] An investigation of this cluster for various values of the parameter γ shows that four phase transitions take place in the interval $\gamma \in [0,1.6]$ [see Fig. 3a]: two second-order transitions (at $\gamma \approx 0.78$ and $\gamma \approx 1.22$, where the lowest normal-mode eigenfrequency is zero) and two first-order transitions (at $\gamma \approx 0.52$ and $\gamma \approx 1.34$). It is evident from Fig. 3a that the occupation numbers of the outer shells change at $\gamma \approx 0.52$). A similar structural transformation at the point $\gamma = 0.52$ is encountered for a cluster of $N=37$ particles. Such abrupt changes in the structure of the clusters are typical of first-order phase transitions. However, an examination of the region of the second first-order phase transition at $\gamma \approx 1.34$ reveals that the structure of the cluster does not undergo any distinct changes. A more detailed analysis shows that in this region the third shell rotates relative to the fourth shell, as is evident in Fig. 3b. To describe these orientational transformations quantitatively, Fig. 3 shows plots of the mutual orientational order parameter $g_{s_1 s_2}$ of different pairs of shells $\{s_1, s_2\}$ [see Eq. (4) below], which is highly sensitive to any change in the mutual orientation of the shells.

A further increase in the parameter γ (decrease in the range of the particle interaction potential) produces two more first-order transitions, which are shown in Fig. 3b. As a result

of the first such transition (at $\gamma \approx 7.015$) one of the particles is interposed between the second and third shells (see Table I and the discussion above). The corresponding structural transformation can be written as

$$\{1,7,13,16\} \rightarrow \{1,6,\overline{1},12,17\}.$$

For $\gamma \approx 19$ the cluster acquires a very distinct faceting and goes over to the most symmetric state {1,6,12,18} with the closest packing. The subsequent increase in the short-range influence of the interaction potential does not produce any more structural changes in the system. Curiously, in the given interval of γ the normal modes having the lowest nonzero frequency $\omega_{\min}$ [see Fig. 3b] correspond to a twofold-degenerate vibration of the entire cluster as a whole in a harmonic trap with frequency $\omega_{\min} = \sqrt{2}$.

The results of investigations of Coulomb and dipole clusters have shown that by far the majority of the ground configurations of these systems are parts of a crystal lattice with hexagonal symmetry.[8,19] In describing and analyzing the properties of such configurations, it is helpful to introduce ''crystal shells'' (Cr_c) in the form of concentric groups of lattice sites of an ideal 2D crystal with c particles located at their center.[3)] An investigation of finite dusty plasma systems has shown (see Figs. 1–3) that the changes in the cluster configurations as the parameter γ is increased (i.e., as the long-range influence of the interaction potential decreases) takes place in the direction along which the maximum number of crystal shells is filled.

We close this section with the observation that a characteristic feature of all the first-order phase transitions we observed in clusters with $N < 40$ and over a wide range of the control parameter γ is an abrupt change in the structure of the cluster. This distinctive attribute shows up as a change in the occupation numbers of the neighboring shells of the system, as occurs in the cases of clusters with $N = 10$ and $N = 33$ particles considered above (see Figs. 1 and 2). However, structural changes can also take place in such a way that the occupation numbers of shells far apart from one another change simultaneously at a first-order phase transition point. This is precisely the situation in the intervals $\gamma \approx 7.015$ and $\gamma \approx 19$ for a system of $N = 37$ particles; see Fig. 3b. Another interesting example of such structural transformations is found in a cluster of $N = 26$ particles, whose behavior is similar to the pattern of changes in a cluster of $N = 10$ particles: As the short-range influence of the particle interaction potential decreases, two first-order transitions occur in succession with simultaneous changes in the number of particles in the first and third shells: $\{3,9,14\} \rightarrow \{4,9,13\}$ for $\gamma \approx 1.6$ and $\{4,9,13\} \rightarrow \{3,9,14\}$ for $\gamma \approx 12$.

3. PHASE TRANSITIONS

An important distinguishing feature of small clusters is the possibility of such systems exhibiting two types of disorder effects:[6,10] intershell disorder (orientational melting of shells s_1 and s_2 at a temperature $\Theta_{s_1 s_2}$) and radial disorder (total melting at a temperature Θ_f). An analysis of the eigenfrequencies shows that for clusters with small values of the lowest eigenfrequencies $\omega_{\min}$ the corresponding eigenvectors are mainly directed along the tangents to the shells and correspond to mutual rotation of the latter.[7] Such clusters will have low intershell disorder temperatures $\Theta_{s_1 s_2}$, when shells s_1 and s_2 rotate relative to one another, losing mutual orientational order.[4)]

It is obvious that the change in the structure of the clusters as the control parameter γ is varied should alter the orientational and total disorder temperatures. It is also possible for the orientational melting effect to vanish altogether when the cluster has a ''well-packed'' structure. The results of our calculations, given below, confirm this hypothesis.

Orientational disorder in clusters is usually investigated by measuring the mean-square relative angular displacements of two shells, which is analogous to mean-square displacements [see Eq. (5) below]. In this approach the orientational melting temperature $\Theta_{s_1 s_2}$ of shells s_1 and s_2 is defined as the temperature at which the mean-square angular displacement changes abruptly. We use a slightly different procedure to determine the orientational melting temperature, defining it as the point at which the mutual orientational order parameter of shells s_1 and s_2 becomes equal to zero.[8] We determine this quantity as follows. For each shell numbered s and containing N_s particles we consider a complex variable ψ_s such that

$$\psi_s = \frac{1}{N_s} \sum_{i_s} \exp(j N_s \varphi_i), \quad j^2 = -1. \tag{3}$$

The sum in Eq. (3) spans all particles belonging to the given shell. We then define the mutual orientational order parameter as

$$g_{s_1 s_2} = \psi_{s_1} \psi_{s_2}^{*}. \tag{4}$$

Clearly, the quantity $\langle g_{s_1 s_2} \rangle$ vanishes at the point of mutual disorder (rotation) of shells s_1 and s_2. The degree of angular order of particles within shell s can be characterized by the quantity $\langle g_{ss} \rangle = \langle |\psi_s|^2 \rangle$. We note that ψ_s and $\langle g_{s_1 s_2} \rangle$ are analogs of the orientational parameter ψ_6 and the correlation function $g_6(r)$ in unbounded 2D systems, where the vanishing of the correlation function, $g_6(r) \rightarrow 0$ in the limit $r \rightarrow \infty$, in the absence of translational order is indicative of the mutual orientational disorder of distant parts of the system.

Figure 4a shows the temperature dependence of the thermodynamic average $\langle g_{21} \rangle$ of the mutual orientational order of a two-shell cluster of $N = 10$ particles for various values of the parameter γ. It is evident from the figure that the change of the system configuration $\{2,8\} \rightarrow \{3,7\}$ (which takes place for $\gamma \approx 1.4$; see Fig. 1) causes the orientational disorder temperature to drop sharply: from $\Theta_{21} \approx 1.3 \times 10^{-4}$ to $\Theta_{21} \approx 0.7 \times 10^{-5}$.

In the range $\gamma > 8.2$, where the cluster is fully packed (see Fig. 1), orientational melting does not occur, and an increase in the temperature results in the exchange of particles between shells, which takes place at $\Theta_f \approx 10^{-3}$. This behavior is evident from the temperature dependence of the mean-square radial displacements u_r^2 of particles in the system:

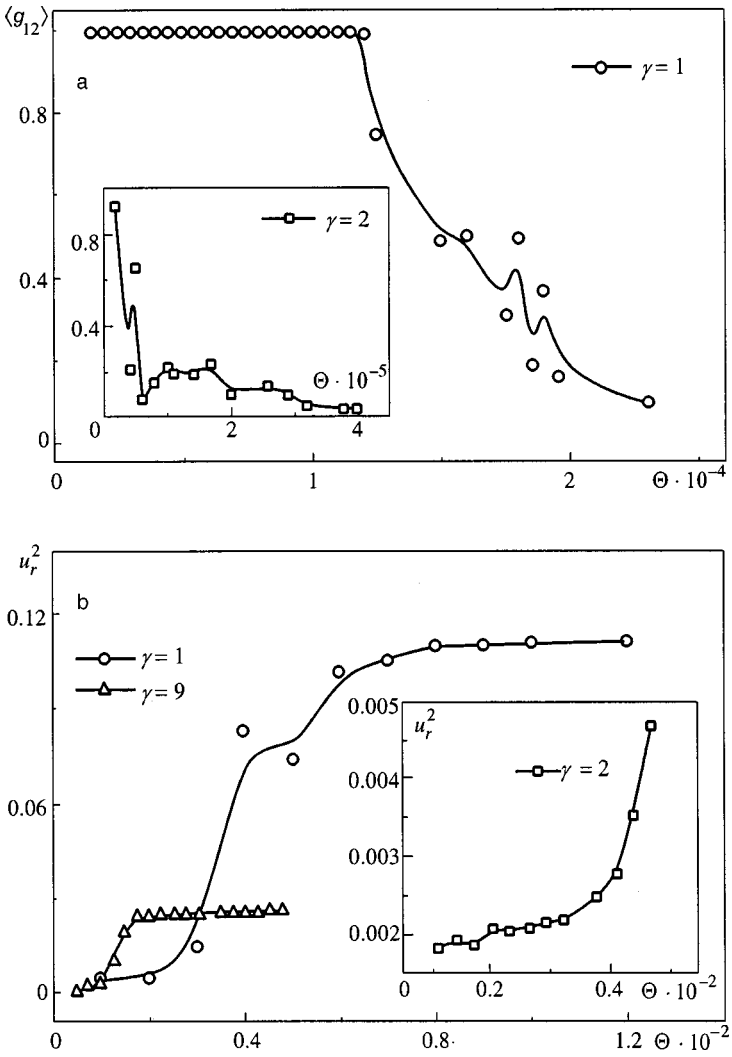


FIG. 4. Two-shell cluster of $N=10$ particles: a) thermodynamic average of the mutual orientational order parameter of the first and second shells for various values of the parameter γ; b) mean-square displacements as a function of the temperature, $u_r^2(\Theta)$.

$$u_r^2=\frac{1}{N}\sum_i\left[\langle|\mathbf{r}_i|^2\rangle-\langle|\mathbf{r}_i|\rangle^2\right]. \tag{5}$$

The corresponding dependence is shown in Fig. 4b. Also shown are analogous curves for systems with $\gamma=1$ and $\gamma=2$. It is evident from the figure that even a slight variation of the control parameter γ can produce an order-of-magnitude change in the total melting temperature in the system.

A variation of the parameters of the interaction potential changes the structure of the isoenergetic surface of the system, which governs the type and characteristic features of the occurring phase transitions. At certain values of the control parameter γ, therefore, it is possible for the system to acquire very interesting thermodynamic properties associated with the distinctive features of such a structure.

Figure 5 shows the temperature dependence of the mean-square radial displacements (5) of a four-shell cluster of $N=33$ particles for $\gamma=3.76$. The plot has a series of plateaus in various temperature intervals. Careful inspection shows that the intervals of abrupt increase in u_r^2 correspond to the successive disorder of different pairs of shells of the cluster: The exchange of particles between the third and fourth shells sets in at a temperature $\Theta^{3,4}\approx 10^{-4}$, and exchange between the second and third shells begins at $\Theta^{2,3}\approx 0.005$. Total melting of the cluster takes place at $\Theta_f\approx 0.01$.

Interesting information about the nature of the processes involved as the temperature increases can be obtained by investigating the distribution of the system among the local minima $\rho(E^{(\mathrm{loc})})$. To estimate the corresponding histogram, we have run several hundred iterations of the gradient descent method in each step of the measurements as a means of determining the nearest local minimum in whose vicinity the system resides, along with the energy of this minimum $E^{(\mathrm{loc})}$. The inset to Fig. 5 shows distributions of a system of $N=33$ particles for $\gamma=3.76$ among the local minima at temperatures $\Theta=10^{-4}$ (ordered state) and $\Theta=8\times 10^{-3}$ (particle exchange between the fourth and third shells). In the fully ordered state the system is always situated in the vicinity of the global minimum of the energy $E=64.795946$, which corresponds to a {1,6,12,14} structure. At $\Theta=8\times 10^{-3}$ the lowest local minimum $E^{(1)}=64.795975$ is occupied with finite probability; in the vicinity of this minimum the configuration of the system can be written as {1,6,11,15} (see Fig. 2).

Looking over these results in conjunction with the above results of searching for the global minimum (see Fig. 2), we

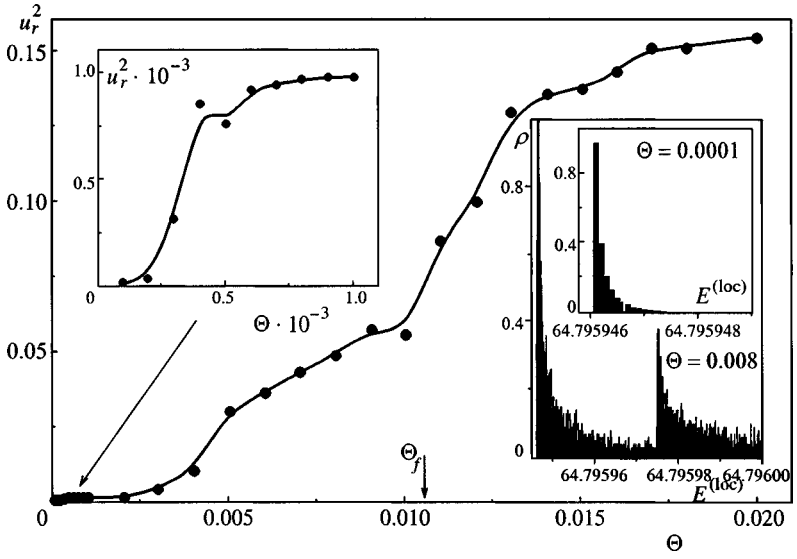


FIG. 5. Four-shell cluster of $N=33$ particles: mean-square radial displacements as a function of the temperature, $u_r^2(\Theta)$. Inset: histograms of the distribution of particles among the local minima $\rho(E^{(\mathrm{loc})})$ in the ordered state (at $\Theta=10^{-4}$) and at the temperature $\Theta=8\times10^{-3}$ at which particle exchange between the third and fourth shells takes place.

can conclude that the first disorder, which is plainly evident on the graph of radial displacements $u_r^2(\Theta)$ in the temperature interval $\Theta\in[10^{-4},10^{-3}]$, corresponds to a finite probability of occupation of the {1,6,11,15} state, which is metastable for the given value of the parameter γ. Such a redistribution of the particles among the shells requires that a potential barrier be overcome, and this fact in combination with the large relative contributions of ''ground'' and ''excited'' states, permits the indicated temperature range to be regarded as the range of dynamical realization of two cluster forms: $\{1,6,12,14\}\leftrightarrow\{1,6,11,15\}$.[21]

4. CONCLUSIONS

In this paper we have investigated a finite system of ''dusty plasma'' particles, which is manifested physically as microparticles in the column of a dc glow discharge or in a weakly isolated, low-pressure rf discharge plasma. For various Debye screening lengths R we have determined the ground-state configurations of clusters consisting of $N\leqslant40$ particles, along with the eigenfrequencies and corresponding eigenvectors of their normal modes. As the screening length varies, the clusters undergo structural transformations, which appear as first-order or second-order phase transitions with respect to the parameter R. The coordinates of the cluster particles change abruptly at the first-order transition points, either by changing the distribution of particles among the shells or as the rotation of pairs of shells relative to one another. At a second-order transition point one of the normal modes softens (vanishes), and the positions of the particles vary continuously.

An investigation of the phase transitions in the system has shown that by varying R (e.g., by varying the density or temperature of the plasma) it is possible to modulate the thermodynamic properties of the system to the extent that the orientational and total disorder temperatures change by orders of magnitude. We have also found that the orientational rotation of different shells of the system can cease for certain Debye screening lengths, when an increase in temperature immediately leads to particle exchange between shells.

An analysis of the variation of the distribution among the local minima as the temperature is varied shows that for certain values of R, in the vicinity of phase transition points, melting in a number of clusters takes place by a multistage mechanism, the disorder of different regions of the cluster occurring in different temperature intervals. The phenomenon of dynamical realization of different cluster forms can be observed in this case.

Author Yu. E. Lozovik is indebted to A. P. Nefedov for valuable discussions. This work has been financially supported in part by a grant from the Russian Fund for Fundamental Research and the program ''Surface Atomic Structures.''

*)E-mail: lozovik@isan.troitsk.ru

1) In the latest experiments the transverse dimensions of the particle ''cloud'' in the plasma are much greater than the Debye shielding length R; consequently, even though the Yukawa potential is associated with the three-dimensional shielding of particle charges by the plasma, it is fully justified in application to the ''two-dimensional'' clusters discussed in the present article.

2) The translational particle temperature, which is a measure of the kinetic motion of massive dust particles, can obviously differ significantly from the temperature of the particles themselves, which depends on the plasma ion and electron temperatures.

3) Obviously, considering the isotropy of the confining potential, we are primarily interested in a finite number of the most symmetric crystal shells, which can be divided into the following groups by the number of sites at the center: Cr_1, Cr_2, Cr_3, Cr_4.

4) In contrast with systems of large numbers of particles $N>40$, we note that orientational melting in small clusters can take place for all pairs of shells, i.e., melting temperatures Θ_{21}, Θ_{32}, Θ_{43},... can exist. Previous calculations of Coulomb, dipole, and logarithmic clusters have shown that only orientational melting of the outer shell is possible in large clusters.

[1] C.-H. Chiang and L. I, Phys. Rev. Lett. **77**, 647 (1996).
[2] W.-T. Juan, Z.-H. Huang, J.-W. Hsu *et al.*, Phys. Rev. E **58**, R6947 (1998).
[3] A. P. Nefedov, Yu. F. Petrov, and V. E. Fortov, Usp. Fiz. Nauk **167**, 1215 (1997); V. E. Fortov, A. P. Nefedov, Yu. S. Vaulina *et al.*, Zh. Éksp. Teor. Fiz. **114**, 2004 (1998) [JETP **87**, 1087 (1998)].
[4] Y. K. Khodataev, S. A. Krapak, A. P. Nefedov, and Yu. F. Petrov, Phys. Rev. E **57**, 7086 (1998).

[5] V. E. Fortov, A. P. Nefedov, Yu. F. Petrov, A. A. Samaryan, and A. V. Chernyshev, Zh. Éksp. Teor. Fiz. **111**, 467 (1997) [JETP **84**, 256 (1997)].
[6] Yu. E. Lozovik, Usp. Fiz. Nauk **153**, 356 (1987) [Sov. Phys. Usp. **30**, 912 (1987)]; Yu. E. Lozovik and V. A. Mandelstam, Phys. Lett. A **145**, 269 (1990); Phys. Lett. A **165**, 469 (1992).
[7] V. M. Bedanov and F. M. Peeters, Phys. Rev. B **49**, 2667 (1994); V. A. Schweigert and F. M. Peeters, Phys. Rev. B **51**, 7700 (1995); I. V. Schweigert, V. A. Schweigert, and F. M. Peeters, Phys. Rev. B **54**, 10827 (1996).
[8] A. I. Belousov and Yu. E. Lozovik, in *Proceedings of the 1998 Conference on Computational Physics (CCP-1998)*; E-print archive cond-mat/9803300; A. I. Belousov and Yu. E. Lozovik, JETP Lett. **68**, 858 (1998).
[9] B. Partoens, V. A. Schweigert, and F. M. Peeters, Phys. Rev. Lett. **79**, 3990 (1997).
[10] Yu. E. Lozovik and E. A. Rakoch, Phys. Rev. B **57**, 1214 (1998).
[11] L. Cardido, J. P. Rino, N. Studart, and F. M. Peeters, J. Phys.: Condens. Matter **10**, 11627 (1998).
[12] D. P. Resendes, J. T. Mendonca, and P. K. Shukla, Phys. Lett. A **239**, 181 (1998); M. Namby, S. V. Vladimirov, and P. K. Shukla, Phys. Lett. A **203**, 40 (1995).
[13] A. Melzer, V. A. Schweigert, I. V. Schweigert, A. Homann, S. Peters, and A. Piel, Phys. Rev. E **54**, R46 (1996).
[14] P. Pieranski, Phys. Rev. Lett. **45**, 569 (1980).
[15] A. T. Skjeltorp, Phys. Rev. Lett. **51**, 2306 (1983).
[16] Yu. E. Lozovik and V. A. Mandelstam, Phys. Lett. A **138**, 204 (1989).
[17] A. P. Nefedov, private communication.
[18] R. Blümel, J. M. Chen, E. Peik *et al.*, Nature (London) **334**, 309 (1988).
[19] M. Drewsen, C. Brodensen, L. Hornekar, and J. S. Hangst, Phys. Rev. Lett. **81**, 2878 (1998).
[20] A. A. Koulakov and B. I. Shklovskii, Phys. Rev. B **57**, 2352 (1998).
[21] D. J. Wales and R. S. Berry, Phys. Rev. Lett. **73**, 2875 (1994); Phys. Rev. Lett. **63**, 1156 (1989); R. M. Lynden–Bell and D. J. Wales, J. Chem. Phys. **101**, 1460 (1994).

Translated by James S. Wood

Light-induced drift of aerosol particles in gas mixtures

V. G. Chernyak*) and O. V. Klitenik

Ural State University, 620083 Ekaterinburg, Russia
(Submitted 13 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1313–1328 (October 1999)

We analyze the motion of an aerosol particle in a gas mixture in which the molecules of one of the components have been selectively excited as to velocity by resonant optical radiation. We derive expressions for the force with which the gas acts on the particle and for the velocity of particle motion in the Knudsen regime. We also examine the dependence of the force and velocity of photophoresis on the offset of the radiation frequency from the center of the absorption line and on the concentration and mass ratios of the molecules of the absorbing and buffer gases. © *1999 American Institute of Physics.* [S1063-7761(99)01310-4]

1. INTRODUCTION

Studies of phenomena related to the effect of optical radiation on submicrometer particles are of fundamental importance in astrophysics and the physics of the atmosphere, and are of interest in connection with the analysis and cleaning of gas–dust flows.

In the theory of radiometric photophoresis,[1] which results from nonuniform heating of an aerosol particle that absorbs light, the spectral composition of the light is of interest because it is related to the absorptivity of the particle, and accordingly to the temperature field at the particle surface. The light interacts with the particle but not with the gas.

Also of interest is the study of the motion of aerosol particles in the field of optical radiation whose frequency is close to the frequencies of the quantum transitions of the molecules comprising the gas. In this case the light is absorbed by the gas.

Because of the Doppler effect, the radiation is absorbed selectively as to molecular velocity. This gives rise to light-induced fluxes of matter and energy.[2] The difference in the scattering of excited and unexcited molecules by the particle surface and the difference in the transport cross sections of their interaction lead to the emergence of a resonant photophoresis force on the particle.[3,4]

Photophoresis of a particle in a one-component gas is studied in Refs. 3 and 4. There the radiation is taken in the form of a traveling light wave whose frequency is close to that of an electronic or vibrational–rotational transition of the molecules of the surrounding gas. This leads to a drift of the aerosol particle due to the difference in the scattering of excited and unexcited molecules by the particle surface. To obtain the velocity distribution functions of excited and unexcited particles, Derjaguin and Roldughin[3] used the 13-moment approximation of Grad,[5] in which a distribution function is approximated by a finite number of its moments. Thus, the distribution functions of excited and unexcited molecules depend on quantities averaged over the velocity space and neglect significant singularities over a narrow range of velocities, which is determined by the condition for selectivity of the absorbed light.[2] The assumption that far from the particle the gas is spatially homogeneous was used in Ref. 4. There only the accommodation mechanism of photophoresis was examined. This mechanism emerges due to the difference in the momentum accommodation coefficients of the excited and unexcited molecules of the gas. The bulk mechanism, which is due to the difference in the transport characteristics of the molecules in the ground and excited states, was not considered in Ref. 4.

If the aerosol particle is immersed in a mixture of a light-absorbing gas and a buffer gas, there is sure to be an additional factor that determines photophoresis. The total momentum transferred to the aerosol particle also proves to be unbalanced because of the difference in the masses of the molecules comprising the gas mixture, in which the macroscopic fluxes of the absorbing and buffer gases travel in opposite directions.[2] It is also obvious that the value of the photophoresis force depends on the concentration of the components of the gas mixture.

In the present paper we study the photophoresis motion of a particle that does not absorb light or absorbs it homogeneously. The particle is immersed in a binary gas mixture in which the molecules of the absorbing gas are excited selectively as to velocity by resonant optical radiation. We examine the Knudsen regime, in which the particle is much smaller than the mean free path of the molecules.

2. STATEMENT OF THE PROBLEM

We examine a spherical aerosol particle suspended in a binary gas mixture. The system is in thermodynamic equilibrium at temperature T_0.

We irradiate the system with a monochromatic traveling light wave propagating along the z axis (Fig. 1).

Let the radiation be absorbed by the molecules of one of the components in an electronic or vibrational–rotational transition from the ground state n to the first excited state m. The frequency ω of the radiation is slightly offset from ω_{mn}, the center of the absorption line, i.e., $\Omega=\omega-\omega_{mn}$ $\times(|\Omega|\ll\omega,\omega_{mn})$.

Due to the Doppler effect, the radiation is absorbed selectively as to molecular velocity by molecules whose veloc-

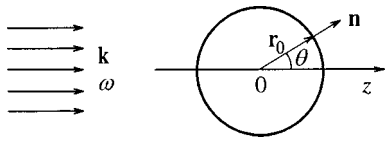


FIG. 1. Problem geometry.

ity projection v_z in the direction of light propagation is close to the resonant value $v_{\rm res}=\Omega/k$, where k is the magnitude of the wave vector $\mathbf{k}$. The effective resonant velocity range $\Delta\mathbf{v}$ is determined by the condition that $\mathbf{k}\cdot\Delta\mathbf{v}\sim\Gamma$, where Γ is the homogeneous halfwidth of the absorption line. The molecules that absorb radiation become excited. As a result, the velocity distribution of the unexcited molecules of the absorbing gas in the range $\Delta\mathbf{v}$ has a Bennett dip, while the corresponding fraction of molecules are in the excited state, forming a Bennett peak.[6] The position of the Bennett peak and dip is determined by the sign and magnitude of the offset Ω of the radiation frequency from the center of the absorption line.

If excited and unexcited vapor molecules interact differently with the surface of the aerosol particle and the molecules of the buffer gas, the decay rate of the Bennett peak and smoothing rate of the Bennett dip will differ. As a result, the overall velocity distribution function for the molecules of the active gas will be non-Maxwellian. At finite offsets Ω, this leads to light-induced drift of the absorbing gas.[2]

Resonant interaction between light and the absorbing gas destroys the state of equilibrium. If the excited and unexcited molecules have different transport characteristics, macroscopic fluxes of excited and unexcited molecules of the absorbing gas as a whole and of the buffer gas emerge in the gas mixture collinear with the wave vector $\mathbf{k}$. These macroscopic fluxes experience different resistances when flowing around the particle because the excited and unexcited molecules of the absorbing gas and the molecules of the buffer gas interact with the particle surface differently and have different transport characteristics and masses. Thus, the gas transfers an unbalanced momentum to the particle, i.e., there is a force on the particles (the photophoresis force). The symmetry of the problem implies that this force is parallel to the z axis (see Fig. 1).

The subscripts 1 and 2 label quantities that refer to the absorbing and buffer gases, respectively, and the subscripts m and n label quantities that refer to excited and unexcited molecules. The absorbing gas consists of molecules with the same mass m_1 but differing effective diameters, $d_n\neq d_m$.

If the velocity distribution functions for the excited (f_m) and unexcited (f_n) molecules of the absorbing gas and for the molecules of the buffer gas (f_2) are known, the photophoresis force can be calculated:

$$F_R=2\pi r_0^2\int_0^\pi d\theta\sin\theta\sum_{i=m,n,2}m_i\int v_r v_z f_i d\mathbf{v}. \tag{1}$$

When the absorbing molecules are approximated by a two-level model, the velocity distribution functions f_m, f_n, and f_2 satisfy the equations of motion[6,7]

$$\frac{\partial f_m}{\partial t}+\mathbf{v}\cdot\nabla f_m=\frac{\chi(\mathbf{v})\Gamma_m(f_n-f_m)}{2}-\Gamma_m f_m+S_m, \tag{2}$$

$$\frac{\partial f_n}{\partial t}+\mathbf{v}\cdot\nabla f_n=-\frac{\chi(\mathbf{v})\Gamma_m(f_n-f_m)}{2}+\Gamma_m f_m+S_n, \tag{3}$$

$$\frac{\partial f_2}{\partial t}+\mathbf{v}\cdot\nabla f_2=S_2. \tag{4}$$

Here $S_i=S_{im}+S_{in}+S_{i2}$ $(i=m,n,2)$,

$$\chi(\mathbf{v})=\frac{4}{\Gamma_m}G^2\frac{\Gamma}{\Gamma^2+(\Omega-\mathbf{k}\cdot\mathbf{v})^2},\quad G=\left|\frac{d_{mn}E_0}{2\hbar}\right|, \tag{5}$$

where Γ is the homogeneous halfwidth of the absorption line, Γ_m is the rate of radiative decay of the excited level, $\chi(\mathbf{v})$ is the probability of absorption per unit time for molecules with a given velocity $\mathbf{v}$ (the absorption rate for such molecules), d_{mn} is the dipole matrix element of the $m-n$ transition, E_0 is the amplitude of the electric field of the light wave, S_{ij} are Boltzmann integrals corresponding to collisions between molecules of species i and j, and $\hbar$ is Planck's constant.

To specify the boundary conditions for Eqs. (2)–(4) we must specify the distribution functions $f_i^+(\mathbf{r}_0,\mathbf{v})$ $(i=m,n,2)$ for molecules emitted from the particle surface. Generally, the temperature at the surface differs from the equilibrium temperature T_0, a fact that can be explained by the absorption of light by the particle. We assume that the particle either does not reflect light, or does reflect it but with its temperature remaining uniform. In that event, no radiometric photophoresis takes place. Henceforth we assume that the particle temperature is that of the gas mixture, T_0.

Let the fraction ε_i of molecules of species i be reflected by the particle surface diffusely with a Maxwellian distribution f_i^s corresponding to the temperature T_0 and number density n_{is}, while the fraction $(1-\alpha_i)$ is elastically reflected. Then, ignoring inelastic collisions with the surface, for molecules of species i we have

$$f_i^+(\mathbf{r}_0,\mathbf{v})=\varepsilon_i f_i^s+(1-\varepsilon_i)f_i^-(2(\mathbf{v}\cdot\mathbf{n})\mathbf{n}),$$

$$i=m,n,2,\quad \mathbf{v}\cdot\mathbf{n}>0,$$

$$f_i^s=n_{is}\left(\frac{m_i}{2\pi k_B T_0}\right)^{3/2}\exp\left\{-\frac{v^2}{\overline{v_i}^2}\right\},\quad \overline{v_i}=\left(\frac{2k_BT_0}{m_i}\right)^{1/2}, \tag{6}$$

where $\mathbf{n}=\mathbf{r}_0/r$ is the outward-directed unit normal to the particle surface (Fig. 1), $f_i^-(\mathbf{v})$ is the distribution function of the molecules incident upon the surface, m_i is the mass of a molecules of species i, and k_B is Boltzmann's constant.

In (6), the unknown number densities n_{is} of the molecules reflected from the surface of the aerosol particle can be found from the nonpercolation conditions

$$N_i^+=|N_i^-|,\quad i=m,n,2. \tag{7}$$

Here N_i^+ and $|N_i^-|$ give the number of molecules of species i incident upon and reflected by unit surface area per unit time:

$$N_i^+=\int_{\mathbf{n}\cdot\mathbf{v}>0}v_r f_i^+(\mathbf{r}_0,\mathbf{v})d\mathbf{v},$$

$$N_i^- = \int_{\mathbf{n}\cdot\mathbf{v}<0} |v_r| f_i^-(\mathbf{r}_0,\mathbf{v})\,d\mathbf{v}. \tag{8}$$

The distribution function f_i^- of the molecules incident upon the particle can be found by solving the equations of motion (2)–(4).

We limit ourselves to analysis of the Knudsen regime, in which the mean free path of molecules is much greater than the particle radius. We can then neglect the perturbation of the distribution functions for the molecules incident upon the particle due to collisions with emitted molecules.

If we assume that the transverse size of the light beam is much greater than the particle radius and the mean free path of the molecules, the intensity of the light will be position-independent, and the distribution functions produced by intermolecular collisions, i.e., far from the particle, will be spatially homogeneous. Then in Eqs. (2)–(4), which the distribution functions f_i^- of the incident molecules must satisfy, we can neglect spatial derivatives.

Let the light be continuous and the state of the gas time-independent. Then in the equations of motion (2)–(4) we can neglect time derivatives.

We examine the case of small values of $\chi(\mathbf{v})$, which is common for vibrational–rotational transitions, while for electronic transitions this case applies to low light intensities.

Note that for $\chi(\mathbf{v}) \ll 1$ the state of the system is close to equilibrium, so that the molecular velocity distribution functions can be written as perturbed Maxwellian distributions:

$$f_i(\mathbf{v}) = f_{i0}[1+h_i(\mathbf{v})],$$
$$f_{i0} = n_i\left(\frac{m_i}{2pk_BT_0}\right)^{3/2}\exp\left\{-\frac{v^2}{\overline{v_i}^2}\right\}, \tag{9}$$

where n_i is the number density of molecules of species i.

Under these assumptions, the linearized variants of Eqs. (2)–(4) assume the form

$$\frac{1}{2}\Gamma_m\chi(\mathbf{v})\left(\frac{n_n}{n_m}-1\right) - \Gamma_m(1+h_m) + L_{mm} + L_{mn} + L_{m2} = 0, \tag{10}$$

$$-\frac{1}{2}\Gamma_m\chi(\mathbf{v})\left(1-\frac{n_m}{n_n}\right) + \frac{n_m}{n_n}\Gamma_m(1+h_m) + L_{nm} + L_{nn} + L_{n2} = 0, \tag{11}$$

$$L_{2m} + L_{2n} + L_{22} = 0, \tag{12}$$

where L_{ij} are the linearized collision integrals.

3. MOLECULAR VELOCITY DISTRIBUTION FUNCTIONS

We consider only elastic collisions between molecules, and for the linearized collision integrals L_{ij} we make approximations that ensure accurate values for the first 13 moments of L_{ij} (see Ref. 8):

$$L_{ij} = -\gamma_{ij}h_i + A_{ij}, \tag{13}$$

where γ_{ij} is the effective collision rate of molecules of species i and j, which is independent of the molecular velocities, and A_{ij} is the approximating inverse-collision integral.

We write an expression for A_{ij} with allowance for the fact that the mean free path of the molecules is much greater than the particle radius. Hence the equations of motion (10)–(12) describe distribution functions produced only by the interactions of the molecules with each other and with light. Then, far from the particle, the vector fluxes of matter and energy, which enter into the approximating inverse-collision integral A_{ij} (Ref. 8), are collinear with the z axis. We also allow for the fact that far from the particle there are no tangential stresses in the gas, but there is light-induced anisotropy of the partial pressures,[7] i.e., $p_{xx}^i = p_{yy}^i \neq p_{zz}^i$. However, the term in the complete expression for A_{ij} (Ref. 8) that contains the stress tensor is proportional to $3\cos^2\theta - 1$, so integrating it with respect to the angle θ in the expression (1) yields no contribution to the force F_R. Hence the term in A_{ij} containing the stress tensor can be neglected.

Due to selective cooling and heating of the components of the gas mixture,[7] the temperatures of these components differ, but they are spatially homogeneous.[9] Hence the terms that take into account this difference in temperatures of the gas components contribute nothing to the photophoresis force and can be neglected in the expression for A_{ij}.

Thus, we can simplify the expression for A_{ij} in Ref. 8:

$$\begin{aligned} A_{ij} = 2c_{iz}[\,&\gamma_{ij}u_i - (u_i - \sqrt{m_i/m_j}\,u_j)\,\nu_{ij}^{(1)} \\ &- (H_i - (m_j/m_i)^{3/2}H_j)\,\nu_{ij}^{(2)}] \\ &+ \frac{8}{5}c_{iz}\left(c_i^2 - \frac{5}{2}\right)\Big[(\gamma_{ij} - \nu_{ij}^{(5)})H_i + \nu_{ij}^{(6)}H_j \\ &- \frac{5}{8}(u_i - \sqrt{m_i/m_j}u_j)\,\nu_{ij}^{(2)}\Big], \end{aligned} \tag{14}$$

where

$$u_i = \frac{U_i}{\overline{v_i}} = \frac{1}{\pi^{3/2}}\int c_{iz}h_i\exp\{-c_i^2\}d\mathbf{c}_i,$$

$$i,j = m,n,2, \quad \mathbf{c}_i = \frac{\mathbf{v}}{\overline{v_i}}, \tag{15}$$

$$H_i = \frac{1}{2\overline{v_i}p_0^i}\left(q_i - \frac{5}{2}\frac{k_BT_0}{m_i}j_i\right) = \frac{1}{2\pi^{3/2}}\int c_{iz}\left(c_i^2 - \frac{5}{2}\right)h_i\exp\{-c_i^2\}d\mathbf{c}_i. \tag{16}$$

Here U_i is the velocity of the ith component of the gas mixture, q_i and j_i are the thermal and diffusive flux of the ith component, p_0^i is the equilibrium partial pressure of the ith component, and the expressions for the rates $\nu_{ij}^{(k)}$ are given in Appendix A.

The solution of Eqs. (10)–(12) with allowance for (13) and (14) takes the form

$$h_m = \frac{\Gamma_m}{\Gamma_m + \gamma_m}\left(\frac{1}{2}\frac{n_n}{n_m}\chi(\mathbf{v}) - 1 + \frac{A_m}{\Gamma_m}\right), \tag{17}$$

$$h_n=\frac{\Gamma_m}{\Gamma_m+\gamma_m}\frac{\gamma_m}{\gamma_n}\left(-\frac{1}{2}\chi(\mathbf{v})+\frac{n_m}{n_n}\right)+\frac{\Gamma_m}{\Gamma_m+\gamma_m}\frac{n_m}{n_n}\frac{A_m}{\gamma_n}+\frac{A_n}{\gamma_n}, \tag{18}$$

$$h_2=\frac{A_2}{\gamma_2}. \tag{19}$$

Here we have introduced the notation

$$A_i=\sum_{j=m,n,2}A_{ij}, \quad \gamma_i=\gamma_{im}+\gamma_{in}+\gamma_{i2}, \tag{20}$$

where γ_i is the effective collision rate of molecules of species i in the gas mixture.

The unknown macroscopic quantities u_i and H_i in Eqs. (17)–(19) can be calculated using (15) and (16). Here we must bear in mind that in an approximation linear in $\chi(\mathbf{v})$ the concentration of the excited molecules is low, i.e., $n_m \ll n_n$ and $n_n \approx n_1$. Furthermore, in (14) we can neglect second-order terms, which are proportional to the rate $\nu_{ij}^{(2)}$ and determine thermal diffusion and diffusive heat transfer (the Dufour effect), if the gas mixture is nonuniform in temperature and concentration.

Under such conditions we obtain

$$u_m=\frac{\Gamma_m}{\Gamma_m+\nu_{mn}^{(1)}+\nu_{m2}^{(1)}}\frac{n_1}{n_m}\frac{\Psi_0}{\pi^{3/2}},$$

$$\Psi_0=\int c_{1z}\chi(\mathbf{v})\exp\{-c_1^2\}d\mathbf{c}_1, \tag{21}$$

$$H_m=\frac{\Gamma_m}{\Gamma_m+\nu_{mn}^{(5)}+\nu_{m2}^{(5)}}\frac{n_1}{n_m}\frac{\Psi_1}{4\pi^{3/2}},$$

$$\Psi_1=\int c_{1z}\left(c_1^2-\frac{5}{2}\right)\chi(\mathbf{v})\exp\{-c_1^2\}d\mathbf{c}_1. \tag{22}$$

The quantities Ψ_0 and Ψ_1, which depend on the offset and broadening parameters $x=\Omega/k\overline{v}_1$ and $y=\Gamma/k\overline{v}_1$, have a simple form in the presence of inhomogeneous ($y\ll 1$) and homogeneous ($y\gg 1$) spectral-line broadening. For intermediate values of y we have done a numerical calculation (see Appendix B).

The dimensionless drift velocity of the absorbing gas far from the particle is

$$u_1=\frac{1}{n_1}(n_nu_n+n_mu_m)=\frac{n_mu_m}{n_1}\frac{n_2}{n_2+(m_1/m_2)^{3/2}n_1}\frac{\nu_{n2}^{(1)}-\nu_{m2}^{(1)}}{\nu_{n2}^{(1)}}. \tag{23}$$

In accordance with momentum conservation for the gas mixture as a whole, the expression for the dimensionless velocity of the buffer gas is

$$u_2=-\left(\frac{m_1}{m_2}\right)^{1/2}\frac{n_1}{n_2}u_1. \tag{24}$$

For the H_i we obtain

$$H_1=\frac{\lambda_2}{\lambda_n\lambda_2-\nu_{n2}^{(6)}\nu_{2n}^{(6)}}\frac{n_mH_m}{n_1}\left(\Delta\nu-\frac{\nu_{2n}^{(6)}}{\lambda_2}(\nu_{n2}^{(6)}-\nu_{m2}^{(6)})\right), \tag{25}$$

$$H_2=\frac{\nu_{2n}^{(6)}}{\lambda_2}\left(H_1-\frac{n_mH_m}{n_1}\frac{\nu_{n2}^{(6)}-\nu_{m2}^{(6)}}{\nu_{n2}^{(6)}}\right), \tag{26}$$

where

$$\lambda_n=\nu_{2n}^{(5)}+\nu_{nn}^{(5)}-\nu_{nn}^{(6)}, \quad \lambda_2=\nu_{2n}^{(5)}+\nu_{22}^{(5)}-\nu_{22}^{(6)},$$
$$\Delta\nu=(\nu_{n2}^{(5)}-\nu_{m2}^{(5)})+(\nu_{nn}^{(5)}-\nu_{mn}^{(5)})+(\nu_{mn}^{(6)}-\nu_{nn}^{(6)}).$$

For the gas mixture as a whole we have

$$H=\frac{n_1H_1+n_2H_2}{n}, \quad n=n_1+n_2. \tag{27}$$

Thus, Eqs. (9) and (17)–(26) completely determine the velocity distribution functions of the molecules incident upon an aerosol particle.

4. FORCE AND VELOCITY OF PHOTOPHORESIS

Using the distribution functions $h(\mathbf{v})$ given by Eqs. (17)–(19) and the adopted boundary conditions (6), we obtain an expression for the photophoresis force (1):

$$F_R=4\pi r_0^2p_1\left(G_\varepsilon\Delta\varepsilon+G_\gamma\frac{\Delta\gamma}{\gamma_n}+G_uu_1+G_HH\right). \tag{28}$$

Here

$$G_\varepsilon=-\frac{\pi^{1/2}}{6}\frac{n_m}{n_1}u_m,$$

$$G_\gamma=-\frac{1}{2\pi}\frac{\Gamma_m}{\Gamma_m+\gamma_m}\left(\frac{10\pi^{1/2}}{3}\Psi_0+\frac{16}{15}\pi^{1/2}\Psi_1+\frac{\overline{\Psi}_2}{\pi^{1/2}}\right)$$

$$G_u=-\frac{1}{6\pi^{1/2}}\left\{20\left[\left(\frac{m_1}{m_2}\right)^{1/2}-1\right]+\pi\left[\left(\frac{m_1}{m_2}\right)^{1/2}\pi\varepsilon_2-1\right]\right\},$$

$$G_H=\frac{8}{15\pi^{1/2}}\frac{n}{n_1}, \tag{29}$$

$$\overline{\Psi}_2=\int_0^\pi d\theta\sin\theta\int_{c_{1r}<0}c_{1r}c_{1z}\chi(\mathbf{v})\exp(-c_1^2)d\mathbf{c}_1,$$

$$\Delta\gamma=\gamma_n-\gamma_m, \quad \Delta\varepsilon=\varepsilon_n-\varepsilon_m, \tag{30}$$

with $p_1=n_1k_BT_0$ the partial pressure of the absorbing gas.

The calculation of $\overline{\Psi}_2$ as a function of the parameters $x=\Omega k\overline{v}_1$ and $y=\Gamma/k\overline{v}_1$ is carried out in Appendix B.

In Eq. (28), the kinetic coefficient G_s characterizes the surface mechanism of photophoresis, which is related to the difference in the condensation coefficients for excited and unexcited molecules of the absorbing gas. The difference in scattering of excited and unexcited molecules by the surface leads to unbalanced momentum transfer to the particle.

The kinetic coefficient G_γ characterizes the contribution related to the differing collision rates of excited and unexcited vapor molecules. A change in the collision rate is related to a change in the effective diameters of molecules participating in optical excitation. This leads to a change in

the mean free path. The difference in the mean free paths of molecules in the ground and excited states leads to a situation in which oppositely directed fluxes of excited and unexcited molecules experience differing resistances when flowing around the particle. As a result, the photophoresis force acquires a frequency-dependent component.

The third and fourth terms in the expression (28) for the photophoresis force, which are proportional to the velocity of light-induced drift and H, also reflect the differing collision rates of molecules of the absorbing gas in the ground and excited states. However, numerical estimates show that the contribution of the fourth term in (28) does exceed 3% in all cases, and for low buffer-gas concentrations, i.e., for $n_2 \ll n_1$, it is less than 1%.

The photophoresis force F_R accelerates an initially static particle, and does so as long as the drag F_c of the gas mixture is less (in absolute value) than F_R. The condition $|F_c| = F_R$ leads to an expression for the terminal velocity U_f of the aerosol particle.

According to Waldmann,[10] the drag in the free-molecule regime with allowance for the low concentration of excited molecules, $n_m \ll n_n$, is

$$F_c = -\frac{8\pi^{1/2}}{3} r_0^2 m_1 n_1 \overline{v}_1 \alpha U_f,$$

$$\alpha = 1 + \varepsilon_n \frac{\pi}{8} + \frac{n_2}{n_1}\left(\frac{m_2}{m_1}\right)^{1/2}\left(1 + \varepsilon_2 \frac{\pi}{8}\right). \tag{31}$$

Thus, the terminal particle velocity U_f with respect to the center of mass of the gas mixture is

$$U_f = \frac{3\pi^{1/2}}{4\alpha}\overline{v}_1\left(G_\varepsilon \Delta\varepsilon + G_\gamma \frac{\Delta\gamma}{\gamma_n} + G_u u_1 + G_H H\right). \tag{32}$$

If we take the effective rates of collisions of the $i-j$ type to be in the form $\gamma_{ij} = \nu_{ij}^{(1)}$ and assume that the interacting molecules are hard elastic spheres, the expressions for the dimensionless velocity of light-induced drift u_1 and the kinetic coefficient G_ε become

$$u_1 = \frac{\Gamma_m}{\Gamma_m + \gamma_m} \frac{n_2}{n_2 + (m_1/m_2)^{3/2} n_1} \frac{\Psi_0}{\pi^{3/2}} \frac{\gamma_{n2} - \gamma_{m2}}{\gamma_{n2}},$$

$$G_\varepsilon = -\frac{\Gamma_m}{\Gamma_m + \gamma_m} \frac{\Psi_0}{6\pi}. \tag{33}$$

Assuming that the values of the effective diameters of excited and unexcited molecules of the absorbing gas are close, i.e., $(d_n - d_m)/d_n \ll 1$, we obtain

$$\frac{\Delta\gamma}{\gamma_n} = A\frac{\Delta d}{d_n}, \quad \frac{\gamma_{n2} - \gamma_{m2}}{\gamma_{n2}} = \frac{\Delta d}{d_{n2}},$$

$$\Delta d = d_n - d_m, \quad d_{ij} = \frac{d_i + d_j}{2}, \tag{34}$$

where

$$A = \frac{\sqrt{m_1/2}\, n_1 d_n^2 + \sqrt{m_{12}}\, n_2 d_n d_{n2}}{\sqrt{m_1/2}\, n_1 d_n^2 + \sqrt{m_{12}}\, n_2 d_{n2}^2}, \quad m_{12} = \frac{m_1 m_2}{m_1 + m_2}. \tag{35}$$

Then the expression (29) for the photophoresis force can be written

$$F_R = 4\pi r_0^2 p_1\left(G_\varepsilon \Delta\varepsilon + G_d \frac{\Delta d}{d_n}\right), \tag{36}$$

where

$$G_d = AG_\gamma + G_u \frac{\Gamma_m}{\Gamma_m + \gamma_m} \frac{n_2 d_n/d_{n2}}{n_2 + (m_1/m_2)^{3/2} n_1} \frac{\Psi_0}{\pi^{3/2}}. \tag{37}$$

The second term in parentheses in (36) reflects the bulk mechanism of photophoresis, which is related to the differing transport properties of the excited and unexcited molecules of the absorbing gas.

The particle velocity is

$$U_f = \overline{v}_1\left(\mu_\varepsilon \Delta\varepsilon + \mu_d \frac{\Delta d}{d_n}\right), \tag{38}$$

where μ_ε and μ_d are the kinetic coefficients characterizing, respectively, the surface and bulk mechanisms of photophoresis velocity:

$$\mu_\varepsilon = \frac{3\pi^{1/2}}{4\alpha} G_\varepsilon, \quad \mu_d = \frac{3\pi^{1/2}}{4\alpha} G_d. \tag{39}$$

The expressions for the kinetic coefficients G_ε and G_d become much simpler for inhomogeneous and homogeneous line broadening.

4.1. Inhomogeneous broadening

Such broadening ($y \ll 1$) is typical of rarefied gases. In this case we obtain

$$G_\varepsilon = -\frac{2\pi G^2}{3(\Gamma_m + \gamma_m) k \overline{v}_1} x \exp(-x^2), \tag{40}$$

$$G_\gamma = -\frac{2G^2}{(\Gamma_m + \gamma_m) k \overline{v}_1} \times \left[\frac{2}{3}\left(\frac{13}{5}\pi^{3/2} - 1\right)x - \frac{2\pi^{3/2} - 1}{3} x^3 + \cdots\right]. \tag{41}$$

The term in the expression (29) for G_γ containing $\overline{\Psi}_2$ provides a contribution of at most 5%. Neglecting it, we obtain instead of (41)

$$G_\gamma = -\frac{4\pi^{3/2} G^2}{15(\Gamma_m + \gamma_m) k \overline{v}_1}(13 + 8x^2) x \exp(-x^2). \tag{42}$$

The coefficient G_d then becomes

$$G_d = -\frac{2\pi^{1/2} G^2}{(\Gamma_m + \gamma_m) k \overline{v}_1}\left[A\frac{2\pi}{15}(13 + 8x^2) - G_u \frac{2 n_2 d_n/d_{n2}}{n_2 + (m_1/m_2)^{3/2} n_1}\right] x \exp(-x^2). \tag{43}$$

4.2. Homogeneous broadening

In this case ($y \gg 1$) for $|x| \ll y$ we obtain

$$G_\varepsilon = -\frac{2\pi^{1/2} G^2}{3(\Gamma_m + \gamma_m)\Gamma} \frac{x}{y^2},$$

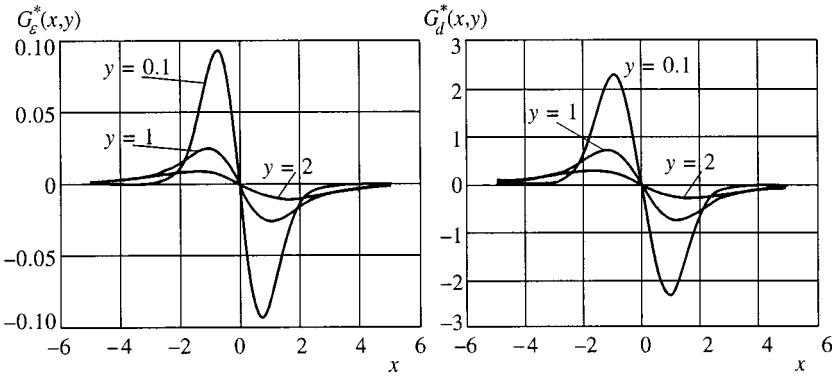


FIG. 2. Dependence of the modified kinetic coefficients G_ε^* (a) and G_d^* (b) on the offset and broadening parameters (x and y, respectively) at $n_1 \approx n_2$, $m_1 \approx m_2$; and $d_n = d_2$.

$$G_\gamma = -\frac{2G^2}{3\pi^{1/2}(\Gamma_m+\gamma_m)\Gamma}(10\pi^{3/2}-4)\frac{x}{y^2},$$

$$G_d = -\frac{2G^2}{(\Gamma_m+\gamma_m)\Gamma}\left[A\,\frac{10\pi^{3/2}-4}{3\pi^{1/2}} - G_u\,\frac{2n_2 d_n/d_{n2}}{n_2+(m_1/m_2)^{3/2}n_1}\right]\frac{x}{y^2}. \quad (44)$$

5. DISCUSSION

The kinetic coefficients G_ε and G_d, which characterize the surface (accommodation) and bulk mechanisms of photophoresis of an aerosol particle, are proportional to the radiative intensity, and depend on the ratio of the rate of radiative decay of the excited level, Γ_m, to the intermolecular collision rate γ_m, the frequency offset Ω from the center of the absorption line, the component concentrations, and the molecular mass ratio.

Figure 2 depicts the dependence of the modified kinetic coefficients on the offset parameter $x=\Omega/k\overline{v_1}$ and the broadening parameter $y=\Gamma/k\overline{v_1}$:

$$G_\varepsilon^* = \left(\frac{8G^2}{k\overline{v_1}(\Gamma_m+\gamma_m)}\right)^{-1} G_\varepsilon,$$

$$G_d^* = \left(\frac{4G^2}{k\overline{v_1}(\Gamma_m+\gamma_m)}\right)^{-1} G_d.$$

As in the phenomenon of light-induced drift,[2] the directions of the force and velocity of photophoresis depend on the sign of the offset Ω, i.e., the kinetic coefficients G_ε and G_d and hence μ_ε and μ_d are odd functions of Ω (Fig. 2). Here, in the case of inhomogeneous broadening, the photophoresis velocity peaks at $\Omega \approx k\overline{v_1}$. When $|\Omega| > 3k\overline{v_1}$, the effect is negligible, since the condition for resonance is not satisfied and only a few molecules are excited via light absorption.

The kinetic coefficients G_ε and G_d are positive if the offset parameter $x=\Omega/k\overline{v_1}$ is negative. Conversely, when $x>0$ both G_ε and G_d are negative.

Thus, the directions of the surface component of the force and velocity of photophoresis are determined by the signs of the difference of the coefficients of accommodation of the unexcited and excited molecules, $\Delta\varepsilon = \varepsilon_n - \varepsilon_m$, and the frequency offset Ω. If $\Delta\varepsilon > 0$, the direction of the surface component coincides with the direction of light propagation when $\Omega<0$, and is opposite the direction of light propagation when $\Omega>0$.

To explain this result, we examine a specific case in which the accommodation coefficient of the molecules that have absorbed light decreases, $\Delta\varepsilon>0$, and the frequency offset is positive, $\Omega>0$. This leads to a situation in which oppositely directed fluxes of excited and unexcited molecules are generated. The first flux is collinear with the wave vector ($u_m>0$), while the second flux travels in the opposite direction ($u_n<0$). In view of the fact that $\varepsilon_n > \varepsilon_m$, the flux of the unexcited molecules experiences greater resistance when flowing around the aerosol particle than the flux of the excited molecules. The resultant force on the particle points in the same direction as the flux of unexcited particles, i.e., opposite the direction of light propagation.

The directions of the bulk component of the force and velocity of photophoresis are determined by the signs of the difference of the effective diameters of the unexcited and excited molecules, $\Delta d = d_n - d_m$, and the frequency offset Ω. If $\Delta d>0$, the direction of the bulk component coincides with the direction of light propagation when $\Omega<0$ and is opposite that direction when $\Omega>0$.

By way of explanation, let Ω be positive and the cross section of molecular collisions increase as a result of optical excitation. This reduces the mean free path of molecules that have absorbed light. Then the flux of excited molecules, pointing in the same direction as the wave vector, experiences a greater resistance when flowing around the particle than the flux of molecules in the ground state, so the aerosol particle is dragged in the direction of light propagation.

We now consider the dependence of the force F_R and velocity U_f of an aerosol particle on the ratios of the number densities, $z=n_2/n_1$, and the masses, $b=m_1/m_2$, of the molecules of the absorbing and buffer gases. Figure 3 depicts the results of numerical calculations of the U_f vs. z dependence. Clearly, as the concentration of the buffer gas increases at some given number density of the mixture ($n=n_1+n_2=\text{const}$), the velocity of the aerosol particle decreases.

Combining (38) and (39), we obtain an approximate expression for the photophoresis velocity at a given pressure of the gas mixture:

$$U_f(z) \approx \frac{U_f(z=0)}{1+b^{-1/2}z}, \quad (45)$$

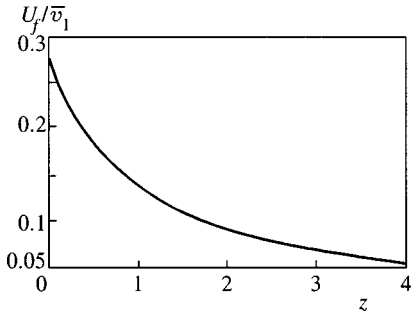


FIG. 3. Dependence of the photophoresis velocity on the ratio of the concentrations of the buffer and absorbing gases for $x=0.75$, $m_1 \approx m_2$, $\gamma_m \ll \Gamma_m$, $d_n \approx d_2$, $\varepsilon_n=\varepsilon_2=1$, and $\Delta\varepsilon=\Delta d/d_n=-0.1$.

where $U_f(z=0)$ is the photophoresis velocity in the absence of a buffer gas. Note that Eq. (45) provides a satisfactory approximation to the results of numerical calculations.

The dependence of the photophoresis velocity on the ratio of molecular masses of the absorbing and buffer gases, $b=m_1/m_2$, is shown in Fig. 4. Clearly, the particle velocity increases monotonically with b for $b \leqslant 20$. For a low-mass absorbing gas and high-mass buffer gas ($m_1 \ll m_2$), the photophoresis velocity is proportional to $\sqrt{m_1/m_2}$. For a heavy absorbing gas and a light buffer gas, on the other hand, the photophoresis velocity is independent of b, and is equal to the particle velocity in a pure absorbing gas.

We now make some numerical estimates for a specific system under typical experimental conditions. We consider two cases: an aluminum particle of radius $r_0 \sim 1\ \mu$m immersed in sodium vapor, and the same particle immersed in an equimolar mixture of sodium vapor and the inert gas argon; in both cases the temperature is $T_0=700$ K. The saturated vapor pressure is 306 Pa.[11]

The molecular characteristics of Na and Ar are $m_1=3.82\times10^{-26}$ kg, $m_2=6.64\times10^{-26}$ kg, $d_1=3.0$ Å, and $d_2=3.4$ Å; the mean velocity of the sodium molecules is $\overline{v}_1 \approx 711\ \mathrm{m\,s^{-1}}$.

The source of light is a tunable dye laser emitting in the vicinity of the D_1 and D_2 lines of sodium. The wavelength λ is 600 nm.

In this case the frequency of the incident light turns out to be much greater than the electron collision rate in the metal, and less than the plasma frequency. Langmuir screening[12] therefore takes place, i.e., light is reflected by the particle with essentially no absorption. Hence, under these conditions the model of a nonabsorbing particle usually adopted in theoretical studies is realized.

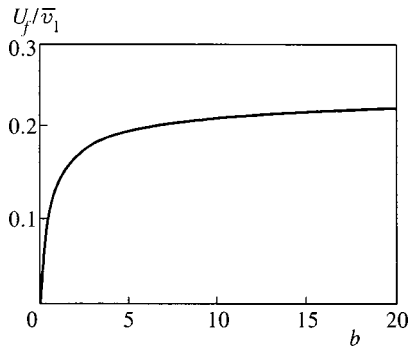


FIG. 4. Dependence of the photophoresis velocity on the ratio of the masses of the molecules of the buffer and absorbing gases for $x=0.75$, $z=1$, $\gamma_m \ll \Gamma_m$, $d_n \approx d_2$, $\varepsilon_n=\varepsilon_2=1$, and $\Delta\varepsilon=\Delta d/d_n=-0.1$.

Let the radiative power be 10 mW and the beam diameter about 1 mm. At that intensity the Rabi frequency G is approximately 10^8 Hz. The radiative decay rate is $\Gamma_m=6\times10^7$ Hz (see Ref. 13). Finally, the Doppler broadening, $k\overline{v}_1 \approx 7.3\times10^9$ Hz, corresponds to inhomogeneous line broadening, $\Gamma/k\overline{v}_1 \approx 0.01$.

1. *Pure gas* ($n_2=0$). In this case the collision rate of the excited molecules and the Knudsen number Kn are $\gamma_m \approx 7.1\times10^6$ Hz and $\mathrm{Kn} \approx 100$, which corresponds to the free-molecule regime.

Numerical estimates of the kinetic coefficients yield

$$G_\varepsilon = -0.042x\exp(-x^2),$$

$$G_d = -0.030(13+8x^2)x\exp(-x^2).$$

Let $x=\Omega/k\overline{v}_1=0.8$ and $\Delta\varepsilon=\Delta d/d_n=-0.1$. Then $G_\varepsilon \approx -0.016$ and $G_d \approx -0.228$. For the photophoresis force we obtain $F_R \approx 9.3\times10^{-11}$ N. Note that the gravitational force on the particle in question is $Mg \approx 1.1\times10^{-14}$ N, i.e., about 0.01% of the photophoresis force.

Assuming $\varepsilon_n=1$, the velocity acquired by the particle is then $U_f \approx 0.03\overline{v}_1 \approx 20\ \mathrm{m\,s^{-1}}$, which is about 6% of the speed of sound.

2. *Equimolar mixture* Na+Ar ($n_1=n_2$). The numerical values of the main parameters are the same as in the previous case. Then Eq. (38) yields an estimate for the particle velocity, $U_f \approx 7\ \mathrm{m\,s^{-1}}$.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 99-01-00143).

APPENDIX A

The expressions for the rates $\nu_{\alpha\beta}^{(k)}$ that enter into Eq. (13) have the form[8]

$$\nu_{\alpha\beta}^{(1)} = \frac{16}{3}\frac{m_{\alpha\beta}}{m_\alpha} n_\beta \Omega_{\alpha\beta}^{(1,1)},$$

$$\nu_{\alpha\beta}^{(2)} = \frac{64}{15}\left(\frac{m_{\alpha\beta}}{m_\alpha}\right)^2 n_\beta \left(\Omega_{\alpha\beta}^{(1,2)} - \frac{5}{2}\Omega_{\alpha\beta}^{(1,1)}\right),$$

$$\nu_{\alpha\beta}^{(5)} = \frac{64}{15}\left(\frac{m_{\alpha\beta}}{m_\alpha}\right)^3 \frac{m_\alpha}{m_\beta} n_\beta \left[\Omega_{\alpha\beta}^{(2,2)} + \left(\frac{15}{4}\frac{m_\alpha}{m_\beta} + \frac{25}{8}\frac{m_\beta}{m_\alpha}\right)\Omega_{\alpha\beta}^{(1,1)} - \frac{1}{2}\frac{m_\beta}{m_\alpha}(5\Omega_{\alpha\beta}^{(1,2)} - \Omega_{\alpha\beta}^{(1,3)})\right],$$

$$\nu_{\alpha\beta}^{(6)} = \frac{64}{15}\left(\frac{m_{\alpha\beta}}{m_\alpha}\right)^3 \left(\frac{m_\alpha}{m_\beta}\right)^{3/2} n_\beta \left[-\Omega_{\alpha\beta}^{(2,2)} + \frac{55}{8}\Omega_{\alpha\beta}^{(1,1)} - \frac{5}{2}\Omega_{\alpha\beta}^{(1,2)} + \frac{1}{2}\Omega_{\alpha\beta}^{(1,3)}\right].$$

Here $m_{\alpha\beta}=m_\alpha m_\beta/(m_\alpha+m_\beta)$ is the reduced mass of molecules of species α and β, and the $\Omega_{\alpha\beta}^{(l,r)}$ are the Chapman–Cowling integrals,[14] which depend on the shape of the pair-

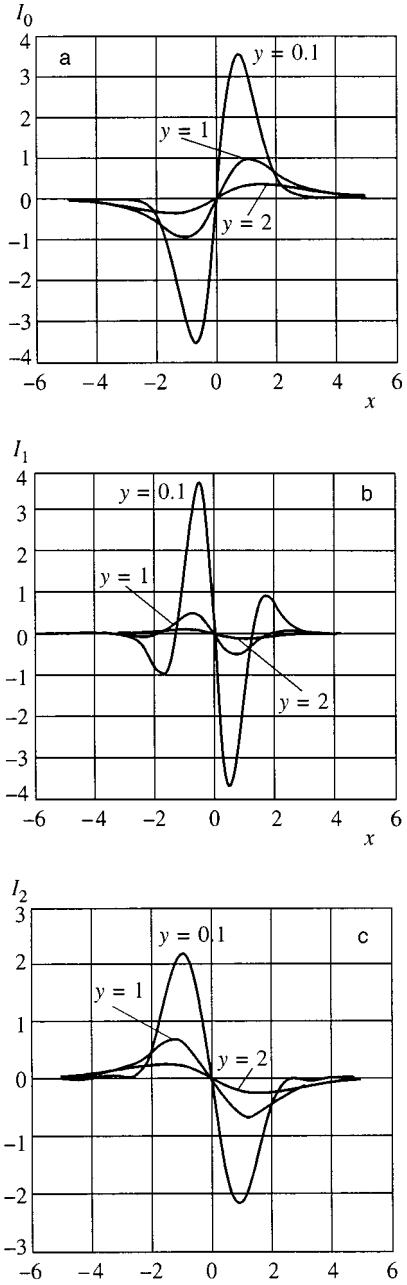


FIG. 5. Dependence of the quantities I_0 (a), I_1 (b), and I_2 (c) on the parameters x and y.

wise potential between molecules of species α and β. In particular, for the molecular model of hard spheres of diameter d_α we have[14]

$$\Omega_{\alpha\beta}^{(l,r)}=\left(\frac{k_BT}{2\pi m_{\alpha\beta}}\right)^{1/2}\frac{(r+1)!}{2}\left[1-\frac{1+(-1)^l}{2(l+1)}\right]\pi d_{\alpha\beta}^2,$$

$$d_{\alpha\beta}=\frac{d_\alpha+d_\beta}{2}.$$

APPENDIX B

Here we find the dependence of Ψ_0, Ψ_1, and $\overline{\Psi}_2$, which enter into Eqs. (21), (22), and (30), on the offset and broadening parameters $x=\Omega/k\overline{v}_1$ and $y=\Gamma/k\overline{v}_1$. Let

$$\Psi_0=\frac{4G^2}{\Gamma_m k\overline{v}_1}I_0(x,y),\quad \Psi_1=\frac{4G^2}{\Gamma_m k\overline{v}_1}I_1(x,y),$$

$$\overline{\Psi}_2=\frac{4G^2}{\Gamma_m k\overline{v}_1}I_2(x,y).$$

For inhomogeneous broadening ($y\ll 1$) we have the approximate expression

$$\chi(\mathbf{v})=\frac{4G^2}{\Gamma_m}\pi\delta(\Omega-\mathbf{k}\cdot\mathbf{v}).$$

Simple calculations yield

$$I_0(x,y)=\pi^2 x\exp(-x^2),$$

$$I_1(x,y)=\pi^2 x\left(x^2-\frac{3}{2}\right)\exp(-x^2),$$

$$I_2=\frac{\pi^{3/2}}{3}(-2x+x^3+\cdots).$$

For homogeneous broadening ($y\gg 1$) with $x/y\ll 1$, to within terms linear in x/y,

$$\chi(\mathbf{v})=\frac{4G^2}{\Gamma_m\Gamma}\left[1-\left(\frac{x-c_{1z}}{y}\right)^2\right],\quad c_{1z}=\frac{v_z}{\overline{v}_1},$$

$$I_0=\pi^{3/2}\frac{x}{y^3},\quad I_1=0,\quad I_2=-\frac{4\pi}{3}\frac{x}{y^3}.$$

The results of numerical calculations of I_0, I_1, and I_2 for various values of the parameters x and y are plotted in Fig. 5.

*)E-mail: vladimir.chernyak@usu.ru

[1] V. Chernyak and S. Beresnev, J. Aerosol Sci. **24**, 857 (1993).
[2] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[3] B. V. Derjaguin and V. I. Roldughin, J. Colloid Interface Sci. **122**, 537 (1988).
[4] V. G. Chernyak, Zh. Éksp. Teor. Fiz. **96**, 878 (1989) [Sov. Phys. JETP **69**, 498 (1989)].
[5] H. Grad, Commun. Pure Appl. Math. **2**, 331 (1949).
[6] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).
[7] F. Kh. Gel'mukhanov and L. V. Il'ichev, Khim. Fiz. No. **3**, 1544 (1984).
[8] F. J. McCormack, Phys. Fluids **16**, 2095 (1973).
[9] V. G. Chernyak and O. V. Klitenik, Zh. Éksp. Teor. Fiz. **113**, 1036 (1998) [JETP **86**, 565 (1998)].
[10] L. Waldmann, Z. Naturforsch. Teil A **14**, 589 (1983).
[11] N. B. Vargaftik, *Tables on the Thermophysical Properties of Liquids and Gases: In Normal and Dissociated States*, 2nd ed., Halsted Press, New York (1975).

[12] E. A. Turov, *Constitutive Equations of Electrodynamics* [in Russian], Nauka, Moscow (1972).
[13] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions*, Springer-Verlag, Berlin (1985).
[14] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, 3rd ed., Cambridge Univ. Press, New York (1991).

Translated by Eugene Yankovsky

Experimental indication of macroscopic polarization parallel to the tilt plane in free-standing films of ferroelectric liquid crystals 8SI* and DOBAMBC

P. O. Andreeva and V. K. Dolganov*)

Institute of Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

C. Gors and R. Fouret

Laboratoire de Dynamique et Structures des Matériaux Moleculaire, Université de Lille 1, UFR de Physique, 59655 Villeneuve d'Ascq, France

(Submitted 25 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1329–1334 (October 1999)

Optical reflectivity and x-ray studies have been conducted on free-standing films and bulk samples of ferroelectric liquid crystals near the smectic-*A*–smectic-*C** transition. A tilt plane rotation with respect to the direction of an applied electric field is found in the ferroelectric films above the bulk transition temperature. Whereas the macroscopic polarization is perpendicular to the tilt plane at low temperature, it is parallel to the tilt plane at elevated temperature. The temperature dependence of the average tilt angle is measured. © *1999 American Institute of Physics.* [S1063-7761(99)01410-9]

The smectic-*C** (Sm*C**) phase in a free-standing film is a subject of much interest.[1–8] In this phase the direction of the long molecular axis is tilted with respect to the normal to the smectic layers. In the ferroelectric Sm*C** phase, spontaneous polarization is perpendicular to the tilt plane and to the **c**-director (**c** is parallel to the tilt plane and the layer). Recently, the first observation of longitudinal (i.e., parallel to the tilt plane) polarization has been reported in free suspended films of an antiferroelectric smectic phase[9] and ferroelectric phase.[10,11]

In this paper we demonstrate an unusual behavior of the well-known ferroelectric compounds 8SI* and DOBAMBC in Sm*C** free-standing films the polarization can be parallel to the tilt plane, i.e., the molecular tilt plane is parallel to an electric field. The temperature dependence of the tilt angle is determined for structures with transverse and longitudinal polarization.

Our samples were the compounds (S)-4-(2′-methylbutyl) phenyl 4-(*n*-octyl)biphenyl-4-carboxylate (8SI*) and *p*-decyloxybenzylidene *p*-aminocinnamic acid 2-methylbutyl ester (DOBAMBC). In the bulk sample of 8SI* the following phase sequence was observed: Sm*I**–(66 °C)–Sm*C**–(82 °C)–SmA–(134 °C)–Ch–(140 °C)–I. DOBAMBC showed transition temperatures Sm*I**–(73 °C)–Sm*C**–(90.5 °C)–Sm*A*–(115 °C)–I.

Thick freely suspended films were prepared by drawing the liquid crystal in the smectic state over a 6-mm hole in a glass plate. Thin films were prepared by layer-by-layer thinning.[12–14] In a free-standing film the smectic layers are aligned parallel to the film surface. An electric field of 3 to 20 V/cm in the plane of the film was used to align the tilt direction in the Sm*C** state. The incident beam was linearly polarized and was perpendicular to the film surface. An electric field could be applied in two mutually perpendicular directions [parallel ($E_\parallel$) and perpendicular ($E_\perp$) to the plane of polarization of the light]. For thick films, the layer number N was determined from the spectral dependence of the optical reflection in the Sm*A* phase [15]

$$I(\lambda)=\frac{(n^2-1)^2\sin^2(2\pi nNd/\lambda)}{4n^2+(n^2-1)^2\sin^2(2\pi nNd/\lambda)}, \tag{1}$$

where d is the interlayer spacing. In the Sm*A* phase, $n=n_0$ (n_0 is the ordinary index of refraction). For thin films, the number of layers was determined from the relative intensity of reflections for films with a different number of layers:

$$I(\lambda)\simeq N^2\pi^2d^2(n^2-1)^2/\lambda^2. \tag{2}$$

In the Sm*C** phase, two values of the optical reflection (I_p and I_0) have been measured [for the plane of polarization of the light oriented parallel (I_p) and perpendicular (I_0) to the tilt plane]. For our calculations of the average tilt angle θ we used the value of the birefringence in the Sm*A* phase measured by Musevic *et al.*[16] X-ray diffraction studies on bulk samples were made using a curved linear position-sensitive multidetector ($\lambda=1.5406$ Å).

Figure 1 shows optical reflection intensities ($I_\parallel, I_\perp$) from 2- and 20-layer films: the electric field was applied parallel ($\mathbf{E}_\parallel$) and perpendicular ($\mathbf{E}_\perp$) to the direction of polarization of the light. As previously observed,[1,2] the Sm*C**–Sm*A* transition temperature is a function of film thickness ($T\approx 112$ °C for $N=2$; $T\approx 97$ °C for $N=20$). For two-layer film the observed change in intensity, when the direction of the electric field is switched ($\mathbf{E}_\parallel \leftrightarrow \mathbf{E}_\perp$), corresponds quantitatively to our conventional view of the Sm*C** phase: the ferroelectric polarization is perpendicular to the tilt plane. Since the refractive index n_p (polarization of the light is parallel to the tilt plane) is greater than n_0, the larger value of the reflection intensity (Fig. 1, $N=2$, $T=81$ °C and $T=106$ °C) corresponds to the direction perpendicular to the field [see Eqs. (1) and (2)]. At low temperatures, all films

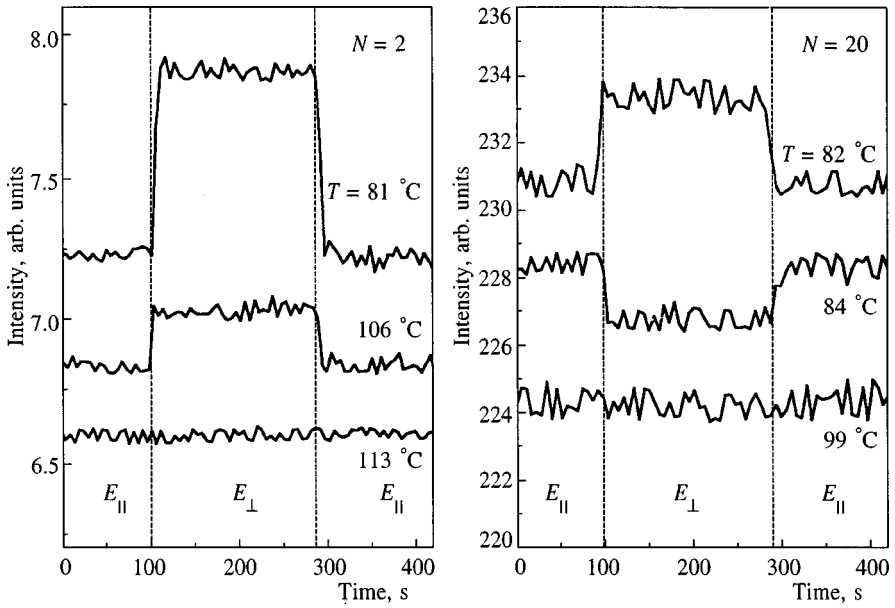


FIG. 1. Optical reflection intensity for thin ($N=2$) and thick ($N=20$) ferroelectric films in an electric field applied parallel ($\mathbf{E}_{\|}$) and perpendicular ($\mathbf{E}_{\perp}$) to the direction of polarization of the light ($\lambda=600$ nm). For thick film ($N=20$), the change in reflection intensity when the direction of the electric field is changed is different at low ($T=82\,^{\circ}$C) and high ($T=84\,^{\circ}$C) temperatures (8SI*).

show the same change in the reflected intensities (Fig. 1, $N=20$, $T=82\,^{\circ}$C). Quantitatively different behavior is observed for thick SmC^* films at high temperature (Fig. 1, $N=20$, $T=84\,^{\circ}$C). The reflection intensity $I_{\perp}$ is less than $I_{\|}$. This means that the tilt plane and **c**-director are oriented in the **E**-field direction.

Figure 2 provides additional evidence for anomalous field-induced orientation of the **c**-director. The wavelength λ_m of the reflectivity minimum for thick films depends on the index of refraction: the larger value of λ_m corresponds to the larger value of n [$\lambda_m=2Nnd$; see Eq. (1)]. At the temperatures below $T_i\approx 91\,^{\circ}$C (Fig. 2a, DOBAMBC), the wavelength of reflectivity minimum λ_m for the field $\mathbf{E}_{\perp}$ is greater than λ_m for $\mathbf{E}_{\|}$ (**c**-director oriented perpendicular to the **E**-field). At $T_i\approx 91\,^{\circ}$C, the reflection spectra belonging to $\mathbf{E}_{\perp}$- and $\mathbf{E}_{\|}$-field exchange position (Fig. 2b). This interchange of the spectra indicates that at temperatures above $T_i\approx 91\,^{\circ}C$ the **c**-director is parallel to the field.

We observed anomalous orientation in ferroelectric films with thickness between 12 and some hundreds of layers. For $N>40$-layer films, T_i was nearly the same as the bulk transition temperature (SmC^*–SmA). In thinner films, T_i was shifted to higher temperatures, but could not be precisely determined because of temperature hysteresis. Figure 3 shows the temperature dependence of the relative optical reflection intensities $I_{\perp}/I_{\|}$ for films 2, 5, and 20 smectic layers thick. An anomalous orientation (**c**-director is parallel to the field) corresponds to $I_{\perp}/I_{\|}<1$. The arrows show the behavior of the relative intensities $I_{\perp}/I_{\|}$ in the temperature range for hysteresis (on heating and cooling). At temperatures above 90 °C, the anomalous orientation is observed in a very weak electric field (<10 V/cm) without a transition between

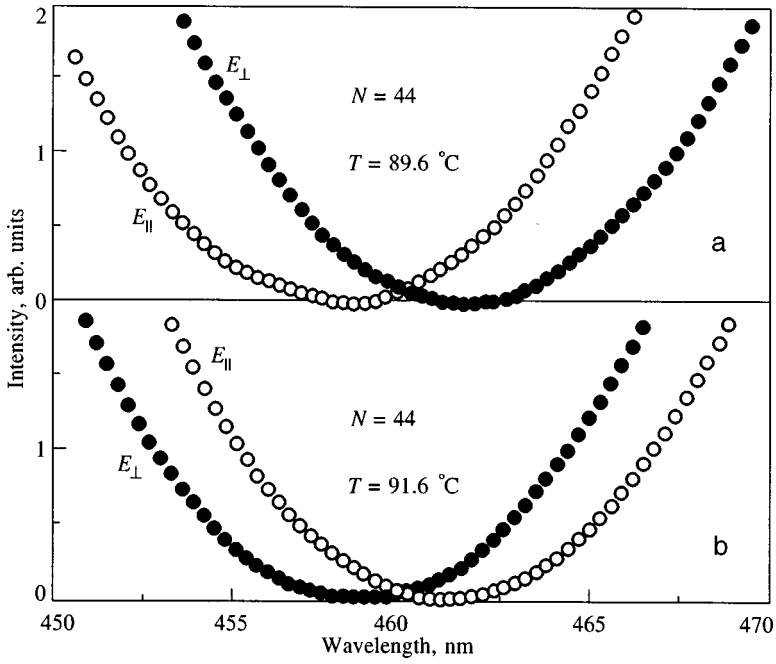


FIG. 2. Reflection spectra at the temperature below (a) and above (b) the bulk transition temperature. The relative location of the reflectivity minima λ for $\mathbf{E}_{\perp}$- and $\mathbf{E}_{\|}$-field directions differs for temperatures below and above the bulk transition (DOBAMBC). In the figure, the intensities are shown with respect to the intensities in the minima of the curves.

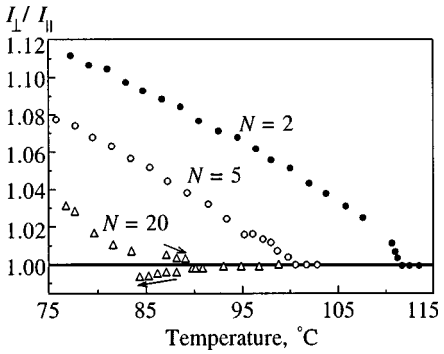


FIG. 3. Temperature dependence of the optical reflection intensities $E_\perp / E_\parallel$ for films with 2, 5, and 20 smectic layers ($\lambda = 600$ nm). Arrows show the behavior of the intensity for a thick film ($N = 20$) in the temperature range of hysteresis (on heating and cooling), 8SI*.

the two orientations. The temperature range of hysteresis can be reduced by turning an electric field on and off. At high enough fields, no perfectly oriented film is observed because of convection instability.[16] This precludes determining the **E**-field threshold of reorientation in the temperature range for hysteresis.

Our results can be interpreted with the aid of Fig. 4. Here we indicate the smectic layers in the SmC^* phase (a), SmA phase (d), and the layer structure above the bulk transition temperature (b,c; $T_c > T_b$). The possibility of spatial variation of the tilt angle across the film thickness (Figs. 4b and 4c) is related to the surface ordering of the smectic layers.[1–8] At high temperatures (Fig. 4c), only some surface layers are tilted. When a free-standing film is cooled, all layers become tilted (Fig. 4b) because of the diverging correlation length ξ of the SmC^* ordering.

The variation of the tilt angle across the film must play an important role for the analyses of the origin of the spontaneous polarization above the bulk transition temperature.[10,11] The change in molecular tilt angle implies S-shaped bending of the liquid crystal director n (Fig. 4b). This should lead to macroscopic flexoelectric polarization. The polarization density in a nematic liquid crystal is given by[17]

$$\mathbf{P}_f = e_1(\mathbf{n}\,\mathrm{div}\,\mathbf{n}) + e_3[(\mathrm{curl}\,\mathbf{n})\mathbf{n}], \tag{3}$$

where e_1 and e_3 are the flexoelectric constants.

In smectic liquid crystals, other kinds of deformations can cause flexoelectric polarization. However, for our estimate we use Eq. (3). For a bend deformation, the second term is dominant. Using $e_3 = 10^{-11}$ C/m[17] and a variation of the tilt angle from the surface layer to the center of the film $\Delta\theta \approx 0.3$ rad, we obtain from Eq. (3) that $P_f = 10^{-4}$ C/m^2 ($N = 20$). Our estimate shows that above the bulk transition temperature, the flexoelectric polarization in the tilt plane P_f and the bulk ferroelectric polarization in the SmC^* phase P_0[18] are of the same order of magnitude.

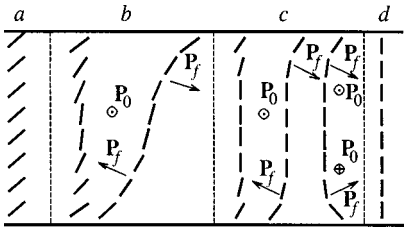


FIG. 4. Orientation of molecules in a smectic film: SmC^* phase (a), S-like orientation of the director near the bulk SmC^*–SmA transition temperature (b), S- and C-like orientation well above the bulk transition temperature (c), SmA phase (d).

With increasing temperature, two effects take place: first, P_0 decreases and P_f can be greater than P_0; second, the tilt angle at the center of the film becomes zero (or small). The top and bottom parts of the film, from the point of view of the molecular tilt, are not (or weakly) coupled to each other and can be oriented independently. When $P_f > P_0$, a C-shaped orientation (Fig. 4c) is favored in an electric field. Reversal of the molecular tilt at the bottom of the film results in reversal of the ferroelectric (P_0) and flexoelectric (P_f) polarization (Fig. 4c). For $P_f \gg P_0$ the net polarization is parallel to the **c**-director, and, a $S \leftrightarrow C$ transition leads to a 90 °C reorientation of the tilt planes with respect to the direction of the **E**-field.

Optical reflectivity is a convenient vehicle for determining of the average molecular tilt θ.[11] The value I_p/I_0 depends on the reflection n_p index in the tilt plane [Eqs. (1), (2)] and on the layer spacing in the SmC^* phase d_C [Eq. (1)]. Considering the molecules to be rigid rods, d_C can be taken as

$$d_C = d_A \cos\theta, \tag{4}$$

where d_A is the layer spacing in the SmA phase. Near the temperature of the bulk phase transition, d_A was determined to be 3.0 nm (x-ray data, 8SI*, 83 °C). Using results derived for uniaxial crystals,[15] as in the case of ellipsometric studies,[3] the relation between θ and n_p is taken to be

$$\cos^2\theta = \frac{n_0^2(n_e^2 - n_p^2)}{n_p^2(n_e^2 - n_0^2)}. \tag{5}$$

When n_0, n_e, and d_A are known, I_p/I_0 alone is sufficient to determine the average tilt angle θ. It should be underscored that $I_p/I_0 = I_\perp/I_\parallel$ for the usual orientation and $I_p/I_0 = I_\parallel/I_\perp$ for anomalous orientation. Equations (1), (4), and (5) [or (2) and (5) for thin films] enable one to determine θ. The values of the average tilt angle θ resulting from the $I_\perp/I_\parallel$ values (Fig. 3) are shown in Fig. 5. Figure 5 also shows the temperature dependence of θ obtained from our x-ray measurements on a bulk sample. At low temperatures, the data for 20-layer film are in good agreement with measurements of the bulk sample. For the 20-layer film the change in the orientation of molecules is observed at an average angle of $\theta \approx 5$ °C. It is clear that the tilt angle is quite small at the center of the film. On the other hand, comparing the data for films with 2 and 5 smectic layers (Fig. 5), it is obvious that the tilt angle is nonzero at the center of the 5-layer film at all temperatures. This is the reason why we do not observe anomalous orientation in thin films.

In an electric field, Galerne and Liebert[19] observed the orientation of SmO films floating on the free surface of droplets in the isotropic phase. Unlike the SmC^*, the SmO phase

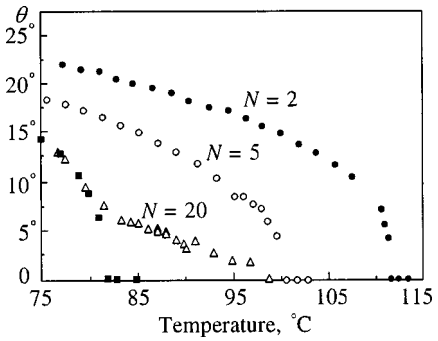


FIG. 5. Temperature dependence of the average tilt for the film thickness 2, 5, and 20 smectic layers. The filled-square symbols are x-ray data for bulk samples (8SI*).

has a herringbone arrangement of molecules. In this case, the weak dipole moment results from the reduced polarizability of the aliphatic tips of the first-layer molecules in contact with air.[19] In SmC^* films the polarization P_f could originate from the flexoelectricity.

Adapting the conclusions of Ref. 1 about the SmC^* order to our model we can distinguish ''low'' and ''high'' temperatures. At ''low'' temperatures the tilt angle is nonzero throughout the film ($dN \ll 2\xi$, net polarization perpendicular to the tilt plane). At ''high'' temperatures the tilt angle is zero at the center of the film (net polarization parallel to the **c**-director). The crossover temperature T^* is defined by $2\xi(T^*) \sim dN$. We can speculate that T_i is defined by $T^*(T_i \sim T^*)$.

In summary, we report an observation of an anomalous orientation of ferroelectric films of 8SI* and DOBAMBC in an electric field. Above the bulk transition temperature the net ferroelectric polarization in thick films is parallel to **c** and the tilt plane. The applied field orients the tilt plane parallel to the electric field. A novel method for determining the tilt angle in ferroelectric films is described.

This work was supported in part by a grant from the Russian Fund for Fundamental Research (98-02-16639), the Science and Technology program ''Statistical Physics,'' and the French National Education Ministry (Laboratoire de Dynamique et Structures des Matériaux Moleculaires, URS, CNRS No. 801).

*)E-mail: dolganov@issp.ac.ru

[1] S. Henekamp, R. A. Pelcovits, E. Fontes, E. Yi Chen, and R. Pindak, Phys. Rev. Lett. **52**, 1017 (1984).
[2] S. M. Amador and P. S. Pershan, Phys. Rev. A **41**, 4326 (1990).
[3] Ch. Bahr and D. Fliegner, Phys. Rev. A **46**, 7663 (1992).
[4] Ch. Bahr and D. Fliegner, Phys. Rev. Lett. **70**, 1842 (1993).
[5] I. Kraus, P. Pieranski, E. Demikhov, and H. Stegemeyer, Phys. Rev. E **48**, 1916 (1993).
[6] T. Stoebe and C. C. Huang, Int. J. Mod. Phys. B **9**, 2285 (1995).
[7] Ch. Bahr, C. J. Booth, D. Fliegner, and J. W. Goodby, Phys. Rev. E **52**, R4612 (1995); Phys. Rev. Lett. **77**, 1083 (1996).
[8] D. Schlauf, Ch. Bahr, and C. C. Huang, Phys. Rev. E **55**, R4885 (1997).
[9] D. R. Link, J. E. Maclennan, and N. A. Clark, Phys. Rev. Lett. **77**, 2237 (1996).
[10] P. O. Andreeva, V. K. Dolganov, and K. P. Meletov, JETP Lett. **66**, 442 (1997).
[11] P. O. Andreeva, V. K. Dolganov, C. Gors, R. Fouret, and E. I. Kats, Phys. Rev. E **59** (1999) (in press).
[12] T. Stoebe, P. Mach, and C. C. Huang, Phys. Rev. Lett. **73**, 1384 (1994).
[13] E. I. Demikhov, V. K. Dolganov, and K. P. Meletov, Phys. Rev. E **52**, R1285 (1995).
[14] V. K. Dolganov, E. I. Demikhov, R. Fouret, and C. Gors, Phys. Lett. A **220**, 242 (1996).
[15] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, New York (1964).
[16] I. Musevic, M. Skarabot, R. Blinc, W. Schranz, and P. Dolinar, Liq. Cryst. **20**, 771 (1996).
[17] P. G. de Gennes, *The Physics of Liquid Crystals*, Clarendon Press, Oxford (1994).
[18] B. I. Ostrovskii, A. Z. Rabinovich, A. S. Sonin *et al.*, JETP Lett. **25**, 70 (1977).
[19] Y. Galerne and L. Liebert, Phys. Rev. Lett. **64**, 906 (1990).

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

Spin-wave susceptibility of partially randomized ferromagnetic superlattices

V. A. Ignatchenko,*) Yu. I. Man'kov, and A. V. Pozdnyakov

L. V. Kirenskĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 25 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1335–1345 (October 1999)

This paper is a theoretical investigation of the effect of inhomogeneities in the period of a ferromagnetic superlattice on the high-frequency superlattice susceptibility. The calculations are done for a model in which the uniaxial magnetic anisotropy is taken as the physical parameter that characterizes both the ideal superlattice and a partially randomized superlattice. It is found that as the inhomogeneities become more intense, the two resonance peaks corresponding to the splitting of the spectrum at the edge of the Brillouin zone of the superlattice broaden, move closer to each other, and finally merge into one. The height of this peak increases and the peak width decreases as the intensity of the inhomogeneities increases further. The effect of inhomogeneities on the susceptibility differs dramatically in the two limits of short- and long-wave inhomogeneities: in the latter case (in contrast to the former) the dependence of the separation of the susceptibility peaks on the intensity and correlation properties of the inhomogeneities is nonmonotonic. The possibility of observing these effects in spin-wave resonance experiments involving multilayer magnetic films is also discussed. [S1063-7761(99)01510-3]

1. INTRODUCTION

The problem of propagation of waves in partially or completely randomized multilayer structures (one-dimensional superlattices) has lately received much attention. There are several approaches in developing a theory: modeling stochastization by a random arrangement of layers of two different materials;[1] computer simulation of random deviations of the surface between layers from their initial periodic arrangement;[2] and introduction of a doubly periodic dependence (with incommensurate periods) of physical parameters on the coordinate along the superlattice axis[3] (only some typical papers on this subject have been cited here, since there are many publications devoted to each approach).

One approach has been developed in our two papers, Refs. 4 and 5. A brief discussion of the results obtained for one-dimensional inhomogeneities described by a correlation function with an exponential decay of correlations can be found in Ref. 4. In Ref. 5, we systematically develop a method for one-, two-, and three-dimensional inhomogeneities of the sublattice period. Our approach differs from methods used earlier in that we do not postulate the correlation properties of the superlattice—we derive them from very general assumptions concerning the nature of stochastic spatial modulation of the sublattice period. We then find the spectrum and decay of waves by studying the averaged Green's function containing the correlation function established earlier. The theory is developed for spin, elastic, and electromagnetic waves.

During recent years extensive experimental research has been in progress in which spin-wave resonance is observed in multilayer ferromagnetic films.[6,7] As is known, the spin wavelength λ_s in thin films is determined by a size effect, which allows generating spin waves with $\lambda_s \leqslant d$, where d is the thickness of the ferromagnetic film, by using electromagnetic fields with a wavelength $\lambda \gg d$. This makes it possible to meet the conditions in which both the frequency and the wavelength of the generated wave coincide with the corresponding parameters of the edge of the superlattice Brillouin zone. It is at this edge that the spectrum of spin waves is most sensitive to inhomogeneities in the superlattice structure. Since the physical parameter observed in the experiments is the high-frequency magnetic susceptibility, it would be interesting to theoretically study this characteristic of a multilayer system. In the present paper we investigate the high-frequency susceptibility of a superlattice for a model system, whose correspondence to a real system is discussed below. We assume that the initial superlattice is a magnetic structure with a harmonic dependence of uniaxial magnetic anisotropy along the z axis, with the direction of the anisotropy axis remaining constant and parallel to the z axis. This is the simplest model for a theoretical study. It also provides the possibility of demonstrating the main features of the modification of the spin-wave spectrum in superlattices.

2. HIGH-FREQUENCY SUSCEPTIBILITY

The dynamics of a ferromagnetic system is described by the Landau–Lifshitz equation

$$\dot{\mathbf{M}} = -g(\mathbf{M} \times \mathbf{H}_{\text{eff}}), \tag{1}$$

where $\mathbf{M}$ is the magnetization, g is the gyromagnetic ratio, and $\mathbf{H}_{\text{eff}}$ is the effective magnetic field.

In the geometry corresponding to spin-wave resonance, the external magnetic field $\mathbf{H}$ is directed along the reciprocal superlattice vector $\mathbf{q}$ parallel to the z axis, i.e., perpendicular

to the magnetic film, while the high-frequency field **h** lies in the xy plane of the film. Accordingly, the components of the effective magnetic field $\mathbf{H}_{\text{eff}}$ are

$$H_{\text{eff}}^x=\alpha\frac{d^2M_x}{dz^2}+h_x,\quad H_{\text{eff}}^y=\alpha\frac{d^2M_y}{dz^2}+h_y,$$

$$H_{\text{eff}}^z=\alpha\frac{d^2M_z}{dz^2}+H_i+\beta(z)M_z, \tag{2}$$

where α is the exchange parameter, β is the magnetic anisotropy parameter, and $H_i=H-4\pi M_z$ is the internal constant magnetic field, which allows for the demagnetizing field of the film. If we take into account the symmetry of the current problem, all dynamic demagnetizing fields vanish.

We write the anisotropy parameter $\beta(z)$ in the form

$$\beta(z)=\beta[1+\gamma\rho(z)], \tag{3}$$

where β is the average value of anisotropy, γ is its rms fluctuation, and $\rho(z)$ is the centered ($\langle\rho\rangle=0$) and normalized ($\langle\rho^2\rangle=1$) function describing a superlattice with a stochastically modulated period.

Linearizing Eq. (1) in the usual manner ($M_z\approx M$ and $M_x, M_y\ll M$), we obtain an equation for the resonant circular projections $m=M_x+iM_y$ and $h=H_x+iH_y$:

$$\nabla^2 m+(\nu-\epsilon\rho(z))m=-\frac{\hbar}{\alpha}, \tag{4}$$

where we have introduced the notation

$$\nu=\frac{\omega-g(H+\beta M-4\pi M)}{\alpha g M},\quad \epsilon=\frac{\gamma\beta}{\alpha}, \tag{5}$$

with ω being the frequency of the external electromagnetic field. At frequencies used in spin-wave experiments ($\omega/2\pi\sim 10^{10}$ Hz), the wavelength ($\lambda\sim 1$ cm) is much greater than the thickness of films being studied ($d\sim 10^{-5}$ cm). Hence the amplitude h of the high-frequency field on the right-hand side of Eq. (4) may be assumed constant (time-independent). For such a field to excite standing spin waves in the film, the magnetic moment must be at least partially fixed at the surfaces of the film. We assume that this moment is completely fixed at the surfaces:

$$m(z)\big|_{z=\pm d/2}=0; \tag{6}$$

the origin of the z axis is chosen at the center of the film. Such conditions may be created by depositing additional layers of a magnetically hard alloy on both surfaces of the film.

Kittel[8] was the first to solve for the spectrum of spin waves in a thin homogeneous film (see, e.g., Gurevich and Melkov's monograph[9]), essentially by solving Eq. (4) with the boundary conditions (6) at $\epsilon=0$. The spectrum is given by the expressions

$$\nu=k_n^2,\quad k_n=\frac{\pi n}{d}, \tag{7}$$

with the field h exciting only symmetric vibrations $m\sim\cos k_n z$ corresponding to an odd number of half-waves that fit into the film thickness, i.e., $n=1,3,5,\ldots$.

To investigate Eq. (4) for $\epsilon\neq 0$, we expand m, ρ, h in the eigenfunctions of the unperturbed problem:

$$m(z)=\sum_{n=-\infty}^{+\infty} m_n\cos k_n z,\ \text{ etc.} \tag{8}$$

Then for the Fourier transforms of the corresponding functions, m_n, ρ_n, and h_n, we obtain the equation

$$(\nu-k_n^2)m_n=\epsilon\sum_{l=-\infty}^{+\infty} m_l\rho_{n-l}+\frac{h_n}{\alpha}, \tag{9}$$

where the Fourier transforms of the high-frequency field are given by the expression

$$h_n=\begin{cases}\dfrac{2h}{\pi n}\sin\dfrac{\pi n}{2}, & n\neq 0,\\ 0, & n=0.\end{cases} \tag{10}$$

We see that the equation does not contain the term in the series with $l=n$, since $\rho(z)$ is a centered function. Substituting the formal solution of Eq. (9) into the right-hand side of the same equation and averaging over the ensemble of random realizations of the function $\rho(z)$, we obtain

$$(\nu-k_n^2)\langle m_n\rangle=\epsilon^2\sum_{l\neq n}\sum_{l_1\neq l}\frac{\langle m_l\rho_{l-l_1}\rho_{n-l}\rangle}{\nu-k_l^2}+\frac{h_n}{\alpha}. \tag{11}$$

Now we decouple the correlator on the right-hand side of the equation in the approximation of the first nonvanishing perturbative term (the Bourret approximation[10]),

$$\langle m_l\rho_{l-l_1}\rho_{n-l}\rangle\approx\langle m_l\rangle\langle\rho_{l-l_1}\rho_{n-l}\rangle, \tag{12}$$

and use an identity valid for all homogeneous random functions (see, e.g., Ref. 11),

$$\langle\rho_{l_1}\rho_{l_2}\rangle=\langle|\rho_{l_1}|^2\rangle\delta_{l_1 l_2}. \tag{13}$$

This leads us to a solution of the form

$$\langle m_m\rangle=\frac{h_n}{\alpha}\left\{\nu-k_n^2-\epsilon^2\sum_{l\neq n}\frac{\langle|\rho_{n-l}|^2\rangle}{\nu-k_l^2}\right\}^{-1}. \tag{14}$$

Experimenters measure the response, averaged over the film volume, of the system to the high-frequency field:

$$\overline{m}=\frac{1}{d}\int_{-d/2}^{d/2}m(z)dz=\sum_{n=-\infty}^{+\infty}{}'\left(\frac{2}{k_n d}\sin\frac{\pi n}{2}\right)m_n, \tag{15}$$

where the prime on the sum indicates the absence of the term with $n=0$.

Thus, the average susceptibility observed in spin-wave experiments is the sum of partial susceptibilities,

$$\chi=\frac{\langle\overline{m}\rangle}{h}=\sum_{n=-\infty}^{+\infty}{}'\chi_n, \tag{16}$$

where

$$\chi_n=\left(\frac{2}{\pi n}\sin\frac{\pi n}{2}\right)^2\frac{1}{\alpha}\left\{\nu-k_n^2-\epsilon^2\sum_{l\neq n}\frac{\langle|\rho_{n-l}^2|\rangle}{\nu-k_l^2}\right\}^{-1}. \tag{17}$$

The study of an expression containing a sum over discrete k_n poses a serious problem. We therefore limit ourselves to the study of the continuous analog of this expression (replacing summation by integration):

$$\chi(\nu,k)=a(k)\left\{\nu-k^2-\epsilon^2\int\frac{S(k-k_1)dk_1}{\nu-k_1^2}\right\}^{-1}, \qquad (18)$$

where $a(k)=\alpha^{-1}(2/kd)^2\sin^2(kd/2)$. Here we must bear in mind that this expression describes the frequency dependence of χ only near the discrete values of the wave number, $k=k_n$. The function $S(k)$ is the spectral density of the random function $\rho(z)$ and is related to the superlattice correlation function $K(r)$ via the Fourier transform (the Wiener–Khinchin theorem)

$$K(r)\equiv\langle\rho(z)\rho(z-r)\rangle=\int S(k)\exp\{ikr\}dk. \qquad (19)$$

So far we have not made any assumptions concerning the function $\rho(z)$, except that it is a centered and normalized homogeneous random function. In accordance with Ref. 5, we write

$$\rho(z)=\sqrt{2}\cos[q(z-u(z))], \qquad (20)$$

where $q=|\mathbf{q}|$ is the wave number of the initial superlattice, and $u(z)$ is a random function describing the inhomogeneity of the period of this superlattice.

Thus, we will examine a model in which the physical parameter characterizing the superlattice has, in this initial state (at $u(z)\equiv 0$), a harmonic dependence on z. A method for finding the correlation function of such a superlattice was developed in Ref. 5. It amounted to a generalization to the case of partially randomized superlattices of the theory of stochastic frequency (or phase) modulation of a periodic radio signal, well-known in radiophysics.[11,12] The correlation properties of the superlattice are expressed in this method in terms of the stochastic characteristics of the function $u(z)$ (more precisely, of the derivative du/dz). Here the shape of the correlation function $K(r)$ of the superlattice is independent, in the limits of long- and short-wave inhomogeneities, of the shape of the correlation functions modulating the initial inhomogeneities du/dz. At the same time, the shape of $K(r)$ is extremely sensitive to the correlation length of the inhomogeneities. The correlation function obtained in Ref. 5 for the two limits, long- and short-wave inhomogeneities, corresponding to a random shift in the boundaries separating the layers, has the form

$$K(r)=\cos(qr)\begin{cases}\exp\left\{-\dfrac{k_{c1}^2r^2}{2}\right\}, & p_0\ll 1,\\ \exp\{-k_{c2}r\}, & p_0\gg 1,\end{cases} \qquad (21)$$

where $p_0=k_\parallel/\sigma q$, and $k_{c1}=\sigma q$ and $k_{c2}=(\sigma q)^2/k_\parallel$ are the effective correlation wave numbers of the superlattice, with σ and $k_\parallel$ the rms fluctuation and the correlation wave number of the random function du/dz. Thus, irrespective of the shape of the correlation function modeling the properties of the initial inhomogeneities du/dz, the correlation function of the superlattice has a Gaussian decay of correlations for long-wave inhomogeneities and an exponential decay for short-wave inhomogeneities.

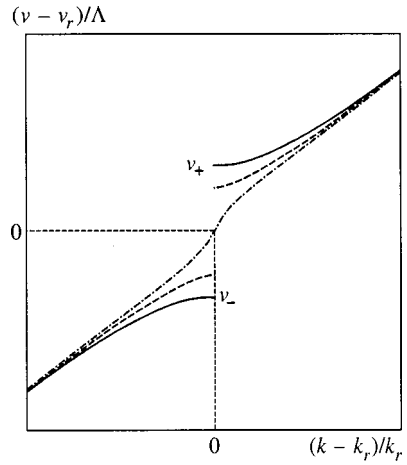


FIG. 1. Dispersion relation for the superlattice near the edge of the Brillouin zone (for more details see the main body of the text).

Using the Fourier transform to express the spectral density $S(k)$ corresponding to the correlation function (21) and substituting the result into (18), we obtain a formula for the superlattice susceptibility:

$$\chi=\frac{a(k)}{\nu-k^2-(\Lambda^2/4)F(L_-+L_+)}, \qquad (22)$$

where $\Lambda=\epsilon\sqrt{2}$, and the functions F and $L_\pm$ are determined for each limit, $p_0\gg 1$ and $p_0\ll 1$, by different expressions. We begin with the limit $p_0\gg 1$, corresponding to short-wave inhomogeneities. Then

$$F=1-\frac{ik_{c2}}{\sqrt{\nu}}, \quad L_\pm=\frac{1}{(\sqrt{\nu}-ik_{c2})^2-(k\pm q)^2}, \qquad (23)$$

By equating the denominator in (22) to zero we obtain the dispersion equation for averaged spin waves in the superlattice (this equation has been studied in Refs. 4 and 5). The qualitative behavior of the results is schematically depicted in Fig. 1. At the edge of the Brillouin zone, corresponding to $k=k_r\equiv q/2$, the spectrum of the initial ideal ($k_{c2}=0$) superlattice exhibits a gap, $\Delta\nu\equiv\nu_+-\nu_-=\Lambda$ (the solid curve in Fig. 1). As k_{c2} increases, the gap decreases (the dashed curve) according to

$$\Delta\nu=\sqrt{\Lambda^2-G_2^2}, \qquad (24)$$

where $G_2=qk_{c2}=\sigma^2q^3/k_\parallel$ is the parameter characterizing the decay due to inhomogeneities. When $G_2>\Lambda$, the dispersion relation for the averaged waves is continuous and has a point of inflection at $k=k_r$ (the dot–dash curve in Fig. 1).

We study the dependence of the susceptibility (22) on the frequency ν at the edge of the Brillouin zone ($k=k_r$). Here we can limit ourselves to the two-wave approximation and discard the nonresonant term L_+. The susceptibility becomes

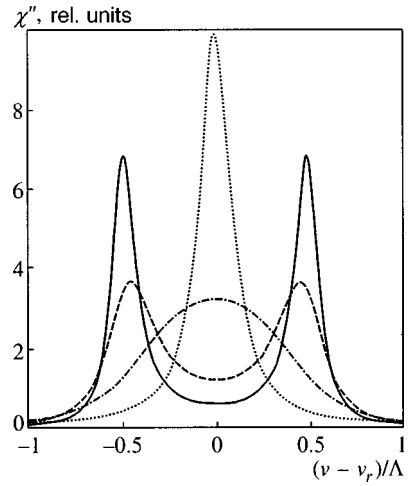


FIG. 2. The imaginary part of susceptibility, χ'', for the case of short-wave inhomogeneities at $G_2/\Lambda=0.15$ (solid curve), 0.3 (dashed curve), 0.8 (dot–dash curve), and 2.5 (dotted curve).

$$\chi=\frac{a(k_r)(\nu-\nu_r-k_{c2}^2-2ik_{c2}\sqrt{\nu})}{(\nu-\nu_r)(\nu-\nu_r-k_{c2}^2-2ik_{c2}\sqrt{\nu})-(\Lambda^2/4)(1-ik_{c2}/\sqrt{\nu})}. \tag{25}$$

The frequency dependence of the imaginary part of the susceptibility, χ'', is depicted in Fig. 2 for four values of the ratio G_2/Λ. Clearly, for small values of G_2 there are two narrow peaks (the solid curve). As G_2 increases (the dashed curve), the height of the peaks decreases, the widths increase, and the peaks move closer to each other, and at a certain value of G_2 they merge into one broad resonance peak (the dot–dash curve). A further increase in G_2 leads to a decrease in the peak width as the peak becomes higher (dotted curve). Figure 3 depicts the decrease in the separation of the resonance peaks, $\Delta\nu_m$, with increasing G_2 (the solid curve). For the sake of comparison, we also give the dependence of the gap width $\Delta\nu$ in the spectrum of the average waves corresponding to (24) (the dashed curve). Clearly, the separation

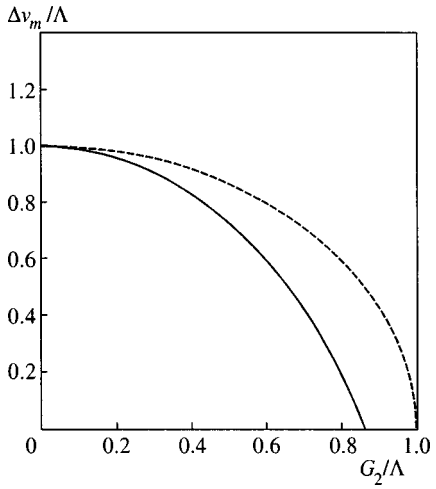


FIG. 3. Separation of the resonance peaks in the susceptibility, $\Delta\nu_m$ (solid curve), and the gap in the spectrum, $\Delta\nu$ (dashed curve) in the presence of short-wave inhomogeneities.

of the peaks, $\Delta\nu_m$, is always less than $\Delta\nu$ and the two peaks merge into one at values of G_2 less than $G_2=\Lambda$, which corresponds to the collapse of the gap in the spectrum. Using these diagrams, we can determine the gap width in the spectrum, $\Delta\nu$, by observing the separation of the resonance peaks, $\Delta\nu_m$.

The imaginary part of the susceptibility can be represented by the sum of two resonances (if we neglect k_{c2}^2 in (25) and allow for the smallness of k_{c2}/k_r):

$$\chi''=\frac{\Gamma_2 a(k_r)}{\Delta\nu}\left[\frac{\Delta\nu+\nu_r-\nu}{(\nu-\nu_r-\Delta\nu/2)^2+\Gamma_2^2}+\frac{\Delta\nu-\nu_r+\nu}{(\nu-\nu_r+\Delta\nu/2)^2+\Gamma_2^2}\right], \tag{26}$$

where $\Gamma_2=G_2/2$ is the width of the resonance peaks, and $\Delta\nu$ is determined by (24).

The separation of these resonance peaks is given by the formula

$$\Delta\nu_m=2\sqrt{\Lambda^2-G_2^2}-\Lambda. \tag{27}$$

The expressions (26) and (27) provide a good approximation to the exact curves in Figs. 2 and 3.

If $G_2\gg\Lambda$, i.e., when there is only one well-resolved central peak, χ'' can be expressed as

$$\chi''=\frac{\Gamma_2' a(k_r)}{(\nu-\nu_r)^2+\Gamma_2'^2}, \tag{28}$$

where the parameter $\Gamma_2'=\Lambda^2/4G_2$ acts as the effective decay constant. The value of this decay decreases with increasing G_2, with the result that the height of the resonance peak increases (Fig. 2).

Now we turn to long-wave one-dimensional inhomogeneities, corresponding to the Gaussian decay of correlations [the upper line in Eq. (21)]. Here

$$F=(2k_{c1}^2\nu)^{-1/2},$$

$$L_\pm=D(u_\pm)+D(v_\pm)+i\frac{\sqrt{\pi}}{2}(\exp\{-u_\pm^2\}+\exp\{-v_\pm^2\}), \tag{29}$$

where $D(x)=\exp(-x^2)\int_0^x\exp(t^2)dt$ is Dawson's integral, and

$$u_\pm=\frac{\sqrt{\nu}-|k\pm q|}{k_{c1}\sqrt{2}},\quad v_\pm=\frac{\sqrt{\nu}+|k\pm q|}{k_{c1}\sqrt{2}}. \tag{30}$$

If we examine the susceptibility near the right-hand boundary of the Brillouin zone, we can neglect the term L_+ in the two-wave approximation, with the result that the susceptibility becomes (from now on we drop the subscript "$-$" in u_- and v_-)

$$\chi=\frac{a(k_r)}{\nu-k^2-(\Lambda^2/4k_{c1}\sqrt{2\nu})[D(u)+D(v)+i(\sqrt{\pi}/2)(\exp\{u^2\}+\exp\{v^2\})]}. \tag{31}$$

The frequency dependence of the imaginary part of this expression (i.e., χ'') is depicted in Fig. 4 for four values of the parameter $G_1=qk_{c1}=\sigma q^2$, which characterizes the decay, due to long-wave inhomogeneities. Clearly, in addition to the features common to both short-wave (Fig. 2) and long-wave (Fig. 4) inhomogeneities, there is an important difference between these two limits. Quantitatively, this difference manifests itself in the height and width of the resonance peaks for the same values of G_1/Λ and G_2/Λ for long- and short-wave inhomogeneities. Qualitatively, the difference amounts to the fact that for long-wave inhomogeneities, an increase in G_1 first leads to an increase in the separation of the peaks, and only after that do the peaks move closer to each other and merge into a single central resonance. A similar effect was obtained in Ref. 5 for the gap in the spectrum in the presence of long-wave inhomogeneities. Figure 5 shows that for the case of long-wave inhomogeneities, the peaks move closer to each other and finally merge at values of G_1 at which the gap $\Delta\nu$ is still only weakly modified by the inhomogeneities.

We studied the expression (31) numerically, but in the two limits the susceptibility is given by simple formulas. In particular, when G_1 is small, we have $u\gg 1$, and for Dawson's integral we have the simple expression

$$D(u)\approx\frac{1}{2u}\left(1+\frac{1}{2u^2}\right). \tag{32}$$

Ignoring the nonresonant terms containing v, we obtain a formula for the imaginary part of the susceptibility as a sum of two susceptibilities:

$$\chi''=\frac{\Gamma_1 a(k_r)}{\Delta\nu}\left[\frac{\nu-\nu_r}{(\nu-\nu_r-\Delta\nu/2)^2+\Gamma_1^2}+\frac{\nu_r-\nu}{(\nu-\nu_r+\Delta\nu/2)^2+\Gamma_1^2}\right], \tag{33}$$

where the gap $\Delta\nu$ and the width Γ_1 of the resonance peaks are now given by the formulas

$$\Delta\nu=\sqrt{\Lambda^2+4G_1^2},\qquad \Gamma_1=\left(\frac{\pi}{2}\right)^{1/2}\frac{\Lambda^2-4G_1^2}{8G_1}\exp\left\{-\frac{\Lambda^2+4G_1^2}{8G_1^2}\right\}. \tag{34}$$

An increase in G_1 moves the resonance peaks apart, so that Eq. (33) corresponds to the initial section of the curve for $\Delta\nu_m$ in Fig. 5.

When the decay is large, Dawson's integral can be neglected in the denominator of (31). Then only an imaginary quantity is left as a factor of Λ^2, and the imaginary part of the susceptibility is given by an expression similar to (28) in which $\Gamma_1'=\sqrt{\pi/2}\Lambda^2/4G_1$ replaces Γ_2'.

3. DISCUSSION

What are the limitations that the simplifying assumptions used in our calculations impose on the comparison of the results obtained in this paper with the data of spin-wave experiments in thin films with a multilayer structure?

The calculations of the effect of inhomogeneities on the high-frequency magnetic susceptibility of the superlattice

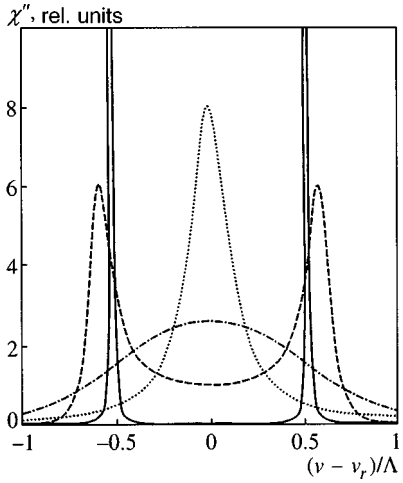


FIG. 4. The imaginary part of the susceptibility, χ'' (in relative units), for the case of long-wave inhomogeneities at $G_1/\Lambda=0.15$ (solid curve), 0.3 (dashed curve), 0.8 (dot–dash curve), and 2.5 (dotted curve).

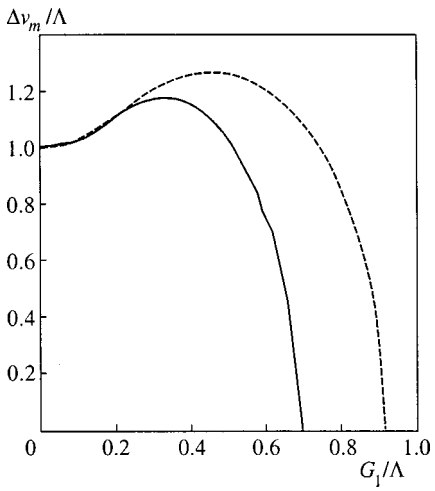


FIG. 5. Separation of the resonance peaks in the susceptibility, $\Delta\nu_m$ (solid curve), and the gap in the spectrum, $\Delta\nu$ (dashed curve) in the presence of long-wave inhomogeneities.

were done here for a model in which uniaxial magnetic anisotropy acts as the physical parameter characterizing both the ideal superlattice and a partially randomized superlattice. In real superlattices the magnetization may also be such a parameter, as well as the exchange parameter and the orientation of the anisotropy axis. The calculations are different for each of these inhomogeneity parameters. However, the comparison of the results obtained in calculations of the effect on the dispersion law of the averaged spin waves in a superlattice with anisotropy inhomogeneities and exchange inhomogeneities[4] show that the modifications of the spectrum at the edge of the Brillouin zone do not differ too much in these two cases. Hence we can expect that the differences will not be too large for the high-frequency susceptibility for the inhomogeneities of the various physical parameters either, with the result that basically only Λ will need to be redefined.

In spin-wave experiments only discrete values of the wave vector, $k_n=\pi n/d$, $n=1,3,5$, can be observed. Consequently, the dispersion relation of the waves, depicted schematically in Fig. 1, can be obtained only at distinct, fairly distant, points, and generally among these there is not a single one that coincides with the edge of the Brillouin zone $k_r=q/2$. Note that our investigation of the frequency dependence of χ'' in this paper was done for $k=k_r$, so a comparison of the results of calculations is valid only with the data of an experiment in which coincidence of k_r with one of the wave numbers from the set $\{k_n\}$ is achieved only by special selection of the film thickness d and the superlattice period $l=2\pi/q$.

In the present paper, as in Ref. 5, we used a model in which the physical parameter characterizing the superlattice varies along the z axis of the initial ideal superlattice by the harmonic law $\rho(z)=\sqrt{2}\cos qz$. This corresponds to the limit of smooth boundaries between the layers of the superlattice, with the thickness of the ''boundary'' being equal to the thickness of the ''layer.'' On the other hand, experiments are done with multilayer structures, in which the boundary-to-layer thickness ratio is usually much less than unity. In this case the function $\rho(z)$ for the initial superlattice is much closer in shape to a train of rectangular pulses of different polarity than to a harmonic function. A theoretical study of the modification of the dispersion law and decay due to inhomogeneities has been carried out for all odd Brillouin zones of a superlattice in Refs. 13–15. It was found that the results obtained for this model differ substantially from those obtained for the model of Ref. 5 with a harmonic ρ vs. z dependence for all Brillouin zones except the first. For the first Brillouin zone the modification is determined by the first harmonic $\rho\sim\cos qz$ in the Fourier expansion of $\rho(z)$, so that the results for the two limiting models differ only by the numerical normalization factor $8/\pi^2$. Since the experimental investigations of spin-wave resonance and the theoretical calculations of χ'' in the present paper were done for the first Brillouin zone, the discrepancy in the models considered here should not have a strong influence on the precision of the comparison of theory and experiment.

Our calculations were done on the assumption that the intrinsic decay of the spin system is much weaker than the decay due to the inhomogeneities in the sublattice period. Only in this situation can the effects described in this paper manifest themselves, and by comparing the theoretical results and the experimental data one can determine the real gap in the spectrum of spin waves (using the diagrams in Figs. 3 and 5) and measure the parameter that determine the correlation properties of the inhomogeneities.

The authors are grateful to A. A. Maradudin for productive discussions. This work was sponsored by the NATO Program and Cooperation Partner Linkage (Grant No. 974573), NATO Networking Supplement (Grant No. 971203), Network Infrastructure (Grant No. 973201), and Krasnoyarsk Regional Science Fund (Project No. 7F0176).

*)E-mail: ignatchenko@theorphys.krascience.rssi.ru

[1] N. Nishiguchi, S. Tamura, and F. Nori, Phys. Rev. B **48**, 2515 (1993).
[2] A. R. McGurn, K. T. Christensen, F. M. Mueller, and A. A. Maradudin, Phys. Rev. B **47**, 13 120 (1993).
[3] S. J. Blandell, J. Phys.: Condens. Matter **6**, 10 283 (1994).
[4] V. A. Ignatchenko, R. S. Iskhakov, and Yu. I. Mankov, J. Magn. Magn. Mater. **140–144**, Part III, 1947 (1995).
[5] V. A. Ignatchenko and Yu. I. Mankov, Phys. Rev. B **56**, 194 (1997).
[6] R. S. Iskhakov, A. S. Chekanov, and L. A. Chekanova, Fiz. Tverd. Tela (Leningrad) **32**, 441 (1990) [Sov. Phys. Solid State **32**, 255 (1990)].
[7] R. S. Iskhakov, I. V. Gavrishin, and L. A. Chekanova, JETP Lett. **63**, 989 (1996).
[8] C. Kittel, Phys. Rev. **110**, 1295 (1958).
[9] A. G. Gurevich and G. A. Melkov, *Magnetization Oscillations and Waves*, CRC Press, Boca Raton (1996).
[10] R. Bourret, Physica (Utrecht) **54**, 623 (1971).
[11] S. M. Rytov, *An Introduction to Statistical Radiophysics* [in Russian], Nauka, Moscow (1994).
[12] A. N. Malakhov, Zh. Eksp. Teor. Fiz. **30**, 884 (1956) [Sov. Phys. JETP **3**, 653 (1957)].
[13] V. A. Ignatchenko, Yu. I. Mankov, and A. A. Maradudin, in *Frontiers in Magnetism of Reduced Dimension Systems*, V. G. Bar'yakhtar and P. E. Wigen (eds.), Kluwer Academic, Dordrecht (1998), p. 217.
[14] V. A. Ignatchenko, Yu. I. Mankov, and A. A. Maradudin, Phys. Rev. B **59**, 42 (1999).
[15] V. A. Ignatchenko, Yu. I. Mankov, and A. A. Maradudin, J. Phys.: Condens. Matter **11**, 2773 (1999).

Translated by Eugene Yankovsky

Spontaneous and field-induced magnetic phase transitions in the intermetallic compounds $(Gd_{1-x}Y_x)\ Mn_2Ge_2$

A. Yu. Sokolov

Moscow Institute of Radio Engineering and Automation, 117454 Moscow, Russia

Guo Guanghua, S. A. Granovskiĭ, and R. Z. Levitin*)

M. V. Lomonosov Moscow State University, 119890 Moscow, Russia

H. Wada and M. Shiga

Kyoto University, Kyoto 606-01, Japan

T. Goto

Institute of Solid State Physics, Tokyo University, Tokyo 106, Japan

(Submitted 23 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1346–1364 (October 1999)

The magnetic properties of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ $(0 \leqslant x \leqslant 0.5)$ (tetragonal crystal structure of the type $ThCr_2Ge_2$) have been measured in fields up to 40 T in the temperature range 4.2–120 K on free powders (whose particles are free to rotate in a magnetic field), fixed powders, and polycrystalline bulk samples. The temperature dependence of the crystal structure parameters of some of these intermetallic compounds was determined from x-ray diffraction measurements. It was found that the temperatures of magnetic disordering of the gadolinium magnetic subsystem and the transition of the manganese subsystem from the ferromagnetic to the antiferromagnetic state decrease as the gadolinium content is decreased, so that in compositions with $x<0.3$ this transition is a first-order phase transition, and with further decrease of the gadolinium content it becomes second-order. In intermetallic compounds of this system with $x=0.3$ and 0.4 another spontaneous first-order transition is observed at lower temperatures. In a magnetic field the magnetization of all the investigated intermetallic compounds except for the composition $x=0.5$ undergo a metamagnetic transition. The $T-x$, $H-x$, and $H-T$ magnetic phase diagrams are constructed. The experimental data are interpreted with allowance for the fact that according to the results of previous studies, the Mn–Mn exchange interaction is antiferromagnetic, and thus the examined intermetallic compounds are two-sublattice ferrimagnets with negative exchange interaction in one of the sublattices. Calculations performed within the framework of this model in the Yafet–Kittel approximation, which assumes that the sublattice with negative exchange can be divided into two sub-sublattices, permit a quantitative description of the experimental results in most cases.
 [S1063-7761(99)01610-8]

1. INTRODUCTION

The ternary intermetallic compounds RMn_2Ge_2 (R is a rare earth or yttrium) have tetragonal crystal structure of the type $ThCr_2Si_2$ (space group $I4/mmm$). This structure consists of alternating layers R–Ge–Mn–Ge–... perpendicular to the tetragonal axis. The magnetic properties of these compounds are determined by the two magnetic subsystems formed by atoms of the rare earth and manganese. It follows from numerous studies (see, for example, the review in Ref. 1) that the largest interaction is the intraplanar Mn–Mn ferromagnetic exchange interaction, which is the main factor determining the magnetic ordering temperature of the manganese subsystem (350–450 K). The exchange interaction between the manganese atoms in neighboring planes is an order of magnitude weaker. The interplanar rare earth–manganese exchange is of roughly the same magnitude. The rare earth–rare earth interaction is another order of magnitude smaller.

It also follows from the experimental data that the interplanar Mn–Mn exchange interaction constant depends strongly on the interatomic distances (mainly on the distance between the manganese atoms in a layer) and changes sign from positive to negative as the crystal structure parameter a decreases below some critical value $a_c \approx 4.045$ Å. Therefore, the natural magnetic ordering of the manganese subsystem is ferromagnetic in the majority of intermetallic compounds of this system with light rare earths, for which $a>a_c$, and antiferromagnetic in intermetallic compounds of this system with heavy rare earths, since $a<a_c$ for the latter.

The foregoing features of the exchange interactions in the intermetallic compounds RMn_2Ge_2 lead to the result that

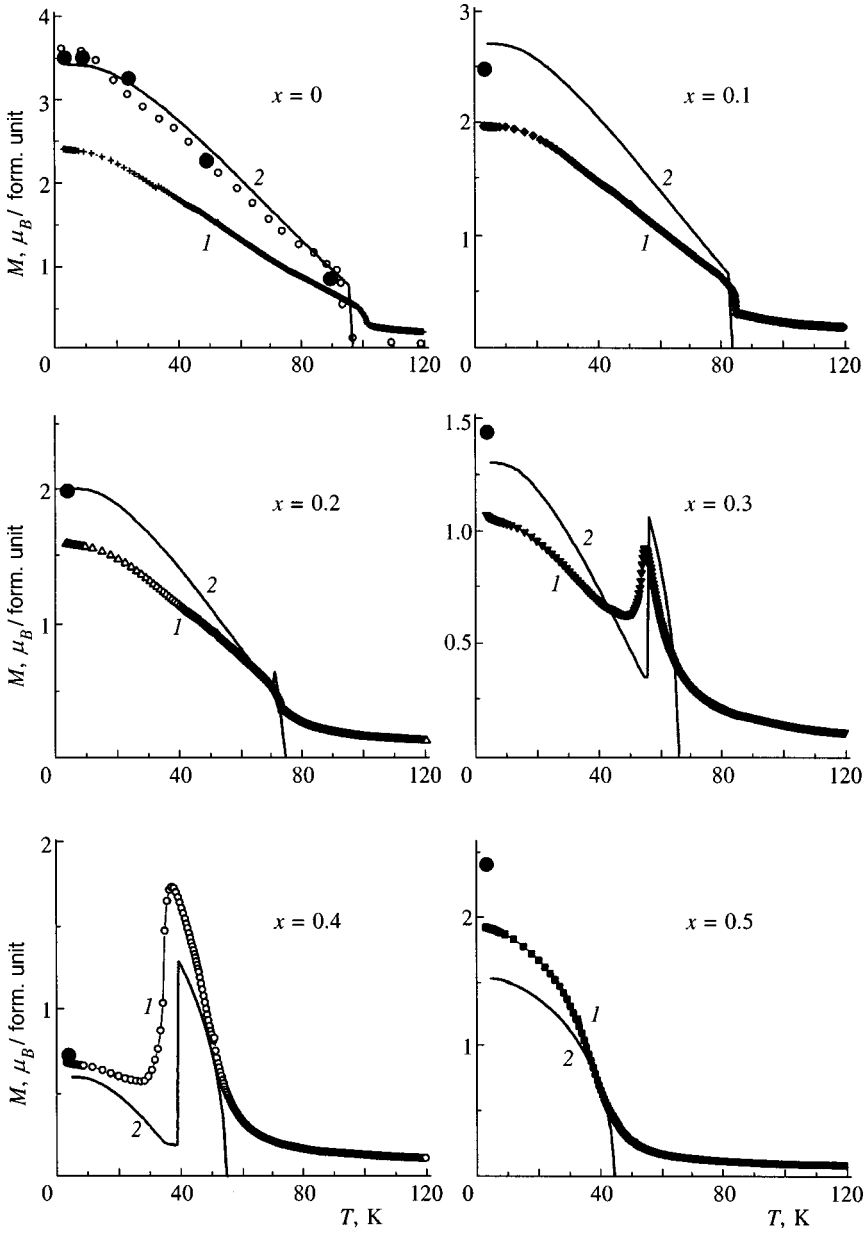


FIG. 1. Temperature dependence of the magnetization of the substituted compounds $Gd_{1-x}Y_xMn_2Ge_2$: *1*—experimental data in a field of 0.83 T; *2*—theoretical dependence of the spontaneous magnetization for a free powder. Values of the spontaneous magnetization (large filled circles) are also shown, derived from measurements in strong magnetic fields (see text and Fig. 4). For the composition with $x=0$, values of the spontaneous magnetization are shown (empty circles) for the single crystal from Ref. 8.

different types of spontaneous magnetic phase transitions arise in the magnetically ordered region.

In $SmMn_2Ge_2$ and in some mixed intermetallics a transition of the manganese subsystem from the ferromagnetic to the antiferromagnetic state is observed when the temperature is lowered. This transition is due to the change in sign of the interplanar Mn–Mn exchange, which in turn is due to thermal expansion.[1,2]

With the emergence of magnetic ordering in the rare-earth subsystem, the manganese subsystem changes from the antiferromagnetic state to the ferromagnetic state, due to the rare earth–manganese exchange interaction. Such transitions have been detected in intermetallics of Gd, Dy, Tb, and a number of mixed compounds.[1,3] Since the rare earth–manganese exchange interaction with heavy rare earths is antiferromagnetic, these compounds are two-sublattice ferrimagnets in their ground state.

Recently, in intermetallic compounds of light rare earths whose manganese subsystem is ferromagnetic at room temperature and lower temperatures, a transition to the antiferromagnetic state was observed when the temperature increased above room temperature.[2,4] The nature of this transition is as yet unclear.

Phase transitions in the intermetallics RMn_2Ge_2 induced by a magnetic field have received considerably less attention. In Ref. 5 it was shown that in a single crystal of an intermetallic compound of gadolinium at temperatures below the magnetic ordering point of the gadolinium subsystem (≈ 95 K) a first-order phase transition takes place in a magnetic field, induced by the field. A similar result was obtained recently on polycrystalline samples in Ref. 6 in which some of the authors of the present paper participated. The field-induced magnetic transition was interpreted as being due to a breakdown of ferromagnetic ordering in the manganese subsystem; however, no quantitative analysis was given.

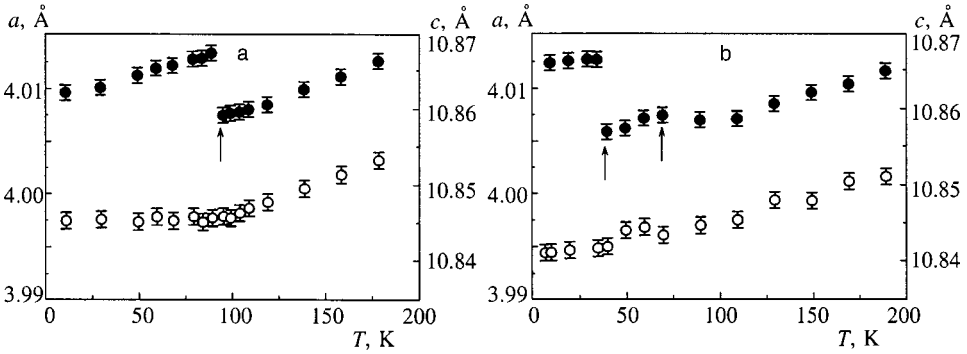


FIG. 2. Temperature dependence of the parameters a (dark circles) and c (empty circles) of the crystal structure of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ for $x=0$ (a) and 0.4 (b). The arrows indicate the temperatures of spontaneous magnetic phase transitions, determined from the magnetic measurements.

In light of this, we have undertaken a study of low-temperature spontaneous and field-induced magnetic phase transitions in the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$. Substitution of gadolinium by nonmagnetic yttrium changes the magnetic moment of the gadolinium subsystem and, accordingly, the energy of the Gd–Mn exchange interaction, which allows one to delineate the dependence of the magnetic behavior on these parameters. At the same time, in this system to first order it is possible to neglect the dependence of the Mn–Mn exchange on the interatomic distances since the crystal structure parameters of gadolinium and yttrium intermetallics are similar. We have also attempted a theoretical interpretation of our experimental results.

The paper is organized as follows. After describing the samples and experimental techniques, we present the experimental data for free powders, whose particles are free to rotate in a magnetic field. Next we lay out the theoretical model that we have used to analyze the experimental results. By comparing the experimental data for free powders with our theoretical calculations we determine the parameters describing the magnetic behavior of the investigated intermetallic compounds and construct theoretical magnetic phase diagrams and the temperature and field dependence of the magnetization for free powders and compare them with the experimental results. Then, within the context of the theoretical model, we analyze the results of our experimental studies of fixed powders and polycrystalline samples, as well as the available experimental data in the literature for single-crystal $GdMn_2Ge_2$. In the Conclusion we present a critical analysis of the theoretical model and its applicability to the magnetic behavior of the investigated system.

2. SAMPLES AND MEASUREMENT TECHNIQUES

Polycrystalline samples of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ ($0 \leqslant x \leqslant 0.5$) were melted together in a cold-hearth arc furnace in an atmosphere of spectrally pure argon from the initial ingredients. The samples were annealed for one week at 750 °C in vacuum. The single-phase character of the samples was monitored by x-ray diffraction measurements.

All magnetic measurements were made in the temperature range 4.2–120 K. The magnetization was measured in static fields up to 0.83 T on a magnetometer, and also by the induction method on fixed and free powders in pulsed fields up to 40 T with a pulse duration of 30 ms. The measurements on powders whose particles, having diameters of 10 $-30\ \mu$m, were free to rotate in a magnetic field followed the technique developed in Ref. 7. We also measured the magnetization of some bulk samples in pulsed fields up to 25 T with a pulse duration of 10 ms. The temperature dependence of the crystal structure parameters for compositions $x=0$ and $x=0.4$ in the temperature range 10–300 K.

3. EXPERIMENTAL RESULTS

Figure 1 shows the temperature dependence of the magnetization of intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ for various values of x measured in a static field of 0.83 T. Although the measurements were performed on free powders whose particles could reorient in the applied field, under these conditions total reorientation of the particles was not achieved. This is indicated by the fact that the magnetization in a field of 0.83 T is significantly less than the saturation magnetization.

The investigated compounds can be divided into three groups according to the temperature dependence of the magnetization. The magnetization of compositions with low yttrium content ($x=0$, 0.1, 0.2) decreases monotonically as the temperature is raised and fall abruptly when a certain temperature is reached. Measurements of the temperature dependence of the lattice parameter of $GdMn_2Ge_2$ show that a discontinuous change in the lattice parameter a occurs at this temperature, i.e., a first-order phase transition is observed (Fig. 2a).

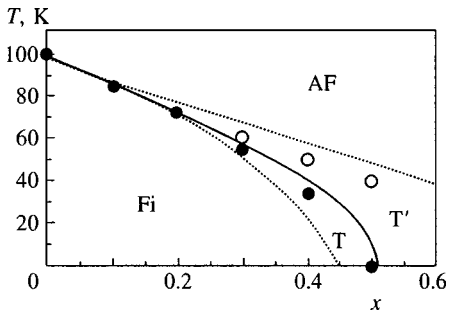


FIG. 3. Magnetic $T-x$ phase diagram of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$. The points are experimental data for second-order (○) and first-order (●) phase transitions. The calculated phase diagrams are represented by dashed (second-order phase transitions) and solid (first-order phase transition) lines. See Fig. 8 for notation of the phases.

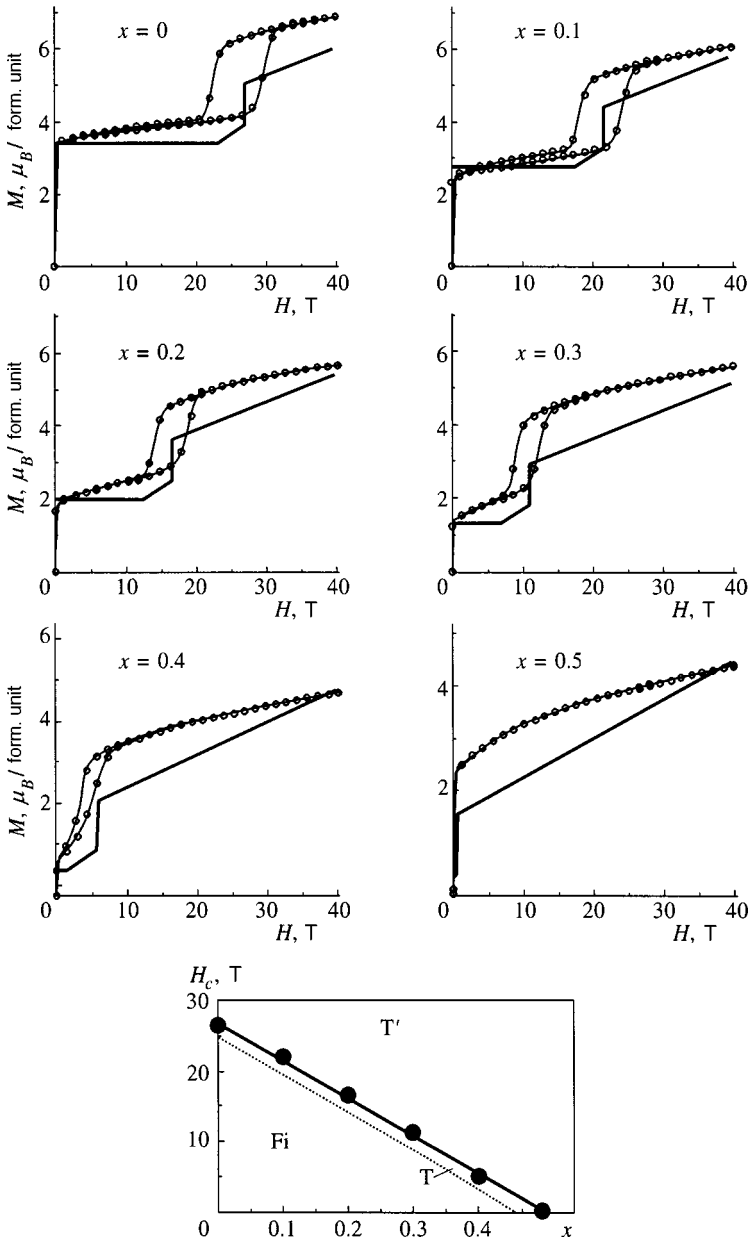


FIG. 4. Field dependence of the magnetization of free powders of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ in fields up to 40 T at 4.2 K. The dots correspond to experimental points, and the lines, to the calculated field dependence of the magnetization. The bottom graph plots the experimental values of the $H-x$ phase diagram at 4.2 K: the points plot experimental values of the H_c fields of the metamagnetic transitions; the lines plot the calculated dependence for the T→T′ (solid) and Fi→T (dotted) transitions.

The magnetization of compositions with $x=0.3$ and 0.4 behaves differently. It varies weakly at low temperatures, and when a certain temperature is reached grows abruptly. The jump in the parameter a at this temperature (see Fig. 2b) indicates that growth of the magnetization is due to a first-order phase transition. Further increase in the temperature leads to a monotonic decrease of the magnetization up to some temperature, above which the magnetization is small and varies only weakly, i.e., a second-order phase transition takes place at this temperature. Note that at this transition a small anomaly is also observed in the temperature dependence of the lattice parameter $a(T)$ (Fig. 2c). Finally, the magnetization for $x=0.5$ decreases monotonically as the temperature is increased and undergoes a second-order phase transition at 40 K.

The $T-x$ magnetic phase diagram of the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ is shown in Fig. 3. The points corresponding to magnetic phase transitions were determined from the maxima of the first derivative of the magnetization with respect to the temperature in a field of 0.83 T.

The field dependence of the magnetization of the investigated intermetallics at 4.2 K in fields up to 40 T, recorded on free powders, is shown in Fig. 4. It can be seen that the magnetization for compositions $x<0.5$ undergoes a metamagnetic transition in a magnetic field, accompanied by significant hysteresis. The field corresponding to this transition, H_c, defined as the mean field of the magnetization jump as the external field is increased and decreased, decreases linearly with increase of the yttrium content x and, as can be seen by extrapolation, is equal to zero for $x\approx 0.5$ (see the bottom graph). The spontaneous magnetization M_s at 4.2 K, determined from the data plotted in Fig. 4 by linear extrapolation of the magnetization curves to zero field from the field region $H<H_c$, decreases linearly as x increases to $x=0.4$ and grows abruptly as the composition $x=0.5$ is approached (Fig. 5). This figure also plots the spontaneous magnetization

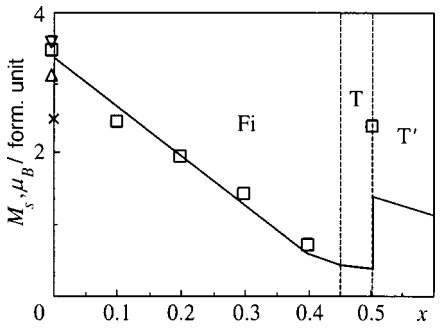


FIG. 5. Dependence of the spontaneous magnetization of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ on the yttrium concentration at 4.2 K: □—measurements for free powders; ×—measurements on fixed powders and bulk polycrystalline samples; Δ and ∇—data for a single crystal from Refs. 5 and 8, respectively. The solid line plots the calculated dependence $M_s(x)$.

that we measured for fixed powders and bulk polycrystalline samples of $GdMn_2Ge_2$, and values of the spontaneous magnetization of single crystals of this compound taken from Refs. 5 and 8. It can be seen that the value of M_s for the free powder is close to the values for the single crystal, while its value for the fixed powder and the bulk polycrystalline sample is significantly lower. This indicates that under the conditions of our experiments the particles of such a powder are ''free,'' and able to reorient in an applied field so as to minimize their magnetic energy.

As follows from Fig. 6, which plots magnetization curves of $GdMn_2Ge_2$ at different temperatures, the field corresponding to the metamagnetic transition, H_c, decreases as the temperature is raised and tends to zero at the temperature of the spontaneous first-order phase transition for this compound. The critical fields of the metamagnetic transition in the intermetallic compound of this system with $x=0.2$ depend on the temperature in a similar way. Figure 7 plots the temperature dependence of H_c for intermetallics of this system with the compositions $x=0$ and 0.2.

4. MODEL

As was noted in the Introduction, intermetallics RMn_2Ge_2 with magnetic rare earths have two magnetic subsystems: the rare-earth subsystem and the manganese subsystem. If R is a heavy rare earth, then the exchange interaction between the manganese and the rare-earth subsystems is negative and at low temperatures they are two-sublattice ferrimagnets. It is necessary, however, to take into account that in intermetallics with heavy rare earths and yttrium the interplanar Mn–Mn exchange is also negative. Thus, the fer-

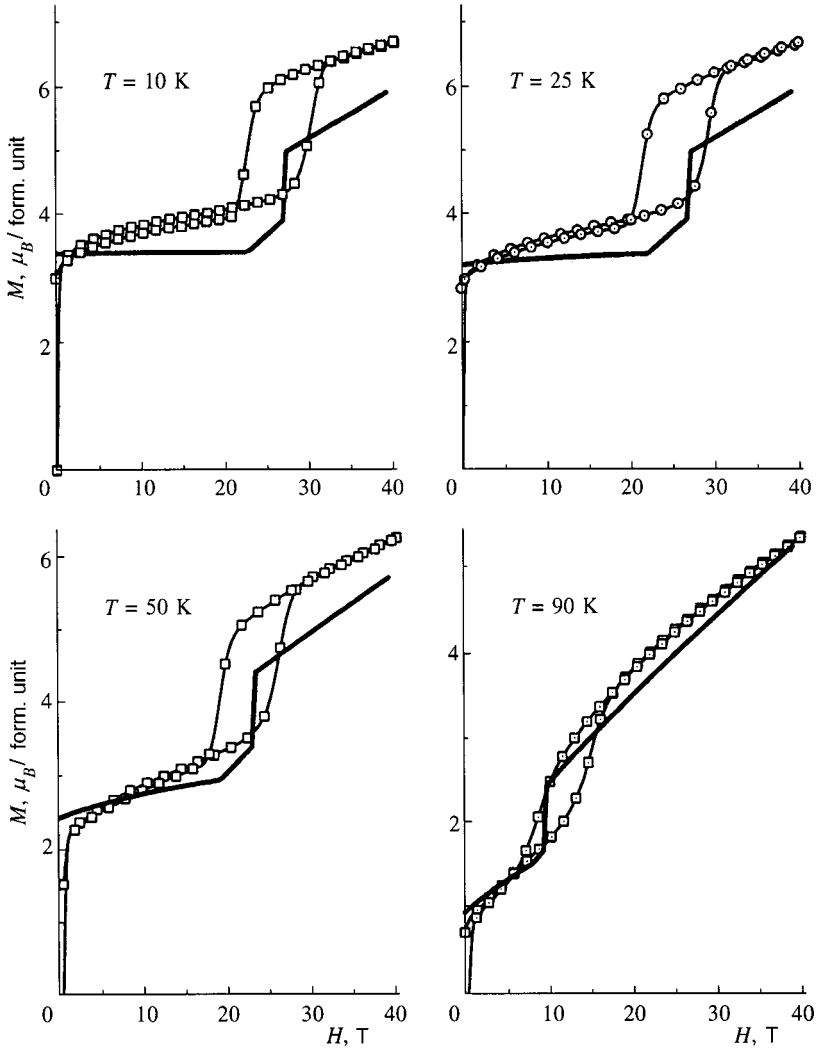


FIG. 6. Field dependence of the magnetization of the intermetallic compound $GdMn_2Ge_2$ at different temperatures. The point symbols are experimental data, the lines are calculated dependences.

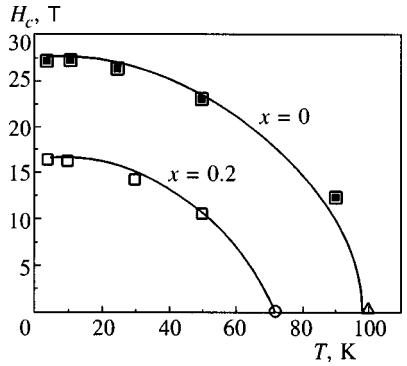


FIG. 7. Temperature dependence of H_c for the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ with $x=0$ and 0.2. Point symbols—experimental data, curves—calculated dependences.

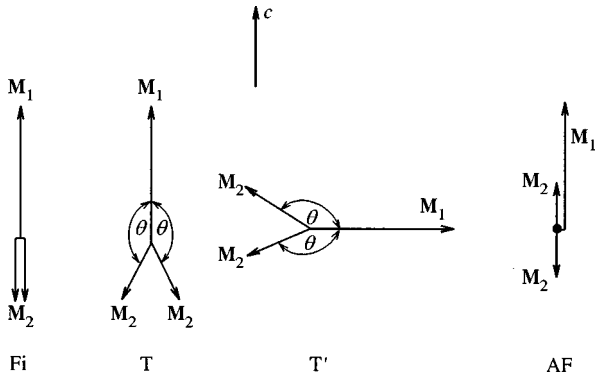


FIG. 8. Possible spontaneous magnetic phases of a ferrimagnet with negative exchange interaction in one of its sublattices.

romagnetic ordering of the manganese subsystem is induced by the rare earth–manganese exchange interaction and is realized only if this interaction is greater than the interplanar Mn–Mn exchange. A decrease in the R–Mn exchange can lead to destruction of collinear ferromagnetic ordering of the manganese subsystem, and the emergence there of noncollinear structures due to competition between the negative R–Mn and Mn–Mn interactions. Application of an external field alters the magnitude of the effective field acting on the manganese subsystem; therefore, it can also cause a change in the magnetic state of this subsystem and the intermetallic compound as a whole.

The theory of ferrimagnets with negative exchange interaction in their sublattices was first considered by Yafet and Kittel[9] within the framework of the theory of Néel ferrimagnetism. However, the authors limited their treatment to spontaneous magnetic structures (in zero field) and ignored magnetic anisotropy. We analyze our experimental results in terms of the Yafet–Kittel model, but with magnetic anisotropy also taken into account, and we also consider the behavior of such a ferrimagnet in an external magnetic field. Since we are interested in the low-temperature properties (at temperatures much less than the magnetic ordering temperature of the intermetallics, which, as was indicated above, is determined by the intraplanar Mn–Mn exchange interaction and is roughly 350–450 K), we take the manganese moment to be independent of temperature and the magnetic field. In addition, we consider only the case in which the resulting moment of the manganese subsystem is less than the moment of the gadolinium subsystem. Taking into account that Gd is an *S* ion with zero orbital angular momentum, we ascribe the magnetic anisotropy of the compounds $Gd_{1-x}Y_xMn_2Ge_2$ to the manganese subsystem, where according to the experimental data of Refs. 5 and 8 we assume that the easy axis of the manganese subsystem is the *c* axis of the crystal. Thus, we are considering here only some special case of the magnetic behavior of a ferrimagnet with negative exchange in one of its subsystems. In general, when these approximations are not invoked, the emergence of magnetic phase diagrams more complicated than the one presented below is possible.

In line with the Yafet–Kittel model we assume that the manganese subsystem in the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ divides into two sublattices with angular momenta $\mathbf{M}_{2'}$ and $\mathbf{M}_{2''}$ of identical magnitude ($M_{2'}=M_{2''}=M_2$), which in general are oriented at some angle with respect to each other and to the magnetic moment of the gadolinium subsystem $(1-x)\mathbf{M}_1$ (triangular magnetic structure). In addition, depending on the sign and magnitude of the magnetic anisotropy and the direction of the magnetic field, the magnetic moments can orient differently relative to the *c* axis of the crystal. The magnetic energy of the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ in the Yafet–Kittel model can be written in the form

$$E=-\lambda_{12}(1-x)\mathbf{M}_1(\mathbf{M}_{2'}+\mathbf{M}_{2''})-\lambda_{22}\mathbf{M}_{2'}\mathbf{M}_{2''}$$
$$-\frac{1}{2}\lambda_{11}(1-x)^2\mathbf{M}_1^2+\varkappa(M_{2'c}^2+M_{2''c}^2)$$
$$-\mathbf{H}[(1-x)\mathbf{M}_1+\mathbf{M}_{2'}+\mathbf{M}_{2''}]. \quad (1)$$

Here the first term describes the exchange interaction of the gadolinium and manganese subsystems ($\lambda_{12}<0$), the second term describes the Mn–Mn exchange ($\lambda_{22}<0$), the third term describes the exchange interaction in the gadolinium subsystem, and the fourth term describes the magnetic anisotropy of the manganese subsystem (M_{2c} is the component of the magnetization in the direction of the tetragonal axis), and the last term is the Zeeman energy.

From the conditions of equilibrium we find that for $\lambda_{22}<0$ in zero external field four different magnetic phases are possible in the system under consideration (Fig. 8).

In the first phase the magnetic moments of the two manganese sub-sublattices are parallel to one another and antiparallel to the magnetic moment of the gadolinium sublattice so that a collinear ferrimagnetic structure is formed (Fi). In the two other phases the magnetic moments of the manganese sublattices make an angle with respect to each other and with the magnetic moment of the gadolinium sublattice so that a triangular magnetic ordering arises. These phases (we denote them T and T′) differ in the orientation of the magnetic moment of the gadolinium sublattice relative to the crystallographic axes: in the first it is parallel, and in the second it is perpendicular to the *c* axis of the crystal. Finally, the fourth phase (AF) corresponds to antiferromagnetic ordering in the manganese sublattice. The energies of the phases depend on the magnitudes of the exchange interactions of the subsystems and the magnetic anisotropy energy.

As these parameters are varied, transitions occur from one phase to another. Thus, for example, for a large enough value of the Gd–Mn exchange in comparison with the Mn–Mn exchange, the Fi phase is stable. A decrease in the magnitude of the Gd–Mn exchange interaction leads to a second-order phase transition to the triangular T phase. Further decrease of the magnitude of the Gd–Mn exchange leads to an increase in the angle between the c axis and the magnetic moments of the manganese subsystem, which leads to growth of the magnetic anisotropy energy and stimulates a first-order transition to the triangular T′ phase, since in this phase the angle between the easy axis (tetragonal axis of the crystal) and the manganese moment $\mathbf{M}_2$ is less than in the T phase, and consequently it is also less than the magnetic anisotropy energy. The AF phase in zero field $H=0$ is realized only in the case of zero Gd–Mn exchange interaction, i.e., when the gadolinium subsystem is disordered.

Let us start with a detailed consideration of the behavior in a field of a free powder whose particles rotate in the field in such a way that the energy is minimized. This case is the most interesting for comparison with the experimental data since the parameters describing the exchange interactions in the investigated intermetallics are most simply determined from such measurements.

In this case, four different magnetic phases are possible in a field: Fi, T, T′, shown in Fig. 8, and the ferromagnetic phase F, in which the magnetic moments of the gadolinium and manganese subsystems are parallel to each other and to the c axis of the crystal. In addition, in the Fi phase, as in an ordinary two-sublattice ferrimagnet, the emergence of an angular phase is possible in which the magnetic moments of the gadolinium and manganese subsystems are oriented at some angle relative to one another and to the external field. However, in the intermetallics under consideration, as our calculations show, the emergence of the angular phase is not energetically favored.

At absolute zero, where it is possible to ignore the dependence of the moment of the gadolinium subsystem on the field, expression (1) easily yields the angular stability of the various magnetic phases for the system $Gd_{1-x}Y_xMn_2Ge_2$:

$$4-2k \leqslant bm(1-x)-2h \tag{2}$$

for the Fi phase,

$$-4+2k \leqslant bm(1-x)-2h \leqslant 4-2k \tag{3}$$

for the T phase,

$$-4-2k \leqslant bm(1-x)-2h \leqslant 4+2k \tag{4}$$

for the T′ phase, and

$$-4+2k \geqslant bm(1-x)-2h \tag{5}$$

for the F phase. Here we have introduced the notation

$$m=\frac{M_1}{M_2}, \quad h=-\frac{H}{\lambda_{22}M_2}, \quad k=-\frac{K}{\lambda_{22}M_2^2}=\frac{2\varkappa}{\lambda_{22}},$$

$$b=\frac{2\lambda_{12}}{\lambda_{22}} \tag{6}$$

(K is the uniaxial anisotropy constant of the manganese subsystem).

Comparing the energies of the various phases, we find that for free powders with weak magnetic anisotropy ($k<2/3$), as the field is increased the following sequence of phase transitions should be observed: Fi→T→T′→T→F. The critical fields of these transitions are given by

$$h=\frac{1}{2}[bm(1-x)-4+2k] \tag{7}$$

for the first-order phase transition Fi→T,

$$h=\frac{1}{2}\sqrt{bm(1-x)-(8-2k^2)} \tag{8}$$

for the first-order phase transition T→T′,

$$h=\frac{1}{2}\sqrt{bm(1-x)+(8-2k^2)} \tag{9}$$

for the first-order re-entrant phase transition T′→T, and

$$h=\frac{1}{2}[bm(1-x)+4-2k] \tag{10}$$

for the second-order phase transition T→F.

If the uniaxial anisotropy constant is increased, then for $2/3<k<2$ the T phase becomes energetically unfavored and the following sequence of phase transitions should be observed: Fi→T′→F, while for $k>2$ the two triangular phases are energetically unfavored and in a field only metamagnetic transitions from the Fi phase to the AF phase and from the latter to the F phase occur. So as not to encumber this paper, we do not give the formulas for the fields corresponding to the phase transitions in these cases, since they are not realized in the system under consideration.

The magnetization in the different phases is described by

$$m_t=m(1-x)-2 \tag{11}$$

for the Fi phase,

$$m_t=m(1-x)\left(1-\frac{b}{4-2k}\right)+\frac{2h}{4-2k} \tag{12}$$

for the T phase,

$$m_t=m(1-x)\left(1-\frac{b}{4+2k}\right)+\frac{2h}{4+2k} \tag{13}$$

for the T′ phase, and

$$m_t=m(1-x)+2 \tag{14}$$

for the F phase.

The above formulas allow one to compare the theoretical and experimental magnetization curves of the investigated system $Gd_{1-x}Y_xMn_2Ge_2$, as well as the phase diagrams of this system at low temperatures. Note that in the fields achieved in the present work, only the first two phase transitions occur, and to observe the two other phase transitions measurements in stronger magnetic fields are needed. Such measurements are being performed at the present time.

At higher temperatures, where it is necessary to take into account the field and temperature dependence of the magne-

tization of the subsystems, the free energies of the various phases must be calculated numerically. Such a calculation was carried out according to the scheme proposed in Refs. 8 and 10. The total free energy of the system was calculated for all possible magnetic phases as a function of temperature and magnetic field as the sum of the internal energy and the entropy term. The magnetization in a given field at a given temperature was calculated for the phase possessing minimum energy under these conditions. This made it possible to calculate the theoretical dependence of the magnetization on the field and temperature and to determine the temperatures of the transitions between the various magnetic phases and compare these results with the experimental data.

Note that at comparatively high temperatures it is possible to describe the magnetic properties of the investigated intermetallic compounds by representing the contribution of the gadolinium subsystem to the total energy of the intermetallic compound within the framework of the Landau theory of phase transitions in the form of the expansion

$$\varepsilon = \frac{\alpha}{2} m^2 + \frac{\beta}{4} m^4. \tag{15}$$

It turns out, however, that calculations within the framework of the Landau theory are less accurate, since this approximation assumes that the moment of the gadolinium subsystem is small, which does not always follow from experiment.

5. COMPARISON OF THE EXPERIMENTAL DATA FOR FREE POWDERS WITH THE RESULTS OF CALCULATIONS

To calculate the theoretical phase diagrams and magnetization curves, it is necessary to know the magnetic moments of the gadolinium and manganese subsystems, the parameters of the exchange interactions describing these subsystems and of the interaction between them, and also the uniaxial magnetic anisotropy constant of the manganese subsystem. For the moment of gadolinium we used the value $7\mu_B$ (as for the trivalent ion), and we took the moment of manganese to be $1.8\mu_B$ according to the available data in the literature for single-crystal $GdMn_2Ge_2$ (Ref. 8). The values of λ_{12}, λ_{22}, and K were determined from our experimental data for the critical field of the metamagnetic transition in $GdMn_2Ge_2$ at 4.2 K ($H_c = 27$ T) and the critical concentration ($x_c = 0.5$), for which $H_c = 0$, and also the magnitude of the initial magnetic susceptibility of $GdMn_2Ge_2$ in the direction of the hard axis (perpendicular to the tetragonal axis) from Ref. 5 ($\chi = 0.62\mu_B/\mathrm{T}\cdot$ formula unit). We assumed that H_c and x_c correspond to a first-order transition from the T phase to the T′ phase, and in the calculation of the anisotropy constant we took account of deformation of the ferromagnetic structure when the field was oriented parallel to the hard axis (for a more detailed treatment, see, e.g., Ref. 11, and also the following section of the present paper). Thus, the exchange parameters and anisotropy constants are given by

$$bm(1-x_c) = \sqrt{2(4-k^2)},$$

$$2h_c = bm - \sqrt{2(4-k^2)},$$

$$\chi' = \frac{(m-2)^2}{2k} + \frac{m}{b}, \tag{16}$$

where $\chi' = -\chi\lambda_{22}$.

The values $\lambda_{12} = -7.7\ \mathrm{T}/\mu_B\cdot$ form. unit and $\lambda_{22} = -10.9\ \mathrm{T}/\mu_B\cdot$ form. unit, found using formulas (16) and the above experimental data, are close to the values obtained in Ref. 8 from an analysis of the temperature dependence of the magnetization of $GdMn_2Ge_2$, whereas the value of the anisotropy constant $K = 15.8\mu_B\cdot$ T/form. unit found by us differs from the values given in Ref. 8, which ignored deformation of the magnetic structure for magnetization along the hard axis and assumed that a large part of the anisotropy is due to the gadolinium subsystem.

We determined the value of the parameter λ_{11} describing the exchange interaction in the gadolinium subsystem by comparing the numerically calculated temperature of the first-order magnetic phase transition in $GdMn_2Ge_2$ with the experimentally determined value (97 K). In line with the data in the literature, we assumed that at this temperature the gadolinium subsystem transitions to the paramagnetic state, and the manganese subsystem becomes antiferromagnetic. The value of λ_{11} turns out to be equal to $2.2\ \mathrm{T}/\mu_B\cdot$ form. unit.

Utilizing the determined parameter values, we calculated the temperature and field dependence of the magnetization of the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ and also the $T-x$, H_c-x, and H_c-T magnetic phase diagrams of this system. They are plotted in Figs. 1, 3–7 along with the experimental data.

Let us first analyze the data for spontaneous magnetization and magnetization in weak fields. As can be seen from the figures, most of the peculiarities in the concentration and temperature dependence of the magnetization of the investigated samples detected experimentally are described within the framework of the proposed model. It follows from Fig. 1 that the temperature dependence of the magnetization of $GdMn_2Ge_2$, calculated theoretically, is in good agreement with our experimental data and with the temperature dependence of the spontaneous magnetization of a single crystal, taken from Ref. 8. The first-order temperature phase transition, experimentally observed in this compound and in compositions with $x = 0.1$ and 0.2 (Fig. 1) and due to demagnetization of the gadolinium subsystem and the onset of antiferromagnetic order in the manganese subsystem, is also described well in the proposed model. The growth of the spontaneous magnetization at 4.2 K of a composition with $x = 0.5$ compared with the magnetization of compositions with $x = 0.3$ and 0.4 is explained by the fact that this intermetallic compound in the ground state is found in the T′ phase, in which the magnetization is greater than in the Fi and T phases (Fig. 1). The peculiarities of the temperature dependence of the magnetization of compositions with $x = 0.3$ and 0.4, according to the the developed model, are due to the fact that as the temperature is raised, a first-order phase transition from the T phase to the T′ phase takes place, accompanied by a jump in magnetization. With a further increase in temperature in these compositions, as in the composition with $x = 0.5$, a second-order phase transition of the manganese subsystem to the antiferromagnetic phase occurs

along with demagnetization of the gadolinium subsystem (the AF phase), see Fig. 1. The experimental $T-x$ diagram, as Fig. 3 shows, is in satisfactory agreement with the theoretical calculations.

Figures 4 and 6 compare some experimental and theoretical field dependences of the magnetization of the intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$. It can be seen that the theoretical values of the fields corresponding to the metamagnetic transitions, calculated theoretically, like the fields corresponding to the first-order transition from the T phase to the T′ phase, are in good agreement with the experimental values. (So as not to clutter the figures, we do not show the theoretically calculated hysteresis loops. Note, however, that the theoretical width of the hysteresis loop, determined from the lability loss fields of the T and T′ phases, is significantly greater than that experimentally observed.) Further evidence of the good agreement of theory with experiment is seen by comparing the theoretical and experimental dependences of $H_c(x)$ at 4.2 K (see the inset to Fig. 4 for $x=0.5$) and the temperature dependence of H_c shown in Fig. 7. At the same time, the theoretically calculated magnitude of the magnetization jump at the metamagnetic transition is considerably less than the experimentally observed value. Also, the theoretically predicted change in the slope of the $M(H)$ curve due to the second-order phase transition from the Fi phase to the T phase is not seen in the experimental curves.

6. MAGNETIZATION OF SINGLE CRYSTALS AND POLYCRYSTALLINE SAMPLES

Above we described experimental data for free powders whose particles are free to rotate in a magnetic field. Naturally, the behavior of fixed samples (single crystals and polycrystalline samples), whose orientation does not change over the course of the experiment, should differ from the magnetic behavior of free powders. It is interesting to analyze how the model developed above describes the magnetic properties of such samples.

Calculations based on the Yafet–Kittel model [formula (1)] show that the for a field aligned with the c axis of the crystal, the problem reduces to a consideration of the antiferromagnetic manganese subsystem in the effective field equal to the sum of the external field and the exchange field with the gadolinium subsystem. As the field is increased, the following sequence of phases should be observed: Fi→T→AF→T→F if the anisotropy is small ($k<2$) and Fi→AF→F if $k>2$. As follows from the above argument, the examined system $Gd_{1-x}Y_xMn_2Ge_2$ is weakly anisotropic ($k<2$). In addition, estimates show that in the investigated range of magnetic fields, only Fi→T and T→AF phase transitions are possible. We therefore restrict our discussion to these transitions.

The field corresponding to the second-order phase transition Fi→T is given by formula (7), and the magnetization in these phases is described by formulas (11) and (12), respectively. The field corresponding to the first-order transition T→AF is

$$h=\frac{1}{2}bm(1-x)-\sqrt{k(2-k)}, \tag{17}$$

and the magnetization in the AF phase is

$$m_t=m. \tag{18}$$

Estimates obtained using these formulas and the parameter values determined above from measurements in free powders yield 24 T for the field corresponding to the transition Fi→T and 38 T for the field corresponding to the transition T→AF for $GdMn_2Ge_2$ at 4.2 K.

For magnetization aligned with the hard axis (field perpendicular to the c axis of the crystal) in fields exceeding the technical saturation field, when the magnetization of the gadolinium subsystem is oriented perpendicular to the tetragonal axis of the crystal, the problem also reduces to a consideration of the behavior of a uniaxial antiferromagnet in an effective field. In this case, the sequence of second-order transitions Fi′→T′→F′ is possible when the field is increased (the prime denotes phases in which the magnetization of the gadolinium subsystem is oriented perpendicular to the tetragonal axis of the crystal). The fields corresponding to these transitions are given by

$$h=\frac{1}{2}[bm(1-x)\pm 2(2+k)], \tag{19}$$

where the minus sign pertains to the transition Fi′→T′, and the plus sign corresponds to the transition T′→F′. As calculations show, for $GdMn_2Ge_2$ at 4.2 K, the transition from the Fi′ phase to the T′ phase should take place in a field of roughly 2 T, and the ferromagnetic phase F′ arises in fields of roughly 100 T. Since according to the data of Ref. 5 the saturation field in this compound (greater than 8 T) is significantly greater than the field corresponding to the transition from the Fi′ phase to the T′ phase, it is clear that the Fi′ phase is not realized for the magnetization aligned with the hard axis and the T′ phase arises in the technical saturation field.

Theoretical estimates and numerical calculations show that when the magnetic moments of the gadolinium and manganese magnetic subsystems rotate from the tetragonal axis to the plane perpendicular to it, a distortion of the collinear ferrimagnetic structure takes place; however, the manganese subsystem remains in the ferromagnetic state if the magnetic moment of the gadolinium subsystem is noncollinear with the direction of the field. At some critical field, the moment of the gadolinium subsystem undergoes a jumplike rotation to the direction of the field accompanied by the emergence of the T′ phase, i.e., a first-order phase transition takes place, accompanied by a jump in the magnetization. In Fig. 9 the numerically calculated magnetization curve of a $GdMn_2Ge_2$ single crystal at 77 K for the field oriented perpendicular to the tetragonal axis is compared with the experimental dependence of the magnetization of the single crystal on the field taken from Ref. 5. It can be seen that the transition to the T′ phase is a first-order phase transition. As can be seen from this figure, the experimental field dependence and the theoretically calculated field dependence of the magnetization are quite close, although the numerical values of the field corresponding to the first-order phase transition differ.

The data obtained above on the theoretical field dependence of the magnetization of a $GdMn_2Ge_2$ single crystal in

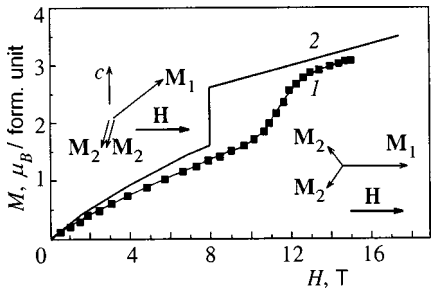


FIG. 9. Experimental (*1*) and theoretical (*2*) field dependence of the magnetization of $GdMn_2Ge_2$ at 77 K. The field is oriented perpendicular to the tetragonal axis. Magnetic structures in weak and strong fields are indicated.

various directions afford a qualitative understanding of the magnetic properties of polycrystalline samples of this compound. Figure 10 plots the field dependence of the magnetization for two different polycrystalline samples of $GdMn_2Ge_2$ at 4.2 K (Ref. 6). It can be seen that a smeared metamagnetic transition is observed in fields of roughly 15–35 T. Note that details of the transition are different for different samples, probably because of a different degree of texture. Comparison with the calculated dependences of $M(H)$ for a single crystal suggests that this transition is associated with the metamagnetic transitions in various crystallographic directions of the crystal described above, and its smearing is due to differences in the orientation of the crystallites relative to the direction of the field.

7. CONCLUSION

Comparison of the experimental data on the magnetization for the mixed intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ with the theoretically calculated magnetic characteristics in the Yafet–Kittel model for ferrimagnets with negative intrasublattice exchange shows that this simple model adequately describes the majority of the experimental data. With its help it is possible to explain the metamagnetic transitions observed both in free powders and in single crystals of the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ in a magnetic field, and also the nature of the spontaneous magnetic phase transitions and peculiarities of the field and concentration dependence of the magnetization of these compounds. Moreover, the present model provides a quantitative description of the experimental $T-x$ and $H-x$ magnetic phase diagrams of this system, the temperature dependence of the fields corresponding to the metamagnetic transition, etc.

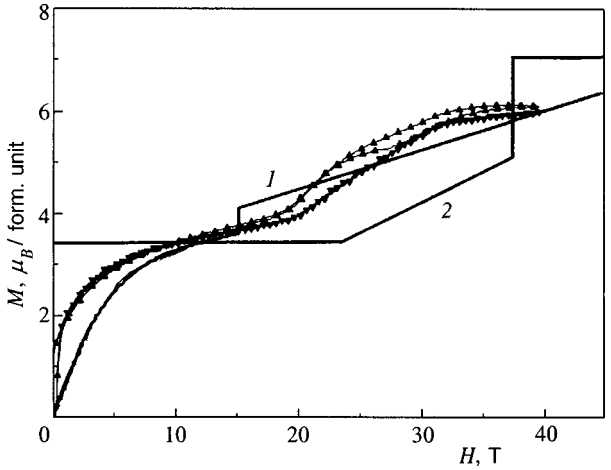


FIG. 10. Theoretical field dependences of the magnetization of a $GdMn_2Ge_2$ single crystal at 4.2 K in various crystallographic directions [field perpendicular (*1*) and parallel (*2*) to the *c* axis] and experimental field dependences of the magnetization (triangles) for two polycrystalline samples of $GdMn_2Ge_2$ at this temperature.

However, it cannot be said that this model provides a completely accurate description of the magnetic properties. We have already noted that the theoretically calculated magnetization jump accompanying a metamagnetic transition is substantially smaller than the value experimentally observed.

We were not able to experimentally detect the T phase (it is possible, however, that this is because it is realized in a narrow range of fields and temperatures and the magnetic parameters of this phase are similar to those of the Fi phase and are therefore hard to distinguish experimentally).

In the experimental dependences of the magnetization of free powders at low temperatures in fields less than the metamagnetic transition, a paraprocess is observed, whereas according to the present model, the susceptibility of the paraprocess should be significantly lower.

In addition, to describe the experimental data using this model, it is necessary to assume that the intrinsic magnetic ordering in the gadolinium subsystem is ferromagnetic ($\lambda_{11}>0$), although in many other type 1-2-2 intermetallics the gadolinium subsystem is ordered antiferromagnetically.[12]

Possibly, all these differences are due to the approximate nature of the model, since it takes into account only exchange interactions between atoms of manganese in neighboring planes and ignores long-range interactions, although they are probably important for a description of long-period magnetic structures observed in the manganese subsystem of some RMn_2Ge_2 intermetallics.[13]

In addition, the first-order transition in temperature in the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ is accompanied by a significant magnetoelastic anomaly (see Fig. 2), and as our measurements have shown, magnetostriction at the metamagnetic transition in $GdMn_2Ge_2$ also has order of magnitude 10^{-3}. Therefore, it is necessary, in principle, to take the magnetoelastic energy into account in the calculations. It is possible to estimate the contribution of the magnetoelastic energy by comparing the change in the exchange energy at the transition of the manganese subsystem from the ferromagnetic to the antiferromagnetic state, $\Delta E_e=2\lambda_{22}M_2^2$, with the change in the elastic energy at this transition, $\Delta E_{\rm el}=c_{11}\varepsilon^2/2$, where ε is the magnitude of the relative elastic deformation at this transition ($\approx 10^{-3}$), and c_{11} is the corresponding elastic modulus (since data on the elastic properties of the examined intermetallics are lacking, we have adopted the standard value for c_{11}: 5×10^{11} dyn/cm^2). These estimates show that in the intermetallics $Gd_{1-x}Y_xMn_2Ge_2$ the magnetoelastic interaction makes a small contribution (on the order of a few percent) to the change in the energy at the transitions of the manganese subsystem from the ferromagnetic to the antiferromagnetic state. However, this transition can be substantial in other intermetallics of type 1-2-2, e.g., in $SmMn_2Ge_2$, in

which the change in the exchange energy at the ferromagnetic–antiferromagnetic transition according to our estimates is at least an order of magnitude smaller than in gadolinium intermetallics.

Finally, we have considered the phase transitions in the system of intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ as purely magnetic. It is possible, however, that a change takes place in the electronic state of the manganese subsystem at these transitions. This is indicated, in particular, by the fact that the magnetic moment of manganese in RMn_2Ge_2 intermetallics with light rare earths, in which the manganese subsystem is ordered ferromagnetically, exceeds $3\mu_B$, which is significantly greater than in intermetallics with heavy rare earths (where it is not greater than $2.4\mu_B$), in which this subsystem is antiferromagnetic (see, e.g., Ref. 14). This problem requires additional study.

The authors thank N.P. Kolmakov for fruitful discussions.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 99-02-17358), by Grant No. 96-15-96429 in support of scientific schools, and by the Russian Ministry of Science in the Fundamental Natural Sciences (Grant No. 97-0-7.3-157). One of us (A.Yu.S.) expresses his gratitude to ISGRS (Kyoto University, Japan) for financial support.

*)E-mail: levitin@plms.phys.msu.su

[1] A. Szytula and J. Leciejewicz, in *Hahdbook Physics and Chemistry of Rare Earths*, Vol. 12, K. A. Gschneidner, Jr. and L. Eyring (eds.), North-Holland, Amsterdam (1989), p. 133.
[2] G. Venturini, B. Malaman, and E. Ressouche, J. Alloys Compd. **240**, 139 (1996).
[3] R. Welter, G. Venturini, E. Ressouche, and B. Malaman, J. Alloys Compd. **218**, 204 (1995).
[4] G. Venturini, J. Alloys Compd. **232**, 133 (1996).
[5] H. Kobayashi, H. Onodera, and H. Yamomoto, J. Magn. Magn. Mater. **79**, 76 (1989).
[6] A. Sokolov, H. Wada, M. Shiga, and T. Goto, Solid State Commun. **105**, 289 (1998).
[7] R. Verhoef, F. R. de Boer, J. J. M. Franse, C. J. M. Denissen, N. H. Jacobs, and K. H. J. Buschow, J. Magn. Magn. Mater. **80**, 41 (1989).
[8] N. Iwata, K. Hattori, and T. Shigeoka, J. Magn. Magn. Mater. **53**, 318 (1986).
[9] Y. Yafet and C. Kittel, Phys. Rev. **87**, 290 (1952).
[10] H. Wada, H. Yamaguchi, and M. Shiga, J. Magn. Magn. Mater. **152**, 165 (1996).
[11] R. Z. Levitin, E. M. Savitskiĭ, V. F. Terekhova, O. D. Chistyakov, and V. L. Yakovenko, Zh. Éksp. Teor. Fiz. **62**, 1858 (1972) [Sov. Phys. JETP **35**, 968 (1972)].
[12] L. D. Tung, J. J. M. Franse, K. H. J. Buschow, P. E. Brommer, and N. E. Thuy, J. Alloys Compd. **260**, 35 (1997).
[13] G. Venturini, R. Welter, E. Ressouche, and B. Malaman, J. Alloys Compd. **210**, 213 (1994).
[14] G. Venturini, B. Malaman, and E. Ressouche, J. Alloys Compd. **241**, 135 (1996)

Translated by Paul F. Schippnick

Nonlinear domain-wall dynamics in a system of two magnetic layers

A. K. Zvezdin

General Physics Institute, Russian Academy of Sciences, 117942 Moscow, Russia

V. V. Kostyuchenko*)

Institute of Microelectronics of the Russian Academy of Sciences, 150007, Yaroslavl'

(Submitted 22 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 1365–1374 (October 1999)

The nonlinear dynamics of the magnetization in a spin-valve structure is investigated. Equations describing the dynamics of the magnetization in such a structure are obtained. The stability of the solution corresponding to a motionless flat domain wall is investigated. The nonlinear domain-wall dynamics are investigated in the approximation of a strong exchange interaction between the magnetic layers and in the approximation of a large magnetostatic energy. In the former case the nonlinear dynamical equations are shown to be similar to the equations describing the dynamics of the magnetization in a weak ferromagnet, and in the latter case they are similar to the equations of motion of a magnetic vortex (i.e., a vertical Bloch line) in a domain wall. © *1999 American Institute of Physics.* [S1063-7761(99)01710-2]

1. INTRODUCTION

The dynamical equations for antiferromagnets are known to differ significantly from the equations derived by Landau and Lifshitz[1] for a ferromagnet (see, for example, Ref. 2). This difference is especially clear in the case of nonlinear magnetization dynamics (see Refs. 3 and 4 and the references therein). It is noteworthy that the dynamics of topological kink solitons (domain walls) in weak ferromagnets have been thoroughly studied by both theoretical and experimental methods.[4] The research performed in this area has revealed many interesting phenomena in the physics of nonlinear waves and magnetic phenomena.

Systems consisting of several magnetic layers with an antiferromagnetic interaction between the layers are closely related to two-sublattice magnets in a certain sense. An investigation of phase transitions in magnetic sublattices that are associated with changes in the magnetization throughout the sublattice showed that the two-sublattice approximation is valid in the limit of a large number of layers.[5] However, it should be noted that superlattices with a finite number of layers can undergo phase transitions that are associated with changes in the magnetization only near the surface, i.e., surface spin-flop transitions.[6–9] In order to avoid the need to take into account the surface-related effects, we shall consider a structure in which the two-sublattice approximation is sure to be valid, i.e., a system of two magnetic layers.

However, there is a fundamental difference between ordinary two-sublattice antiferromagnets and multilayer magnetic structures. In ordinary antiferromagnets there is a clear-cut hierarchy of energy scales: $J \gg K \gg h$, where J is the exchange energy, K is the anisotropy energy, and h is the Zeeman energy. In multilayer magnetic structures all of these quantities can be of the same order. The study of the static properties of these materials has already shown that this difference leads to a richer phase diagram for the phase transitions (see, for example, Refs. 10 and 11). It is natural to expect that this difference also leads to significant changes in the dynamical properties. This conclusion is supported, in particular, by the theoretical investigations of the domain-wall structure in a classical antiferromagnetic spin chain performed in Ref. 12 in the continuum approximation.

In this paper we systematically examine the domain-wall dynamics in a system consisting of two magnetic layers with an antiferromagnetic interaction between the layers. In Sec. 2 we derive equations describing the dynamics of the magnetization in such a system from the expression for the Lagrangian density in a single-sublattice ferromagnet. The system of equations obtained in Sec. 2 is similar to the dynamical equations for a two-sublattice antiferromagnet.

This system of equations allows an exact static solution, which has the form of a Neél wall in each layer. The exchange interaction between the layers leads to coupling of the Neél walls in such a manner that the angle between the magnetization vectors in the layers does not depend on the coordinates directed along the layers and is equal to π. A schematic representation of the structure under consideration is shown in Fig. 1. Such a domain-wall structure is stable only for a small value of the biquadratic exchange energy between the layers. If the energy of the biquadratic exchange interaction between the layers exceeds the Heisenberg exchange energy, a domain wall with a constant angle between the magnetization vectors in the layers no longer satisfies the condition for a minimum of the thermodynamic potential, i.e., spontaneous violation of the domain-wall symmetry occurs. The condition for stability of a symmetric domain wall is explored in detail in Sec. 3.

As a domain wall moves, the dynamical moments appearing because of rotation of the magnetization vector at its center cause deviation of the magnetization vectors from the

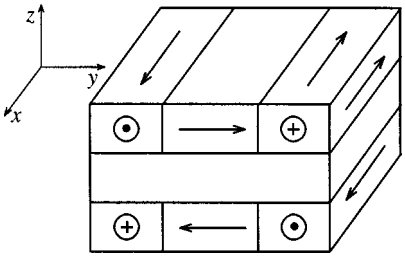


FIG. 1. Schematic representation of the structure of a motionless domain wall.

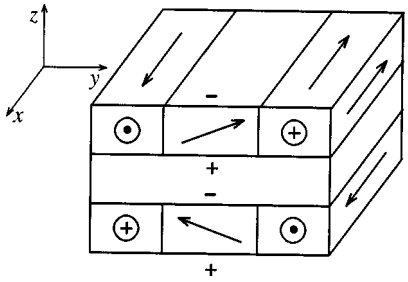


FIG. 3. Schematic representation of the structure of a moving domain wall. Case in which the magnetostatic energy dominates.

plane of each layer. A description of the domain-wall dynamics for the system under consideration in general form is hardly possible, since its dynamical properties vary significantly, depending on which physical mechanism is dominant. In this paper we shall examine two limiting cases, viz., strong exchange coupling between the layers and a large magnetostatic energy.

The case of strong exchange coupling between the layers is considered in Sec. 4. A schematic representation of the structure of a domain wall in motion for this case is presented in Fig. 2. Like magnetic charges appear in this structure near the surfaces of the nonmagnetic intermediate layer. Their interaction leads to an additional increase in the magnetostatic energy. From the standpoint of the physics of the phenomenon, the situation is the same as in the case of a domain wall in a two-sublattice antiferromagnet. The fact that the dynamical equations derived in Sec. 4 are similar to the equations describing the domain-wall dynamics for a weak ferromagnetic supports this conclusion. In this approximation a domain wall is a soliton (kink) of the sine–Gordon equation. Expressions for its thickness and velocity are given in Sec. 4.

The picture of the domain-wall dynamics is entirely different, if the magnetostatic energy dominates, as, for example, in layered Fe/Cr/Fe or Co/Cu/Co systems, in which the magnetostatic energy exceeds the exchange energy by an order of magnitude (see, for example, Ref. 13 and the references therein). In this case the magnetic charges on the surfaces of the nonmagnetic layer have different signs and partially compensate the increase in the magnetostatic energy caused by the deviation of the magnetization vectors from the midplane of the magnetic layer in the domain wall (see Fig. 3). Such a structure of a domain wall in motion is reminiscent of the distribution of the magnetization near its center as a Bloch line moves. This qualitative argument is confirmed by the calculations performed in Sec. 5. In the case of a small deviation of the magnetization from the planes of the layers, the equation describing the domain-wall dynamics are similar to the equations of motion of vertical Bloch lines interacting with one another (see, for example, Ref. 14). If the uniaxial anisotropy is small, the system of equations describing the dynamics of such a formation reduces to a system of coupled sine–Gordon equations. However, in this case the velocity of the domain wall and its thickness differ from those obtained in Sec. 4.

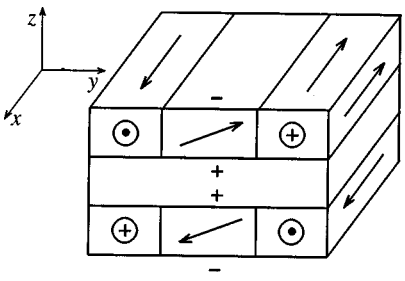


FIG. 2. Schematic representation of the structure of a moving domain wall. Case of a strong exchange interaction between the layers.

2. BASIC EQUATIONS

We assume that the thickness of each magnetic layer (d) is less than the thickness of a domain wall in a bulk sample ($d \ll \Delta$). Let the z axis be directed along a normal to the surface of the layers. Then the dependence of the magnetization on the z coordinate within each layer can be neglected. Within this approximation, the problem of calculating the dependence of the magnetization $\mathbf{M}(x,y,z,t)$ on the three spatial coordinates and time reduces to the problem of calculating the dependence of the two quantities $\mathbf{M}_1(x,y,t)$ and $\mathbf{M}_2(x,y,t)$ on only two spatial coordinates and time, where $\mathbf{M}_i$ is the magnetization in the ith magnetic layer.

To investigate the domain-wall dynamics in such a two-layer structure, we start from the variational principle

$$\delta \int L_S dS dt = 0. \tag{1}$$

It is convenient to express the Lagrangian density L_S in the angular variables θ_i and φ_i, which specify the orientation of the magnetization in the ith magnetic layer. The polar angles θ_i are measured from the z axis, and the azimuthal angles φ_i are measured from the x axis in the xy plane.

In the angular variables θ_i and φ_i the Lagrangian density L_S has the form

$$L_S=\sum_{i=1}^{2}\left[\frac{M}{\gamma}\dot{\varphi}_i\cos\theta_i+2\pi M^2\cos^2\theta_i-MH\sin\theta_i\cos\varphi_i+\frac{1}{2}K\sin^2\theta_i\sin^2\varphi_i+\frac{1}{2}A[(\nabla\theta_i)^2+\sin^2\theta_i(\nabla\varphi_i)^2]\right]+\frac{1}{2}J_1[\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]+\frac{1}{2}J_2[\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]^2, \tag{2}$$

where γ is the gyromagnetic ratio, M is the magnetization in each ferromagnetic layer, H is the external magnetic field parallel to the x axis, K is the uniaxial anisotropy constant, A is the inhomogeneous exchange constant, and J_1 and J_2 are, respectively, the constants of the Heisenberg and biquadratic exchange interactions between the magnetic layers. The relation (2) was obtained from the expression for the Lagrangian density for a single-sublattice ferromagnet with consideration of the exchange interaction between the magnetic layers.

A system of equations which describes the domain-wall dynamics follows from (2):

$$\frac{1}{2}K\sin^2\theta_1\sin(2\varphi_1)+MH\sin\theta_1\sin\varphi_1+\frac{M}{\gamma}\dot{\theta}_1\sin\theta_1-A\nabla(\sin^2\theta_1(\nabla\varphi_1))-\frac{1}{2}J_1\sin\theta_1\sin\theta_2\sin(\varphi_1-\varphi_2)-J_2\sin\theta_1\sin\theta_2\sin(\varphi_1-\varphi_2)[\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]=0, \tag{3}$$

$$\frac{1}{2}\sin(2\theta_1)[K\sin^2\varphi_1+A(\nabla\varphi_1)^2-4\pi M^2]-A\nabla^2\theta_1-MH\cos\theta_1\cos\varphi_1-\frac{M}{\gamma}\dot{\varphi}_1\sin\theta_1+\frac{1}{2}J_1\times[-\sin\theta_1\cos\theta_2+\cos\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]+J_2[\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]\times[-\sin\theta_1\cos\theta_2+\cos\theta_1\sin\theta_2\cos(\varphi_1-\varphi_2)]=0. \tag{4}$$

The equations for θ_2 and φ_2 have a similar form to within accuracy to interchange of the indices $1\leftrightarrow 2$.

The system of equations (3) and (4) has a very large number of diverse types of solutions. In this paper we consider the case where the angle between the magnetization vectors in the layers is close to π. A necessary condition for this is the condition of stability of the antiferromagnetic phase for a homogeneous sample (i.e., in the absence of domain boundaries). This condition has the form (see, for example, Ref. 10)

$$H<\sqrt{2K(J_1-2J_2)+K^2}/M. \tag{5}$$

3. INVESTIGATION OF THE STABILITY OF THE EXACT STATIC SOLUTION

The case where the angle between the magnetization vectors in the layers maintains a constant value equal to π is very important. This condition is satisfied by the exact static solution of the system of equations (3) and (4), which corresponds to a motionless domain boundary,

$$\theta_1^0=\theta_2^0=\frac{\pi}{2},\quad \varphi_1^0=\varphi_2^0-\pi=2\arctan\left[\exp\left(\frac{y}{\Delta}\right)\right], \tag{6}$$

where $\Delta=\sqrt{A/K}$. It is not difficult to see that the condition $H=0$ must be satisfied for such a solution to exist. We substitute $\theta_i=\theta_i^0+\delta\theta_i$ and $\varphi_i=\varphi_i^0+\delta\varphi_i$ into (3) and (4). Then, in the linear approximation with respect to $\delta\theta_i$ and $\delta\varphi_i$ the system of equations (3) and (4) can be represented in the form

$$\frac{d\mathbf{q}}{dt}=\hat{D}\mathbf{q}, \tag{7}$$

where $\mathbf{q}=\{\delta\varphi_1,\delta\varphi_2,\delta\theta_1,\delta\theta_2\}$,

$$\hat{D}=\frac{\gamma}{M}\begin{bmatrix}0 & 0 & K-\hat{P}-4\pi M^2 & -\frac{J_1}{2}+J_2\\ 0 & 0 & -\frac{J_1}{2}+J_2 & K-\hat{P}-4\pi M^2\\ \hat{P} & -\frac{J_1}{2}+J_2 & 0 & 0\\ -\frac{J_1}{2}+J_2 & \hat{P} & 0 & 0\end{bmatrix},$$

$$\hat{P}=K\hat{N}+\frac{J_1}{2}-J_2,\quad \hat{N}=-\frac{\partial^2}{\partial y'^2}+1-2\,\mathrm{sech}\,y',\quad y'=\frac{y}{\Delta}.$$

The solution of the system of equations (3) and (4) is stable if the real parts of the eigenvalues of the operator $\hat{D}$ are nonpositive. The eigenfunctions and eigenvalues of $\hat{N}$ are well known (see, for example, Ref. 15). The discrete spectrum is described by the eigenfunction $f_b=\mathrm{sech}\,y'$ with the eigenvalue $\omega_b=0$, and the continuous spectrum is described by the eigenfunction

$$f_\chi=\frac{(\chi+i\tanh y')\exp(i\chi y')}{\sqrt{1+\chi^2}}$$

with the eigenvalues $\omega_\chi^2=1+\chi^2$. Using these data, we can easily calculate the eigenvalues of $\hat{D}$. The eigenvalues corresponding to the discrete spectrum are

$$\lambda_{1,2}=0,\quad \lambda_{3,4}=\pm\sqrt{(J_1-2J_2)(K-4\pi M^2)}, \tag{8}$$

and the eigenvalues corresponding to the continuous spectrum are

$$\lambda_{1,2}(\chi)=\pm\{2K\sqrt{1+\chi^2}[K(1-\sqrt{1+\chi^2})+2J_2-J_1-4\pi M^2]\}^{1/2},$$
$$\lambda_{3,4}(\chi)=\pm\sqrt{2}\{[K(1-\sqrt{1+\chi^2})-4\pi M^2][J_1-2J_2+K\sqrt{1+\chi^2}]\}^{1/2}. \tag{9}$$

Since the inequality $K<4\pi M^2$ usually holds in thin magnetic layers, the condition for the stability of the solutions of (3) can be written in the form $J_1>2J_2$.

Thus, when the biquadratic exchange energy is small ($J_1>2J_2$) the angle between the magnetization vectors in the adjacent layers is a constant and does not depend on the coordinates in the domain-wall plane. If $J_1<2J_2$, the angle between the magnetization vectors in the adjacent layers is a function of the coordinate perpendicular to the domain-wall plane, and the trajectory of the magnetization in the domain wall has a significantly more complicated form.

Since the inequality $J_1 \gg J_2$ usually holds except in some special cases, we shall assume below that $J_2=0$ and, therefore, the static domain wall described by (3) and (4) is stable.

4. CASE OF A STRONG EXCHANGE INTERACTION BETWEEN LAYERS

In the dynamical case the presence in Eqs. (3) and (4) of the terms proportional to $\dot\theta_i \sin\theta_i$ and $\dot\varphi_i \sin\theta_i$, which are associated with rotation of the magnetization vector at the kink center, causes the magnetization vector to deviate from the plane of the magnetic layer. The influence of this dynamical factor is counteracted by two static factors. One of them is associated with the antiferromagnetic interaction between the layers and acts to orient the magnetization vectors in the layers in an antiparallel configuration. The other factor is associated with the magnetostatic energy, which hinders deviation of the magnetization vectors from the planes of the layers. Thus, the magnetization dynamics are determined by which of these two factors predominates. In this section we consider the case where $J_1 \gg 4\pi M^2$ and $J_1 \gg K$, i.e., the case of strong exchange coupling between the layers. In this situation the angle between the magnetization vectors in the adjacent layers is close to π. Therefore, it is convenient to go over to angular variables

$$\begin{cases} \theta_1=\theta+\varepsilon, \\ \theta_2=\pi-\theta+\varepsilon, \end{cases} \quad \begin{cases} \varphi_1=\varphi+\beta, \\ \varphi_2=\pi+\varphi-\beta, \end{cases} \tag{10}$$

where $\varepsilon,\beta \ll 1$. The Lagrangian density L_S in the lowest order with respect to ε and β is given by the expression

$$\begin{aligned} L_S=&\frac{2M}{\gamma}(\dot\beta\cos\theta-\varepsilon\dot\varphi\sin\theta)+4\pi M^2\cos^2\theta+J_1(\varepsilon^2 \\ &+\beta^2\sin^2\theta)-2MH(\varepsilon\cos\theta\cos\varphi-\beta\sin\theta\sin\varphi) \\ &+K\sin^2\theta\sin^2\varphi+A[(\nabla\theta)^2+\sin^2\theta(\nabla\varphi)^2], \end{aligned} \tag{11}$$

and the dynamical equations can be represented in the form

$$\varepsilon=\frac{MH}{J_1}\cos\theta\cos\varphi+\frac{M}{\gamma J_1}\dot\varphi\sin\theta, \tag{12}$$

$$J_1\beta\sin\theta+MH\sin\varphi+\frac{M}{\gamma}\dot\theta=0, \tag{13}$$

$$\begin{aligned} &-\frac{2M}{\gamma}(\dot\beta\sin\theta+\varepsilon\dot\varphi\cos\theta)-4\pi M^2\sin(2\theta) \\ &+K\sin(2\theta)\sin^2\varphi+J_1\beta^2\sin(2\theta) \\ &+2MH(\varepsilon\sin\theta\cos\varphi+\beta\cos\theta\sin\varphi)+A\sin(2\theta) \\ &\times(\nabla\varphi)^2-2A\nabla^2\theta=0, \end{aligned} \tag{14}$$

$$\begin{aligned} &K\sin^2\theta\sin 2\varphi+2MH(\varepsilon\cos\theta\sin\varphi+\beta\sin\theta\cos\varphi) \\ &+\frac{M}{\gamma}(\dot\varepsilon\sin\theta+\varepsilon\dot\theta\cos\theta)-2A\nabla(\sin^2\theta\nabla\varphi)=0. \end{aligned} \tag{15}$$

If the external magnetic field is weak, we can set $\theta \simeq \pi/2$ in Eqs. (12)–(15). (A stricter condition for the applicability of this approximation will be obtained below.) Then, from Eqs. (12) and (13) we obtain the following expressions for ε and β:

$$\varepsilon=\frac{M\dot\varphi}{\gamma J_1}, \quad \beta=-\frac{MH}{J_1}\sin\varphi. \tag{16}$$

Substituting (16) into (15), we obtain an equation for φ:

$$\tilde\Delta^2\left(\frac{1}{c^2}\ddot\varphi-\nabla^2\varphi\right)+\frac{1}{2}\sin(2\varphi)=0, \tag{17}$$

where $c=\gamma\sqrt{2AJ_1}/M$ and $\tilde\Delta=\Delta/\sqrt{1-(MH)^2/KJ_1}$. Equation (17) is a sine–Gordon equation, whose exact multisoliton solutions are well known. The following solution of the moving kink type corresponds to a domain wall:

$$\varphi=\pm 2\arctan\left[\exp\left(\pm\frac{y-vt}{\tilde\Delta\sqrt{1-v^2/c^2}}\right)\right]. \tag{18}$$

Let us now examine the question of the applicability of the approximations used in this section in greater detail. It follows from the conditions $\varepsilon \ll 1$ and $\beta \ll 1$ that $|MH|/J_1 \ll 1$ and $vM/J_1\gamma\tilde\Delta\sqrt{1-v^2/c^2} \ll 1$. The first of these conditions means that the Zeeman energy should be small compared with the Heisenberg exchange energy between the magnetic layers. The second condition shows that the approximation under consideration is inapplicable near the limiting velocity $v\sim c$. Let us now consider the applicability of the approximation $|\theta-\pi/2| \ll 1$. From (15) we can easily obtain the estimate

$$\left|\theta-\frac{\pi}{2}\right| \simeq \frac{M^2Hv}{J_1\gamma\tilde\Delta\sqrt{1-v^2/c^2}\,|K-4\pi M^2+(MH)^2/J_1|} \ll 1. \tag{19}$$

It is not difficult to see that this condition is automatically satisfied if the previously used approximations are taken into account.

The equations (12)–(15) obtained are similar to the equations describing the dynamics of the magnetization in a two-sublattice antiferromagnet (see, for example, Ref. 16). Therefore, the approximation considered above can be called the antiferromagnet approximation.

5. CASE OF PREDOMINANCE OF THE MAGNETOSTATIC ENERGY

Let us now consider the case of a weak exchange interaction between the layers. If the thickness of the intermediate layer between the magnetic layers amounts to tens of angstroms, the exchange interaction between the layers is small compared with the magnetostatic energy. For example, for the Fe/Cr/Fe system we have $J_1/(4\pi M^2)\sim 0.1$ (see, for example, Ref. 13), i.e., a situation which is the reverse of the situation considered in the preceding section arises.

The magnetostatic energy hinders any deviation of the magnetization vector from the planes of the magnetic layers. Since it dominates in the case under consideration, in (2) it is convenient to go over to the variables

$$\alpha_i = \pi/2 - \theta_i . \tag{20}$$

Since $\alpha_i \ll 1$, the expression for the Lagrangian density (2) can be simplified by leaving only the terms having the lowest order with respect to α_i in each summand. In this approximation the Lagrangian density equals

$$L_S = \sum_{i=1}^{2} \left\{ \frac{M}{\gamma}\alpha_i\dot{\varphi}_i + 2\pi M^2\alpha_i^2 + \frac{1}{2}K\sin^2\varphi_i + \frac{1}{2}A[(\nabla\varphi_i)^2 + (\nabla\alpha_i)^2] - MH\cos\varphi_i \right\} + \frac{1}{2}J_1\cos(\varphi_1-\varphi_2), \tag{21}$$

and the equations of motion have the form

$$-\frac{M}{\gamma}\dot{\alpha}_i + \frac{1}{2}K\sin(2\varphi_i) - A\nabla^2\varphi_i + MH\sin\varphi_i = \frac{1}{2}J_1\sin(\varphi_i-\varphi_{i\pm1}), \tag{22}$$

$$\frac{M}{\gamma}\dot{\varphi}_i + 4\pi M^2\alpha_i - A\nabla^2\alpha_i = 0, \quad i=1,2. \tag{23}$$

If the terms on the right-hand side of Eq. (22) are discarded, the equations obtained are similar to the equations of motion of a vertical Bloch line in a flat domain wall (see, for example, Ref. 17). Therefore, the approximation under consideration is conveniently called the Bloch line approximation.

In the case of a large value for the magnetization, Eqs. (22) and (23) can be reduced to a system of sine–Gordon equations. In fact, if the magnetostatic energy exceeds the anisotropy energy ($4\pi M^2/K \gg 1$), the values of α_i are small and the terms $\sim A\nabla^2\alpha_i$ can be neglected. Then

$$\alpha_i = -\frac{\dot{\varphi}_i}{4\pi M\gamma}. \tag{24}$$

Substituting (24) into (22), we obtain the system of equations

$$\Delta^2\left(\frac{1}{c_1^2}\ddot{\varphi}_i - \nabla^2\varphi_i\right) + \frac{1}{2}\sin(2\varphi_i) + \frac{MH}{K}\sin\varphi_i = \frac{J_1}{K}\sin(\varphi_i-\varphi_{i\pm1}), \tag{25}$$

where $c_1 = 2\gamma\sqrt{\pi A}$. The solutions of this system of equations was investigated in Ref. 18.

6. DISCUSSION OF RESULTS

In magnetic multilayers with a nonmagnetic intermediate layer of small thickness (of the order of several monolayers) the strength of the exchange interaction between the layers can amount to $J = J_1 d \sim 0.1\ \mathrm{erg/cm^2}$. Thus, for materials with a large value for the magnetization (like Fe or Co), i.e., with $M \geqslant 10^3$ G, we have $J_1 \ll 2\pi M^2$. In this case the vertical Bloch line approximation can be used to describe the magnetization dynamics. If, on the other hand, the thickness of the magnetic layers is small, the relation $K \ll 4\pi M^2$ also holds, and the system of coupled sine–Gordon equations (25) can be used to describe the magnetization dynamics. In this case the limiting velocity c_1 is determined by the strength of the exchange interaction within a layer and has a value $c_1 \sim 10^4$ cm/s for Fe and Co with $A \simeq 10^{-6}\ \mathrm{erg/cm^3}$.

In the case of magnetic layers with a small value for the magnetization ($M \leqslant 100$ G) and a strong exchange interaction between the layers ($J \sim 0.1\ \mathrm{erg/cm^2}$) the situation is reversed: $J_1 \gg 2\pi M^2$. The weak-ferromagnet approximation is then valid, and the dynamics of the magnetization in the spin-valve structure correspond to (16) and (17). Under these conditions the magnetization dynamics are described by a sine–Gordon equation. The expression for the limiting velocity c of a solitary wave can be represented in the form

$$c = c_1\sqrt{J_1/2\pi M^2}. \tag{26}$$

Since $J_1 \gg 2\pi M^2$, the limiting domain-wall velocity is several times greater in this case than in the case of dominance of the magnetostatic energy.

The ratio $J_1/2\pi M^2$ decreases both with increasing thickness of the nonmagnetic intermediate layer and with increasing thickness of the magnetic layers, and the condition for applicability of the weak-ferromagnet approximation is violated. Thus, the weak-ferromagnet approximation is valid only for thin layers with a large value for the exchange energy between the layers and a small value for the magnetization in the layers. It can be violated fairly easily. The vertical Bloch line approximation has a significantly broader range of applicability.

7. CONCLUSION

In this paper we have derived equations describing the nonlinear dynamics of the magnetization in a system of two magnetic layers. The system of equations obtained allows an exact static solution, which corresponds to a symmetric flat domain boundary. This solution is shown to be stable if the biquadratic exchange interaction between the layers is weak: $J_1 > 2J_2$. In the case of strong exchange coupling between

the layers the domain-wall dynamics are described by a sine–Gordon equation. If the magnetostatic energy dominates in the system, the dynamical behavior of the domain wall is reminiscent of a magnetic vortex, and the equations of motion are similar to the equations describing the dynamics of a vertical Bloch line in a domain wall.

This work was supported by the Russian Fund for Fundamental Research (Project 99-02-17830), INTAS (Project 97-705), and the ''Integration'' Federal Special-Purpose Program (Project K0573).

*)E-mail: vk@postoff.ics.ac.ru

[1] L. D. Landau and E. M. Lifshitz, Phys. Z. Sowjetunion **8**, 153 (1935).
[2] A. F. Andreev and V. I. Marchenko, Usp. Fiz. Nauk **130**, 39 (1980) [Sov. Phys. Usp. **23**, 21 (1980)].
[3] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, *Nonlinear Magnetization Waves: Dynamic and Topological Solitons* [in Russian], Naukova Dumka, Kiev (1983).
[4] V. G. Bar'yakhtar, M. V. Chetkin, B. A. Ivanov, and C. N. Gadetskii, *Dynamics of Topological Magnetic Solitons*, Springer-Verlag, Berlin–Heidelberg (1994).
[5] A. K. Zvezdin and V. V. Kostyuchenko, Fiz. Tverd. Tela (St. Petersburg) **39**, 178 (1997) [Phys. Solid State **39**, 155 (1997)].
[6] R. W. Wang, D. L. Mills, E. E. Fullerton, J. E. Matson, and S. D. Bader, Phys. Rev. Lett. **72**, 920 (1994).
[7] R. W. Wang and D. L. Mills, Phys. Rev. B **50**, 3931 (1994).
[8] L. Trallori, P. Politi, A. Pettori, M. G. Pini, and J. Villain, J. Phys.: Condens. Matter **7**, L451 (1995).
[9] C. Micheletti, R. B. Griffiths, and J. M. Yeomans, J. Phys. A: Math. Gen. **30**, L233 (1997).
[10] N. S. Almeida and D. L. Mills, Phys. Rev. B **52**, 13 504 (1995).
[11] V. V. Kostyuchenko and A. K. Zvezdin, Phys. Rev. B **57**, 5951 (1998).
[12] N. Papanicolaou, Phys. Rev. B **51**, 15 062 (1995).
[13] A. Schreyer, J. F. Ankner, T. Zeidler, H. Zabel, M. Schäfer, J. A. Wolf, P. Grünberg, and C. F. Majkrzak, Phys. Rev. B **52**, 16 066 (1995).
[14] A. K. Zvezdin, A. F. Popkov, and I. P. Yarema, Zh. Éksp. Teor. Fiz. **98**, 1070 (1990) [Sov. Phys. JETP **71**, 597 (1990)].
[15] A. Bishop, in *Solitons in Action*, K. Lonngren and A. Scott (Eds.), Academic Press, New York (1978) [Russ. transl., Mir, Moscow (1981), p. 72].
[16] A. K. Zvezdin, JETP Lett. **29**, 553 (1979).
[17] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Academic Press, New York (1979) [Russ. transl., Mir, Moscow (1982)].
[18] O. M. Braun, Yu. S. Kivshar, and A. M. Kosevich, J. Phys. C **21**, 3881 (1988)

Translated by P. Shelnitz

On the optimal void fraction in the thermodynamics of a simple liquid

V. S. Vorob'ev*)

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted 19 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1375–1385 (October 1999)

The free energy of a crystal containing a given void fraction is derived in terms of the truncated interparticle Lennard–Jones potential. The free energy is minimized over the void fraction at constant pressure and temperature. It is shown that for all pressures the optimal void fraction remains less than 1% as the temperature is raised. However, at some temperature it grows suddenly and reaches values of the order of the percolation level for voids in a crystal, ~ 0.125. At this point, the crystal transitions to the liquid state takes place. The derived dependence of the transition temperature on the pressure—the melting curve—is in good agreement with experimental data on the melting of solidified inert gases. © *1999 American Institute of Physics.* [S1063-7761(99)01810-7]

1. INTRODUCTION

At present it is commonly assumed[1–4] that the main reason for the difference between the liquid state and the crystalline state is the existence in a crystal of a definite fraction of voids. Thus, using molecular dynamics methods, the authors of Refs. 1 and 2 have investigated the size distribution of voids and their influence on the thermodynamic properties. The presence of voids leads to a decrease in the depth of the total potential energy curve and the number of nearest neighbors, and to an increase in the mean distance between nearest neighbors. Since voids can arrange themselves in various ways, an additional negative contribution to the free energy arises, the so-called configurational component. Under certain conditions, it can compensate the decrease in the depth of the potential well; thus, the state of a crystal with a definite void fraction becomes thermodynamically more favorable than the pure crystalline state.

On the basis of the equations of state of a liquid, based on the experimental data in Refs. 4–7, it became possible to show that the configurational component scaled to the temperature in the existence range of the liquid state is essentially constant and its value, depending on the data, lies in the range $0.7-0.8$, which is near $\ln 2$ (Refs. 6 and 7). This means that if the configurational component is generated by voids, then the void fraction in a liquid should vary only slightly. This contradicts the elementary theory of the formation of vacancies in a crystal,[8,9] according to which the vacancy fraction should depend exponentially on the temperature.

The present paper considers the thermodynamics of a crystal containing a definite void fraction. Calculations are performed for molecular crystals with a face-centered cubic lattice, interacting via a short-range potential. As is well known, in this case all thermodynamic quantities can be represented as dimensionless functions of the mass m, temperature T, depth of the intermolecular potential D, and equilibrium intermolecular distance R. As a consequence, the law of corresponding states applies. The class of materials satisfying these requirements is a broad one, and includes liquefied and solidified inert gases, condensed states of such gases as N_2, O_2, CO, CH_4, fullerite, etc. It is also important that there is a rich body of experimental information on the thermodynamic properties of this group of materials, which can serve as a test for theoretical models.

To start with, this paper considers the influence of voids on the basic characteristics of a crystal: the close packing parameter, the number of nearest neighbors, and the distance between them. An expression is then written down for the configurational component. Then, with the help of the truncated Lennard–Jones potential, an expression is written for the potential energy of the crystal in the presence of voids, and then used to derive an extrapolation formula for the maximum frequency of vibrations—the Debye temperature. The corresponding Grüneisen coefficient is found and thereby the contribution of the vibrations to the free energy. The free energy is minimized with respect to the void fraction at constant pressure and temperature. It is shown that at fixed pressure, up to some temperature its values remain less than 1%. The thermodynamic potential has a clearly defined minimum at this void fraction. Starting at certain temperatures, the void fraction grows abruptly and when a certain maximum temperature is reached it exhibits the opposite behavior. The thermodynamic potential as a function of the void fraction at this value of the temperature is monotonically decreasing, but has a horizontal plateau. A jump occurs in the void fraction here, from small values to values exceeding the percolation threshold. At this point the crystal melts. The optimal void fraction at the transition temperature depends weakly on the pressure, and its values are close to the percolation level. Accordingly, the configurational component of the free energy varies only weakly, and its values correspond to values obtained earlier on the basis of a treatment of the experimental data. The dependence of the transition temperature on the pressure gives the melting curve. The latter is found to be in good agreement with experimental data on the melting of solidified inert gases. It is also shown that the transition from the crystalline to the liquid

state becomes thermodynamically favorable at pressures $p \geqslant 0$ and temperatures $T > 0.58D$, in complete agreement with the experimental data.

2. CRYSTAL PARAMETERS IN THE PRESENCE OF VOIDS

Let some number of voids with relative fraction ν be present in a molecular crystal with a face-centered cubic lattice (fcc). As long as the number of voids is small and they are isolated from one another, their influence on the properties of the crystal is small. Qualitative changes arise when the void fraction reaches the percolation level. For a face-centered crystal the percolation level along bonds is given by the well-known formula[10]

$$\nu_c = \frac{d}{(d-1)Z} = 1/8, \tag{1}$$

where $Z=12$ is the number of nearest neighbors and d is the dimensionality. We will estimate some parameters of the crystal for this void fraction.

First and foremost, we consider the close packing parameter, defined as the fraction of the total volume of the crystal occupied by spheres centered on lattice sites and having radii equal to half the mean distance between nearest neighbors. As is well known, in a molecular crystal with an fcc lattice this parameter is $\phi_k = \pi/3\sqrt{2} = 0.74$. In the presence of voids (for the above void fraction) it is smaller: $\phi_l = \phi_k(1-\nu_c) = 7\pi/24\sqrt{2} \simeq 0.6479$. This latter value is very near the random close packing parameter, which, according to the data presented in Ref. 3, is $\phi_c = 0.644$. Many authors have noted that this packing parameter already characterizes the structural properties of a liquid.

In a fcc lattice, each atom has $Z=12$ nearest neighbors. In a crystal with voids, they are less numerous, with $Z_\nu = 12(1-\nu_c)$. For $\nu_c = 1/8$, we obtain $Z_\nu = 10.5$. This latter value of the number of nearest neighbors in the first coordination sphere is in good agreement with the values given, for example, in Ref. 11, based on the data of numerous experiments. Note that the value $(1-\nu_c)$ can be treated as the probability of finding a particle at a given point, and ν_c as the probability of finding a void.

The presence of voids also somewhat alters the distance between nearest neighbors in a crystal. The volume per particle increases by a factor of $1/(1-\nu_c)$. However, in the calculation of the mean distance between nearest neighbors it is necessary to take into account the close-range order, which is preserved in a system with voids. The increase in this distance may be associated with a decrease in the number of nearest neighbors in the second coordination sphere. In a fcc lattice there are six such neighbors, and in the shell of second neighbors, $6\nu_c$ voids arise. This leads to an increase in the relative volume per nearest neighbor by

$$g(\nu) = 1 + \frac{6\nu}{12} = 1 + \frac{\nu}{2}. \tag{2}$$

For $\nu_c = 1/8$, formula (2) yields $g(1/8) \simeq 1.063$. The specific volumes of the liquid and solid phases of Ar and Ne near zero temperature have a similar relationship. Information about these quantities was obtained in Ref. 7 by processing a set of experimental data.

The arrangement of the voids throughout the crystal can be realized in any of a number of ways. The configurational component of the free energy is proportional in the usual way to the logarithm of that number, and serves as a measure of disorder in the system. We write it as

$$F_c = -kT \ln \frac{[N(1+\nu)]!(2Z)^{N\nu}}{N!(N\nu)!}, \tag{3}$$

where N is the number of particles, $2Z=24$ is the statistical weight of a void, equal to the number of sites that the 12 nearest neighbors surrounding the void can occupy (each neighbor can be found in two positions).

Using Stirling's formula, we obtain

$$F_c = -TN \ln \left[\frac{(1+\nu)^{1+\nu} 24^\nu}{\nu^\nu} \right] = -TN \ln s(\nu). \tag{4}$$

Note that $s(1/8) \simeq 2.166$ and $\ln s \simeq 0.77$. Values of the configurational component close to this value were obtained by various arguments in Refs. 4–7.

Thus, when the void fraction in a crystal reaches the percolation level, we obtain many parameters already typical of the liquid state. We now proceed now to a more consistent construction of the thermodynamics of a crystal with voids.

3. FREE ENERGY OF A SYSTEM WITH VOIDS

Let the particles interact via a Lennard–Jones potential. We take into account only the interaction with nearest neighbors. As the analysis in Ref. 11 shows, such an approximation effectively takes into account the short-range nature of the potential and leads to better agreement with experiment. Thus we write the potential energy per unit mass in the presence of voids as

$$U = \frac{Z_\nu D}{m} \left[\frac{1}{2} \left(\frac{v_0 g}{v} \right)^4 - \left(\frac{v_0 g}{v} \right)^2 \right], \tag{5}$$

where v is the specific volume and v_0 is the specific volume at zero temperature and pressure. This quantity can be expressed in terms of the parameters of the potential: $v_0 = R^3/m\sqrt{2}$.

As can be seen from formula (5), we have taken into account the decrease in the number of nearest neighbors and the increase in the distance between them. In a crystal without voids, $Z_\nu = 12$ and $g=1$, and formula (5) goes over to the usual expression for the potential energy of a crystal with a fcc lattice, given for example in Ref. 9. The determination of the functions Z_ν and g presents a problem. On the basis of arguments presented in the preceding section, we assign these functions in the form

$$Z_\nu = 12e^{-a\nu} = 12f, \tag{6}$$

$$g(\nu) = (1+\nu)/(1+a_1\nu), \tag{7}$$

where a and a_1 are adjustable parameters determined below.

We take account of vibrations of the atoms about their equilibrium positions in the Debye approximation. Then the free energy per unit mass can be written in the form

$$F=\frac{T}{m}\left[3\ln(1-e^{-\Theta/T})-\mathscr{D}\left(\frac{\Theta}{T}\right)\right]-\frac{T}{m}s+U, \tag{8}$$

where Θ is the Debye temperature and $\mathscr{D}(x)$ is the Debye function.

For small specific volumes, the maximum frequency of vibrations, or equivalently the characteristic (Debye) temperature Θ, is related to the potential energy:

$$\Theta=h\left(\frac{3}{4\pi m v}\right)^{1/3} f\sqrt{\frac{v^2}{m}\frac{d^2U}{dv^2}}\,. \tag{9}$$

Substituting the potential (5) in formula (9), we obtain

$$\Theta=h\left(\frac{3}{4\pi m v}\right)^{1/3} f^{3/2}\sqrt{24\frac{D}{m}\left[5\left(\frac{v_0 g}{v}\right)^4-3\left(\frac{v_0 g}{v}\right)^2\right]}\,. \tag{10}$$

Formula (10) is only applicable to small v, and does not possess the necessary extrapolation properties. In the liquid state, the radicand in formula (10) vanishes, or even becomes negative. In addition, it should decrease but remain finite. Taking this into account below, we extrapolate the radicand as follows:

$$5\left(\frac{v_0 g}{v}\right)^4\frac{1}{1+(3a_2/5)(v/v_0 g)^2}=5\left(\frac{v_0 g}{v}\right)^4 f_1, \tag{11}$$

where a_2 is some constant.

On the other hand, the limiting expression for the Debye temperature[7] that ensures that the free energy (7) goes over at large v and high temperatures to the corresponding expression for the free energy of an ideal gas is well known. This can serve as a basis for the following interpolation formula for the characteristic temperature:

$$\Theta=\left(\frac{3}{4\pi}\right)^{1/3}\frac{h}{(mv)^{1/3}\sqrt{m}}\left\{\left(\frac{4\pi s(\nu)}{3}\right)^{1/3}\sqrt{\frac{T}{2\pi}}+f^{3/2}\sqrt{120D\left(\frac{v_0 g}{v}\right)^4 f_1}\right\}. \tag{12}$$

In the vicinity of the condensed state, the temperature term in formula (12) is negligible, and formula (12) yields results similar to formula (10). As the volume increases, the situation reverses, and the temperature term becomes the dominant term. The parameter $\Theta/T\ll 1$, the potential energy in expression (8), can then be neglected, and this expression becomes

$$F=\frac{T}{m}\left[\ln\left(\frac{\Theta}{T}\right)^3-1\right]-\frac{T}{m}\ln s(\nu). \tag{13}$$

Retaining only the temperature term, we see that expression (13) corresponds to the free energy of an ideal gas with all the necessary constants.

It is convenient to transform to dimensionless variables. We express the energies and temperature in terms of the potential depth, i.e., $F'=F/D$, $U'=U/D$, $T'=T/D$, etc., and for the specific volumes we introduce the quantity $v'=v/v_0$ (below we drop the primes). Then the Debye temperature takes the form

$$\Theta=B(1/v)^{1/3}\left\{\left(\frac{4\pi s(\nu)}{3}\right)^{1/3}\sqrt{\frac{T}{2\pi}}+\frac{f^{3/2}g^2}{v^2}\sqrt{120f_1}\right\}, \tag{14}$$

where

$$B=\left(\frac{4\cdot 2^{1/3}}{3\pi}\right)^{1/3}\frac{h}{R\sqrt{Dm}} \tag{15}$$

is the de Bour parameter. For argon $R=3.76$ Å, $D=12.2$ meV (Ref. 11), and the de Bour parameter $B=0.107$.

Analysis of expression (14) shows that as long as the void fraction is small, the characteristic temperature depends on it weakly. In what follows, we neglect this dependence. Note that the weak sensitivity of the characteristic temperature to the void fraction confirms its conjectured[6,7] continuous behavior in the transition from the solid to the liquid state.

Thus, we have constructed an expression for the free energy of a crystal containing voids. This energy is expressed in terms of the parameters of the interparticle interaction potential. We have modeled the dependence of the interaction potential and the characteristic temperature on the void fraction. These dependences contain three variable parameters.

From the free energy it is possible to obtain expressions for various thermodynamic functions. Thus, the pressure in dimensionless units has the form

$$p=\frac{3T\Gamma(v,T)}{v}\mathscr{D}\left(\frac{\Theta}{T}\right)+\frac{24f}{v}\left[\left(\frac{g}{v}\right)^4-\left(\frac{g}{v}\right)^2\right]. \tag{16}$$

The quantity Γ in formula (16) is the Grüneisen coefficient, defined by

$$\Gamma(v,T)=-\frac{\partial\ln\Theta(v,T)}{\partial\ln v}. \tag{17}$$

The Grüneisen coefficient, defined in this way, depends on the specific volume and the temperature. Figure 1 plots

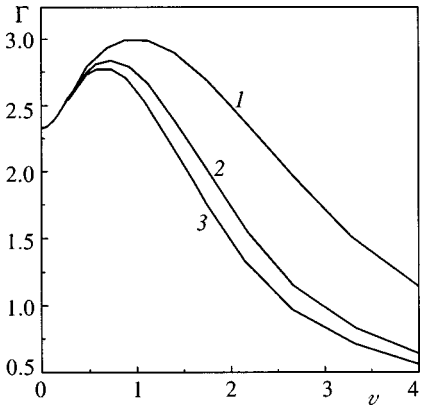


FIG. 1. Dependence of the Grüneisen coefficient on the specific volume at various temperatures: curve *1* corresponds to $T=0.1$, *2* — 1, *3* — 2.

the dependence of the Grüneisen coefficient on the volume for three values of the temperature. These dependences are in line with the existing picture of the behavior of the Grüneisen coefficient in the low-density regime. It follows from formulas (14) and (17) that for small ν the Grüneisen coefficient approaches 7/3. For small ν, regardless of the temperature, the Grüneisen coefficient $\Gamma \to 1/3$, which ensures that formula (16) goes over to the pressure of an ideal gas. Numerical values of this quantity near the maximum are $\Gamma = 2.4-2.9$ and are found to be in agreement with experiment and with calculations relying on other methods.[6,7]

4. DETERMINATION OF THE OPTIMAL VOID FRACTION

The void fraction enters into the free energy (8) as a parameter. The equilibrium state is determined by minimizing the thermodynamic potential $G(\nu,T,\nu)$ with respect to that parameter at constant pressure and temperature

$$\left(\frac{\partial G}{\partial \nu}\right)_{p,T}=0. \tag{18}$$

Using the usual thermodynamic relation $G=F+p\nu$ and assuming constant pressure, we find that this expression is equivalent to

$$\left(\frac{\partial F}{\partial \nu}\right)_{\nu,T}=0. \tag{19}$$

The derivative in Eq. (18) can be calculated explicitly:

$$\left(\frac{\partial F}{\partial \nu}\right)_{\nu,T}=-TL(\nu)+Q(\nu,\nu)=0, \tag{20}$$

where

$$L(\nu)=\ln\left(\frac{24(1+\nu)}{\nu}\right), \tag{21}$$

$$Q(\nu,\nu)=12f\left[\varphi_1\left(\frac{g}{\nu}\right)^4-\varphi_2\left(\frac{g}{\nu}\right)^2\right], \tag{22}$$

$$\varphi_1=2\frac{d\ln g(\nu)}{d\nu}-\frac{a}{2},\quad \varphi_2=2\frac{d\ln g(\nu)}{d\nu}-a. \tag{23}$$

The two equations (20) and (16) contain three unknowns: T, ν, and ν. Starting with Eq. (20), we obtain an explicit expression for the temperature:

$$T=\frac{Q(\nu,\nu)}{L(\nu)}. \tag{24}$$

Substituting this expression into the formula for the pressure (16), we obtain an equation containing only ν and ν (at given p). This latter equation was solved numerically.

5. RESULTS OF CALCULATIONS AND THEIR DISCUSSION

A liquid can exist as a stable phase at temperatures and pressures exceeding the corresponding values at the triple point. For the substances considered here, the characteristic unit of pressure, expressed in terms of the parameters of the potential, is D/R^3, which for argon, for example, is 367 atm. For argon the pressure at the triple point is $p_t=0.69$ atm, and the corresponding dimensionless pressure is very small: $\sim 1.9\times 10^{-3}$. The temperature at the triple point is well known to be $T_t=0.58D$.

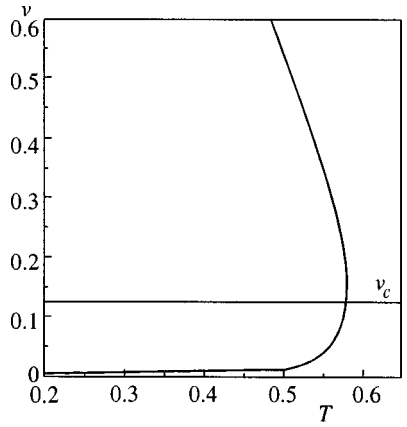


FIG. 2. Temperature dependence of the optimal void fraction at $p=0$, ν_c is the void fraction corresponding to the percolation level.

We first consider the temperature dependence of the optimal number of voids at zero pressure (Fig. 2). This dependence has a temperature maximum at $T\simeq 0.58$, $\nu=0.125$. The curve consists of two branches, with positive and negative slope, respectively. It is easy to show that the lower branch corresponds to states of stable equilibrium (minima of G), and the upper—to states of unstable equilibrium (maxima of G). At low temperatures the void fraction is small, and up to temperatures of order 0.5 essentially remains less than 0.01. However, as the temperature approaches its maximum, it rises abruptly, and at its maximum the void fraction reaches $\nu_m\simeq 0.125$. The maximum temperature corresponds to the experimental value of the temperature at the triple point.

It was possible to achieve this accuracy by specially choosing the values of the adjustable parameters in our model. The results plotted in Fig. 2 were obtained for $a=0.94$, $a_1=0.484$, and $a_2=7$. Recall that the first two are associated with the dependence of the parameters of the potential on the void fraction. Note that the values of these two constants differ only slightly from the values used in Sec. 1 ($a=1$ and $a_1=0.5$). The third constant determines the rate of falloff of the characteristic temperature, and affects the behavior of the Grüneisen coefficient, which we discussed in connection with Fig. 1.

To understand how the thermodynamic potential behaves as a function of the void fraction at $p=0$ and various temperatures, the appropriate curves were constructed and are plotted in Fig. 3. The upper curve corresponds to $T=0.51$. As can be seen, for a small void fraction the curve has a minimum, and with increasing void fraction, a maximum appears, in complete agreement with Fig. 2. A crystal with such a small void fraction differs only slightly in its properties from a solid. Here the equilibrium void fraction is at most 1%. Near the minimum of the thermodynamic potential the crystal is in a state of stable equilibrium. If it is somehow induced to leave that state, it returns to the equilibrium state after the influence is removed. Consequently, it

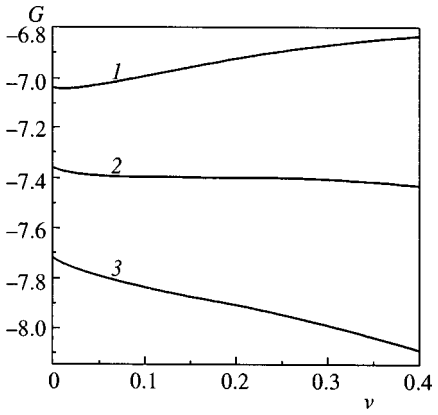


FIG. 3. Isotherms of the free energy at $T=0.58$ for various void fractions for $p=0$ for various temperatures: curve *1* corresponds to $T=0.51$, *2* — 0.58, *3* — 0.65.

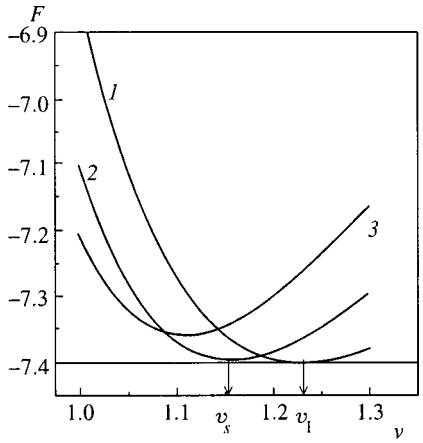


FIG. 4. Isotherms of the free energy at $T=0.58$ for various void fractions: curve *1* corresponds to $\nu=0.15$, *2* — 0.05, *3* — 0. The common horizontal tangent to curves *1* and *2* is drawn, corresponding to $p=0$. Equilibrium volumes of the solid (ν_s) and liquid (ν_l) phases are indicated.

is possible to speak of the elastic properties of the crystal. But such a reaction to an external influence will take place only so long as the action is not so strong as to reach the maximum of G. If the latter occurs, it is possible to speak of a loss of elastic properties and to say that the flow threshold has been reached. When the influence is removed, the void fraction can then even increase.

At $T=0.58$, corresponding to the maximum in Fig. 2, the curve with a minimum and a maximum degenerates into a monotonically falling curve with a horizontal plateau starting at $\nu\simeq 0.05$ and extending to $\nu\simeq 0.23$. It can be said that the thermodynamic potential is in a state of neutral equilibrium with respect to the void fraction, whereupon the maximally disordered state—i.e., the state with the largest void fraction—is thermodynamically more favorable, since that state has maximum entropy. This means that a transition takes place at that temperature from a state with a small void fraction (less than the percolation threshold) to a state with a large void fraction (greater than the percolation threshold), and the crystal melts.

Figure 3 also displays a curve with $T=0.65$. In this case there is no minimum, which is also consistent with the results shown in Fig. 2.

Figure 4 plots isotherms of the free energy at $T=0.58$ for various void fractions. As can be seen, the minimum in the curve with $\nu=0$ lies somewhat above the remaining curves. Starting at $\nu\simeq 0.05$ and extending up to $\nu\simeq 0.15$, the minima lie at the same level. The common tangent to these isotherms, as is well known, determines the phase equilibrium parameters (specific phase volumes and pressure). In the given case, it has zero slope and corresponds to $p=0$. This means that at $p=0$ and $T=0.58$ a transition takes place from a state with a small number of voids ($\nu\simeq 0.05$) corresponding to the solid state, to a state with a large number of voids ($\nu\simeq 0.15$). The latter already corresponds to a liquid. Such a transition is thermodynamically favored: for identical values of the free energy minimum (or equality of the thermodynamic potentials), the higher-entropy phase will be favored. Of course the entropy increases with the number of voids. The values of the relative phase volumes at the tran-

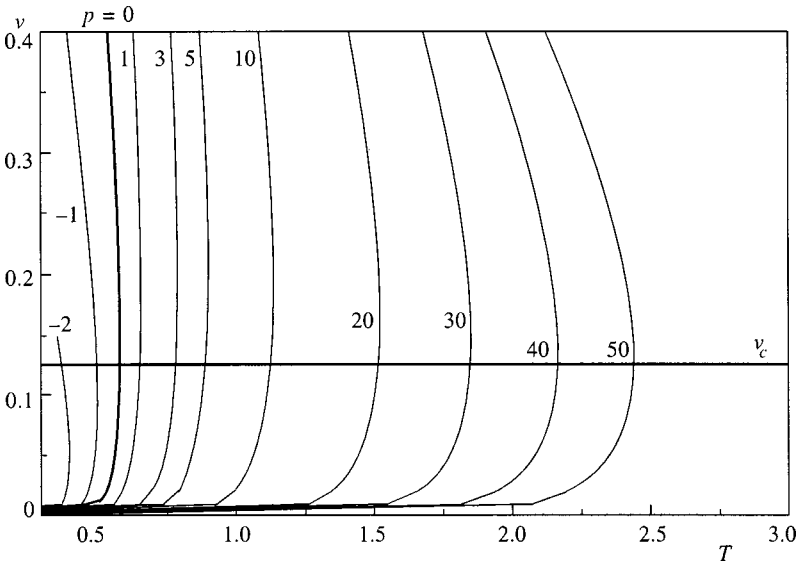


FIG. 5. Temperature dependence of the optimal void fraction at various pressures; ν_c is the void fraction corresponding to the percolation level.

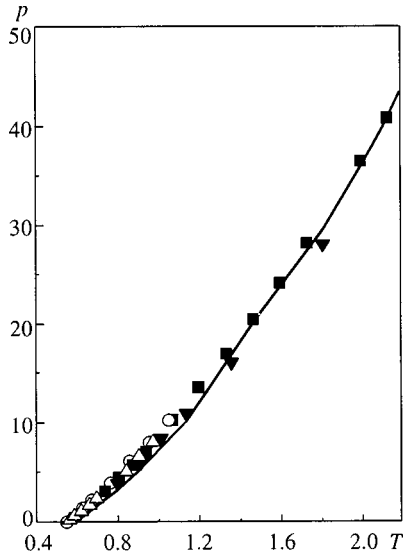


FIG. 6. Temperature dependence of the melting pressure (melting curve). The solid curve was derived in the present work from the points corresponding to the temperature maxima in Fig. 5. Experimental data: ■ — Ar, ○ — Kr, ▼ — Ne, △ — Xe.

sition point are easily found from the plot in Fig. 4. The common horizontal tangent to the isotherms touches the $\nu \simeq 0.05$ curve at $v_s \simeq 1.15$, and the $\nu \simeq 0.15$ curve at $v_1 \simeq 1.23$. The corresponding experimental values, taken from Ref. 12, are roughly 1.1 and 1.24. It can be seen at once that we have good agreement with the experimental data.

We next consider how the optimal void fraction varies with pressure. Figure 5 plots calculated values of the void fraction as a function of temperature at various pressures. As can be seen, their form is similar to the form for $p=0$. The void fraction corresponding to the maximum temperature at a given pressure varies very little, and corresponds roughly to the percolation level. It follows from the foregoing that the points of maximum temperature at given pressure on these curves correspond to points on the melting curve. Therefore, by constructing the dependence of the pressure on the maximum temperature, we should obtain the melting curve.

The constructed curve is plotted in Fig. 6, where we have also plotted experimental data[12] on the melting of solidified inert gases. As can be seen, good agreement is obtained.

I am grateful to B. M. Smirnov for stimulating discussions.

This work was supported by the Russian Fund for Fundamental Research (Project No. 99-02-16596).

*)E-mail: vsv@rinet.ru

[1] F. H. Stillinger, Phys. Rev. E **52**, 4685 (1995).
[2] D. S. Corti, P. G. Debenedetti, S. Sastry, and F. H. Stillinger, Phys. Rev. E **55**, 5522 (1997).
[3] M. D. Rintoul and S. Torquato, Phys. Rev. E **58**, 532 (1998).
[4] D. C. Wallace, Phys. Rev. E **56**, 4971 (1996).
[5] S. M. Stishov, Usp. Fiz. Nauk **114**, 3 (1974) [Sov. Phys. Usp. **17**, 625 (1974)].
[6] V. S. Vorob'ev, JETP Lett. **62**, 574 (1995).
[7] V. S. Vorob'ev, Zh. Éksp. Teor. Fiz. **110**, 683 (1996) [JETP **83**, 368 (1996)]; Teplofiz. Vys. Temp. **34**, 397 (1996).
[8] V. N. Zharkov, V. A. Kalinin, *Equations of State of Solids at High Pressures and Temperatures* [in Russian], Nauka, Moscow (1967).
[9] C. Kittel, *Elementary Solid State Physics*, 5th ed., Wiley, New York (1976).
[10] J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems*, Cambridge University Press, Cambridge (1979), Moskva (1982).
[11] B. M. Smirnov, Usp. Fiz. Nauk **114**, 1165 (1994).
[12] R. K. Crawford, in *Rare Gas Solids*, M. L. Klein and J. A. Venables (eds.), Academic Press, New York (1977), Vol. 2, p. 663.

Translated by Paul F. Schippnick

Existence criteria for Landau–Pekar polarons

É. N. Myasnikov and A. É. Myasnikova*)

Rostov State Teachers' Training University, 344082 Rostov-on-Don, Russia
(Submitted 16 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 1386–1397 (October 1999)

We build the distribution function for a system with coexisting self-localized and delocalized fermions. The distribution function is used to study the behavior of the chemical potential of the carriers in such a system, which is found to differ substantially from the behavior of the chemical potential in a system of delocalized fermions. We also find that as the temperature changes, isostructural first-order phase transitions can emerge in the system of self-localized and delocalized fermions. These transitions, for which changes in the state of the macroscopic number of particles are responsible, manifest themselves in the electrical conductivity, in the contribution of carriers to the specific heat, and in the optical properties of such systems. Formulas are derived that approximate the dependence of the temperature of such a phase transition on the binding energy of the self-localized states of carriers and on the maximum group velocity of phonons participating in the formation of such states. Finally, we show that the special features of the behavior of the chemical potential of the carriers in a system with carrier self-localization lead to the possibility of Bose condensation in a system where bipolaron states are metastable. © *1999 American Institute of Physics.* [S1063-7761(99)01910-1]

1. INTRODUCTION

The problem of electrical conduction in substances with strong Fröhlich interaction between charge carriers and phonons was formulated by Landau and Pekar,[1] who were the first to predict spontaneous violation of the spatial homogeneity of a system accompanied by formation of a large polaron. Their discovery started a new avenue of research in solid-state physics, the study of polarons in semiconductors. However, in many ways the predictions of large-polaron theory were found to differ from the experimental results (e.g., the presence of large polarons was not detected in alkali–halide crystals, while in other substances such polarons were observed only at low temperatures, much lower than the binding energy $E_{\rm pol}$. This was the reason for a skeptical attitude toward large-polaron theory. The reasons for the discrepancies between the theoretical predictions and the experimental data were understood only after spatial dispersion of lattice polarizability had been taken into account.[2]

In this theory, the dispersion of phonons interacting with a carrier is taken into account by the formula $\Omega^2(k)=\Omega^2(0)+u^2k^2$, where u is the minimum phase velocity of phonons and, at the same time, the maximum group velocity of phonons, and it is shown that the band of self-localized states of charge carriers is strongly limited in momenta, in view of which the average velocity in a carrier in self-localized states cannot exceed the maximum group velocity u of the phonons participating in the formation of such states. As a result, the energy width ΔE of this band is much less than $E_{\rm pol}$ (see Ref. 3). Hence polarons cannot have large values of thermal energy, so they cannot exist at temperatures comparable to the polaron binding energy $E_{\rm pol}$.

If in addition to allowing for the spatial dispersion of the lattice polarizability we allow for the finiteness of the volume V_0 within which a carrier is localized in a polaron, we conclude that there exists a limit on the polaron density in the system, i.e., a limit at which the polaron density is the greatest. The volume V_0 can be used to determine the maximum value p_0 of the momentum of a carrier in a polaron by the formula $(4/3)\pi p_0^3=(2\pi\hbar)^3/V_0$, with $p_0 \gg m^*u$ in all substances with strong electron–phonon interaction (here m^* is the effective carrier mass). Hence according to the Pauli exclusion principle, a region of space occupied by a polaron can contain another polaron only if the latter has an average carrier momentum greater than p_0. However, according to what we have just said, the average momentum of a carrier in a polaron cannot be greater than $m^*u \ll p_0$. Therefore, if we allow for carrier spin, the polaron density in the system cannot exceed $2V_0^{-1}$.

The possibility of self-localization also constrains the filling of the delocalized states of carriers. Delocalized carriers with momenta $p<m^*u$ cannot exist even in the absence of polarons in the system, since the time they need to transform into self-localized states is Ω^{-1}, where Ω is the frequency of the phonons with which a carrier interacts and hence becomes self-localized. In other words, in a system with strong electron–phonon interaction such carriers would decay very rapidly (the time Ω^{-1} is of order 10^{-13} s).

The Pauli exclusion principle also suggests that a system with a polaron density $2V_0^{-1}$ can have no delocalized carriers with momenta $p<p_0$, since the entire region with $p<p_0$ of the single-particle phase space (for the carriers) is already occupied by self-localized carriers. As the polaron density in the system tends to zero, the number of possible delocalized states of carriers with $p<p_0$ tends to $2VV_0^{-1}$, where V is the volume occupied by the system. Thus, we have an extremely

unusual situation: the number of states of one kind available for particles depends on the number of states of the other kind occupied by such particles. It is convenient to talk of carriers with $p<p_0$ as ''cold'' carriers and of carriers with $p>p_0$, as ''hot'' carriers. Then the concentration $n_0=2V_0^{-1}$ can be called the maximum concentration of cold carriers. On the other hand, the possible existence of any hot carrier is in no way limited by the occupancy of other states.

These properties of the distribution of the carriers over the states must be properly reflected in the distribution function. Until now, no such distribution function has been known. In the present paper we set up such a distribution function and analyze its properties. Also, by using the inertial mass tensor of a polaron[4] we show that this distribution function makes it possible to study the kinetic properties of systems of self-localized and delocalized fermions.

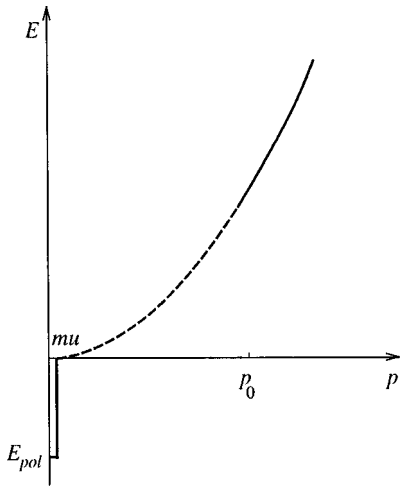


FIG. 1. Dispersion relation for the charge carriers in a medium with one-phonon self-localization of the carriers. The dashed curve indicates the range of states that are occupied if the concentration of self-localized carriers is less than the maximum concentration n_0 of the carriers.

2. DISTRIBUTION FUNCTION FOR CARRIERS IN A SYSTEM WITH ONE-PHONON SELF-LOCALIZATION OF THE CARRIERS

Let the carriers in the system become self-localized as a result of the interaction with one optical phonon branch. Since $p_0 \gg m^*u$, any polaron velocity $v<u$ has essentially no effect on the region of the phase space occupied by the carriers of this polaron. On the other hand, the possibility of different localizations of a polaron in a crystal occupying a volume V is the reason that there are $2VV_0^{-1}$ different polaron states of the carriers (irrespective of the average carrier momentum) in this crystal. Hence, in analyzing the statistical properties of carriers with average momenta $p<m^*u$ in a system with self-localization, one must take the set of states with fixed localization as the basis. In such states the carrier momentum and hence the energy will be uncertain to some extent. To set up the distribution function we employ the Gibbs method, since it has proved its usefulness in deriving the classical Fermi–Dirac and Bose–Einstein distributions.[5] However, the use of a basis that incorporates states with definite localization in constructing the Gibbs distribution function requires a modification of the method, since in its traditional form every state has a well-defined energy.

We examine the properties of systems in which self-localized and delocalized carriers coexist. The dependence of the carrier energy in such a system on the carrier's average momentum[3,6] is depicted in Fig. 1. As noted earlier, if the polarons in such a system occupy all $2VV_0^{-1}$ states, only delocalized carrier states with $p>p_0$ can be filled. In the intermediate case, cold carriers with momenta $p<p_0$ are partially in the self-localized states and partially in the delocalized states. The same can be said of the states of a cold fermion in volume V_0. In the limit, this volume can be occupied by a single self-localized fermion or by a single delocalized fermion or, in the intermediate case, by a partially localized and partially delocalized fermion. Two alternatives are possible here: either there are no cold fermions in the volume V_0 or only one cold fermion occupies this volume. We examine a volume V_0 that contains N cold fermions as a subsystem with a variable number of particles by the Gibbs method. Then the contribution of the first alternative ($N=N_1+N_2=0$) to the normalization condition has the form

$$\frac{\exp[(\Omega-\mu N)/T]}{4\pi p_0^3/3}\left[\int_0^{mu}4\pi p^2dp\,\exp\left(-\frac{E_1}{T}N_1\right)+\int_{mu}^{p_0}4\pi p^2dp\,\exp\left(-\frac{E_2}{T}N_2\right)\right], \tag{1}$$

where N_1 is the number of self-localized fermions, $N_1=0$, and N_2 is the number of delocalized cold fermions, $N_2=0$; $E_1(p)$ is the energy of the self-localized state in which the average carrier momentum is p (here the expression for $E_1(p)$ contains the polaron's effective ''energy'' mass, which in turn is a function of p); and $E_2(p)$ is the energy of a cold delocalized carrier with momentum p, with $E_2(p)=E_{\rm pol}+p^2/2m$. Here $E_{\rm pol}$ is the binding energy of a polaron, and the zero level of the chemical potential coincides with the bottom of the band of self-localized fermions.

When examining the case with $N=1$, we must bear in mind that if states of one type are fully occupied, states of the other type are inaccessible. Then the contribution of this case to the normalization condition is

$$\frac{\exp[(\Omega-\mu N)/T]}{4\pi p_0^3/3}\int_0^1 dN_2\left[(1-N_2)\int_0^{mu}4\pi p^2dp\times\exp\left(-\frac{E_1}{T}\right)+N_2\int_{mu}^{p_0}4\pi p^2dp\,\exp\left(-\frac{E_2}{T}\right)\right], \tag{2}$$

where the possibility of two limiting states and all intermediate states of cold fermions is taken into account by the integral with respect to N_2 from 0 to 1. Thus, the normalization condition has the form

$$\exp\left(\frac{\Omega}{T}\right)\left\{1+\frac{1}{2}\,\frac{\exp(\mu/T)}{4\pi p_0^3/3}\left[\int_0^{mu}4\pi p^2dp\,\exp\left(-\frac{E_1}{T}\right)+\int_{mu}^{p_0}4\pi p^2dp\,\exp\left(-\frac{E_2}{T}\right)\right]\right\}=1, \tag{3}$$

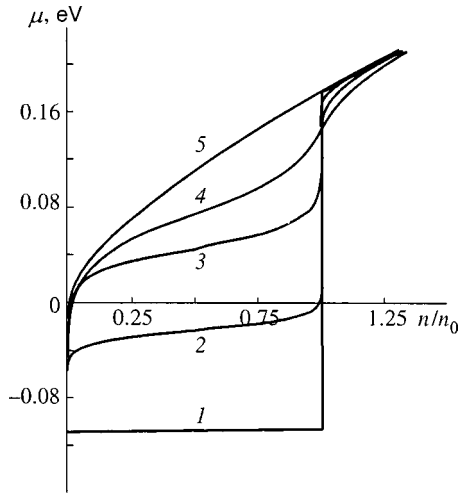


FIG. 2. Chemical potential μ of a system of self-localized and delocalized fermions as a function of the total fermion concentration n/n_0 for different temperatures: curve *1*, $T=0.001$ eV; curve *2*, $T=0.005$ eV; curve *3*, $T=0.01$ eV; and curve *4*, $T=0.02$ eV. Curve *5* demonstrates the concentration dependence of the chemical potential of a system of delocalized fermions at $T=0$ K.

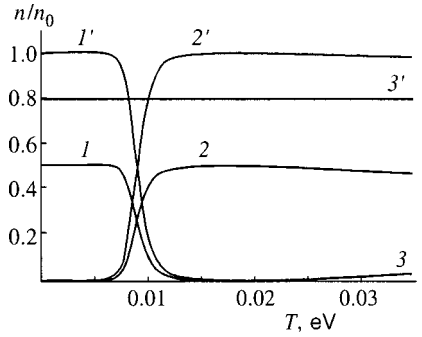


FIG. 3. Temperature dependence of the fermion concentration in each of the states that the fermions can form for two values of the total fermion concentration n. Curves *1* and *1′* correspond to polaron concentrations at $n=0.5n_0$ and $n=1.8n_0$, respectively; curves *2* and *2′* show the behavior of the concentration of cold delocalized carriers at $n=0.5n_0$ and $n=1.8n_0$, respectively; and curves *3* and *3′* show the behavior of the concentration of hot fermions.

which yields the following expression for the average number of cold fermions inside the volume V_0:

$$\overline{N}_{\rm cold}=-2\frac{\partial\Omega}{\partial\mu}=\frac{A\exp(\mu/T)}{1+(1/2)A\exp(\mu/T)}, \tag{4}$$

where

$$A=\frac{V_0}{(2\pi\hbar)^3}\left[\int_0^{mu}4\pi p^2dp\exp\left(-\frac{E_1}{T}\right)+\int_{mu}^{p_0}4\pi p^2dp\exp\left(-\frac{E_2}{T}\right)\right].$$

For the average number of hot fermions inside the volume V_0 we have

$$\overline{N}_{\rm hot}=\frac{2V_0}{(2\pi\hbar)^3}\int_{p_0}^{\infty}\frac{4\pi p^2dp}{\exp[(E_2(p)-\mu)/T]+1}, \tag{5}$$

which follows from the ordinary Fermi–Dirac distribution. Obviously, the total concentration of fermions in the system is the sum of the concentrations of cold and hot carriers:

$$n=\overline{N}_{\rm cold}/V_0+\overline{N}_{\rm hot}/V_0, \tag{6}$$

which leads to a formula for the chemical potential of the fermions as a function of their concentration n and temperature T.

3. PROPERTIES OF THE DISTRIBUTION FUNCTION OF THE CARRIERS IN SYSTEMS WITH ONE-PHONON SELF-LOCALIZATION

Figure 2 depicts the dependence, obtained by using Eqs. (4)–(6), of the chemical potential μ of a system of fermions on the fermion concentration n for several temperatures. Here fermion concentration is measured in units of the maximum concentration of cold fermions, n_0 (this quantity is also the maximum concentration of polarons), which is weakly dependent on the parameters of the medium (just as the polaron volume is). A typical value is $n_0=3\times10^{20}\,{\rm cm}^{-3}$, which can be obtained with the following values of the parameters of the medium used in plotting the diagrams: the reciprocal effective dielectric constant $c=1/\varepsilon_\infty-1/\varepsilon_0=0.27$, the maximum group velocity of the phonons $u=10^5\,{\rm cm\cdot s}^{-1}$, and the effective fermion mass in the crystal $m^*=m_e$. Figure 2 shows that when $n>n_0$, the chemical potentials of systems of free fermions and of systems where free and self-localized fermions can coexist coincide. However, when $n<n_0$, the chemical potential of the system in which self-localization of carriers is possible increases with temperature, while in a system of delocalized fermions the chemical potential always decreases with increasing temperature.

Another characteristic feature of systems where self-localized fermions may be present is the fact that the chemical potential tends to zero as $T\to0$ K, irrespective of the value of the fermion concentration n, provided that $n<n_0$. Figure 2 shows that at carrier concentrations $n<n_0$ and at temperatures below room temperature (the assumption adopted in Fig. 2), the chemical potential of the system is primarily determined by the temperature and is essentially independent of the concentration. The reason, of course, is that at such concentrations ($n<n_0$) and temperatures the band that becomes chiefly populated is the polaron band, and the number of polarons in a state with the same momentum may become macroscopic. This number is the greatest (and equal to $2VV_0^{-1}$) at $T=0$ K, with all the polarons having zero momentum and minimum energy. In view of the macroscopic nature of such degeneracy, we call it condensation of self-localized fermions in the state with the lowest energy.

Figure 3 demonstrates the redistribution of fermions among the self-localized and free states induced by temperature variation for carrier densities $n=0.5n_0$ and $n=1.8n_0$, respectively (the parameters of the medium are the same as in Fig. 2). We see that even at 50–100 K, which is much less than the polaron binding energy (for the adopted values of the parameters, the binding energy $E_{\rm pol}=0.11$ eV), self-localized fermions disappear from the system, transforming

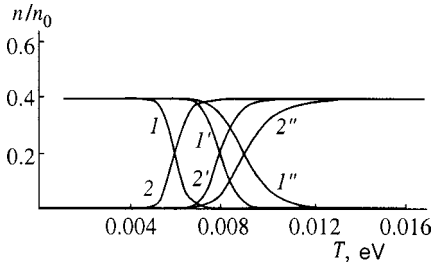


FIG. 4. Variation of the concentration of polarons and cold delocalized fermions with temperature for different values of the polaron binding energy $E_{\rm pol}$ and the maximum group velocity of phonons, u. Curves *1* and *2* correspond to the concentration of carriers in the polaron and cold delocalized states for $u=10^4\ {\rm cm\cdot s^{-1}}$ and $E_{\rm pol}=0.11$ eV; curves *1′* and *2′*, at $u=10^4\ {\rm cm\cdot s^{-1}}$ and $E_{\rm pol}=0.15$ eV; and curves *1″* and *2″*, at $u=10^5\ {\rm cm\cdot s^{-1}}$ and $E_{\rm pol}=0.11$ eV.

into delocalized cold state. Here the temperature at which polarons disappear is independent of the carrier concentration in the system and is determined solely by the polaron binding energy and the value of the maximum group velocity of the phonons participating in the formation of the polaron.

Figure 3 also shows that below room temperature hot fermions appear in the system only if the carrier concentration is greater than the maximum concentration n_0 of cold fermions. This is important since only hot phonons can participate in the screening of the electron–phonon interaction by charge carriers. In particular, this means that the influence of screening effects, which (as shown by Iadonisi *et al.*[7]) substantially changes the binding energy of self-localized states even at concentrations of about $10^{20}\ {\rm cm^{-3}}$, may actually become significant only at carrier concentrations of order $10^{21}\ {\rm cm^{-3}}$.

Figure 4 depicts the binding energy ($E_{\rm pol}$) dependence of the temperature T_c at which the polaron density is halved, as well as its dependence on the maximum group velocity u of the phonons. (The concentration of hot fermions turns out to be negligible at the total concentration $n=0.4n_0$ and the temperatures used in constructing Fig. 4.) The temperature T_c can be called the temperature of an isostructural phase transition, since at this temperature one phase (the macroscopic number of self-localized carriers) disappears, just as liquid disappears in boiling. This phase transition has a nonvanishing latent heat equal to the product of the number of polarons and the polaron binding energy $E_{\rm pol}$. The following formula approximates the dependence of T_c on u and $E_{\rm pol}$ to within $\simeq 5\%$:

$$T_c(u,E_{\rm pol})=1.185E_{\rm pol}1.5^{\log(muc/p_0)}, \tag{7}$$

where c is the reciprocal effective dielectric constant. The width of the transition region always turns out to be of order T_c.

4. DISTRIBUTION OF FERMIONS IN A MEDIUM WITH MANY-COMPONENT POLARIZABILITY

The method of building the distribution function in a system with coexisting self-localized and free fermions can easily be extended to a medium with many-component polarizability (several phonon branches interacting with the charge carriers). As shown Ref. 2, if the minimum phase velocities of these phonon branches differ and the branch with the greatest phase velocity is the one whose interaction with a carrier satisfies the adiabatic condition, the charge carriers in such a medium can be in several different polaron states. Indeed, if the interaction between a carrier and one of the phonon branches leads to the formation of an self-localized state, the polarizability of the medium due to the interaction between the other branches and a polarons is added to the polaron, changing its binding energy, effective mass, and radius.

Let us consider the simplest case of a medium with two-component polarizability, where the interaction between a carrier and a phonon branch leads to the self-localization of the carrier, and the minimum phase velocity of the phonons belonging to this branch, u_2, is greater than the velocity of the phonons of the second branch interaction with the carrier. Two types of polaron can coexist in such a medium. The polarizability cloud of one of the polarons consists of virtual phonons from both branches. When the velocity of such a polaron exceeds the lesser of the two minimum phase phonon velocities (u_1), the polaron's motion leads to emission of real phonons of the respective branch.[2] This results in the polaron losing its polarizability cloud formed by phonons of this branch. Thus, a polaron moving with a velocity v, with $u_1<v<u_2$, has only one polarizability cloud, which consists of virtual phonons with the minimum phase velocity u_2. We call such a polaron an ordinary polaron, while a polaron whose velocity v is less than u_1 we call a double polaron. Obviously, the biding energy E_{DP} and the effective mass m_{DP} of a double polaron are greater than the respective values for an ordinary polaron (E_{OP} and m_{OP}), while the volume is smaller, $V_{DP}<V_{OP}$.

We examine the normalization condition for the Gibbs distribution in a system containing both types of polaron and free carriers. Obviously, the contribution of the case with $N=0$ (as earlier, N is the number of cold fermions in the volume V_0) does not change. If we neglect the difference in the volumes of ordinary and double polarons (i.e., if we assume that $V_{OP}=V_{DP}=V_0$), the contribution of the case with $N=1$ has the form

$$\frac{\exp[(\Omega-\mu N)/T]}{(4/3)\pi p_0^3}\left[\frac{1}{3}\int_0^{mu_1}4\pi p^2dp\,\exp\left(-\frac{E_1}{T}\right)+\frac{1}{3}\int_{mu_1}^{mu_2}4\pi p^2dp\,\exp\left(-\frac{E_1'}{T}\right)+\frac{1}{3}\int_{mu_2}^{p_0}4\pi p^2dp\times\exp\left(-\frac{E_2}{T}\right)\right]. \tag{8}$$

If we assume that the zero of energy and the chemical potential coincides with the bottom of the ordinary-polaron band,

$$E_1'(p)=\frac{p_{OP}^2}{2m_{OP}}=\frac{p^2m_{OP}}{2m^2},$$

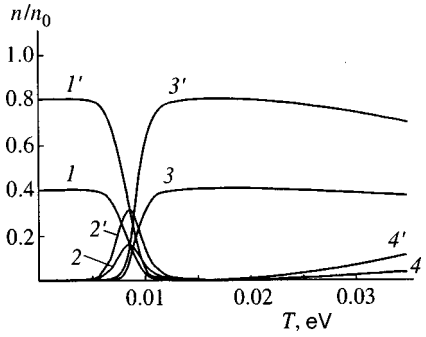


FIG. 5. Temperature dependence of the fermion concentration in each of the states that the fermions can form in a medium with two-component polarizability for two values of the total fermion concentration n: curves *1* and *1′*, a double polaron; curves *2* and *2′*, an ordinary polaron; curves *3* and *3′*, cold delocalized carriers; and curves *4* and *4′*, hot fermions (the ''unprimed'' and ''primed'' curves correspond to $n=0.4n_0$ and $n=0.8n_0$, respectively).

$$E_1(p)=\frac{p_{DP}^2}{2m_{DP}}-(E_{DP}-E_{OP})$$
$$=\frac{p^2m_{DP}}{2m^2}-(E_{DP}-E_{OP}),$$
$$E_2(P)=E_{OP}+\frac{p^2}{2m},$$

where p is the carrier momentum. Using the normalization condition, we obtain an expression for the concentration of cold fermions in a system with two-component polarizability:

$$n_{\text{cold}}=-\frac{2}{V_0}\frac{\partial\Omega}{\partial\mu}=\frac{(2/3)B\exp(\mu/T)}{1+(1/3)B\exp(\mu/T)}, \tag{9}$$

where

$$B=\frac{V_0}{(2\pi\hbar)^3}\left[\int_0^{mu_1}4\pi p^2dp\exp\left(-\frac{E_1}{T}\right)\right.$$
$$+\int_{mu_1}^{mu_2}4\pi p^2dp\exp\left(-\frac{E_2'}{T}\right)+\int_{mu_2}^{p_0}4\pi p^2dp$$
$$\left.\times\exp\left(-\frac{E_2}{T}\right)\right].$$

Equation (9) shows that in such a system the cold-fermion concentration consists of three terms, corresponding to the densities of ordinary and double polarons and of delocalized cold carriers. Adding the concentration of hot carriers [Eq. (5)], we obtain the total concentration of fermions in the system.

Figure 5 depicts the temperature dependence of the concentration of fermions in each of the states that can be realized in a medium with two-component polarizability for two values of the total carrier concentration, $n=0.4n_0$ and $n=0.8n_0$. The parameters of the medium used in building Fig. 5 have the following values: $E_{DP}=0.17$ eV, $E_{OP}=0.11$ eV, $m_{DP}=18m_e$, $m_{OP}=10m_e$, $u_1=10^4$ cm·s^{-1}, and $u_2=10^5$ cm·s^{-1} (in our calculations we did not allow for the dependence of the polaron's effective mass on the polaron velocity, since the energy mass depends very weakly on the polaron velocity for such values of u_1 and u_2; see Ref. 3). Clearly, the temperature at which there is a redistribution of carriers from one state to the other is independent of the concentration n of carriers in the systems. Calculations with other values of the medium parameters show that, as in the case of a medium with one-component polarizability, this temperature is determined solely by the values of u_1 and u_2 and the band gaps separating the bands of different carrier states. Formula (7) yields accurate ($\sim 10\%$) values of the temperature T_c at which the concentration of polarons of one type is halved, provided that for the polaron binding energy E_{pol} we take the band gap separating the band of the carriers whose disappearance temperature is being calculated from a neighboring (on the energy scale) but shallower band. We have not depicted the case with $n>n_0$ in a medium with two-component polarizability, since, as in the case of a medium with one-component polarizability, at temperatures below room temperature (such temperatures are considered here) it differs only in the emergence of hot carriers with a concentration $n-n_0$. Hence, in a system with any number of phonon branches interacting with a charge carrier, hot carriers may emerge at temperatures below room temperature (when $p_0^2/2m\gg T$) only if the concentration n of the carriers exceeds the maximum concentration of cold fermions, n_0.

Of course, experimentally the presence or absence of hot fermions in a system can manifest itself only in the presence or absence of screening of the electron–phonon interaction, which only hot carriers are capable of doing (see Sec. 3). However, the ordinary and double polarons and the delocalized carrier state differ in binding energy, effective mass, lifetime, and especially mobility, which for self-localized carriers differs substantially from that of delocalized carriers.[6] Hence the transition of a macroscopic fraction of the carriers from one state to another, which occurs when the temperature of the system changes, manifests itself in variations of the kinetic coefficients and related quantities. Such a transition also manifests itself in the optical properties of systems with carrier self-localization; in particular, in the characteristic variation of the lifetime of states of light-induced carriers with temperature. (One must bear in mind that not only the lifetimes of carriers excited to one of the polaron bands and to the conduction band are different but so are the relaxation times of the ordinary and double polarons due to the difference in the frequencies of phonon branches.) For instance, Stevens *et al.*[8] observed such a temperature dependence of the lifetime of light-induced carriers.

5. CONCLUSION

Analysis of the distribution function for fermions in a system in which self-localized and delocalized states of fermions coexist leads to the following conclusions concerning the thermodynamic characteristics and the behavior of such systems.

1. The highest concentration of the self-localized fermions in the system is $n_0=2V_0^{-1}$, where V_0 is the polaron volume.

2. The chemical potential of the system is independent of the concentration n of fermions if $n<n_0$ and the temperature $T\to 0$ K (due to the ability of self-localized fermions to condense in a state with zero momentum and minimum energy).

3. The chemical potential of a system of self-localized and free fermions with a fermion concentration $n<n_0$ increases with temperature, in contrast to the behavior of the chemical potential of a system of free fermions.

4. First-order phase transitions related to a change in the state of a macroscopic number of carriers can be observed in systems with self-localization. The latent heat is equal to the product of the number of polarons and the polaron binding energy.

5. The temperature at which polarons of a certain type disappear from the system depends on the maximum group velocity of phonons participating in the formation of these polarons and is usually much less than the polaron binding energy.

6. These phase transitions manifest themselves in the electrical conductivity and in the contribution of the carriers to the specific heat and also in the temperature dependence of the lifetime of the states of light-induced carriers.

These conclusions force us to change the approach to the experimental search for, and study of, polarons.

Allowance for the special features of systems containing self-localized fermions also leads to a substantial modification of our ideas about the role that bipolarons play in the electrical conduction of media with strong electron–phonon interaction. In his monograph (see Ref. 1), Pekar rejected the possibility that bipolarons might exist. Even in the refined theory of polarons,[9] bipolarons prove to have an advantage over polarons (in energy) only for an extremely rare set of parameters of the electron–phonon system. In Ref. 2 it was shown that two-center resonant bipolarons can exist with essentially the same set of parameters. Allowance for the fact that the bipolaron band is bounded in momentum leads to a substantial increase in the temperature T_0 of bipolaron condensation, so that this temperature can easily be 100 K for carrier concentrations in the 10^{20}–$10^{21}\,\mathrm{cm}^{-3}$ range.[2]

The above analysis of the properties of systems of self-localized and delocalized fermions suggests that for a more common set of parameters, for example, those typical of complex oxides, for which bipolarons are metastable (the energy per carrier in a bipolaron is greater than in a polaron), bipolarons are capable of condensing at temperatures in the 100–150 K range and carrier densities in the 10^{20}–$10^{21}\,\mathrm{cm}^{-3}$ range. This is the result of the above-noted ability of the chemical potential of a system of self-localized and delocalized carriers to increase with temperature. Indeed, the Bose nature of bipolarons makes it possible to account for them by adding to (6) the number of carriers in the bipolaron state, N_{bipol}, obtained from the ordinary Bose distribution of bipolarons with allowance for the fact that the bipolaron band is bounded in momentum:[2]

$$\overline{N}_{\mathrm{bipol}}=\frac{2V_0}{(2\pi\hbar)^3}\int_0^{Mu}\frac{4\pi p^2 dp}{\exp[(E_{\mathrm{bipol}}(p)-2\mu)/T]-1}, \quad (10)$$

where M is the bipolaron mass. It turns out that as the temperature increases, the chemical potential of carriers in a system that at $T\to 0$ K and $n<n_0$ is in the polaron band can reach $\mu=E_{\mathrm{bipol}}(0)/2$. The Bose condensate produced in the process can disappear only at high temperatures, transforming into a Bose gas and delocalized states of carriers. But if $n>n_0$, the chemical potential coincides with the bottom of the bipolaron band even at $T=0$ K, so that the temperature interval within which a Bose condensate of bipolarons can exist is $0\ \mathrm{K}<T<T_0$.

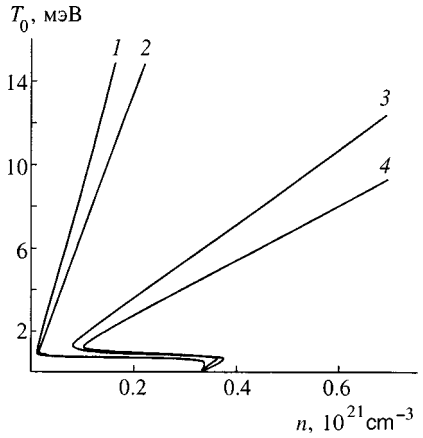


FIG. 6. The Bose condensation temperature for bipolarons as a function of carrier concentration for the case where the polaron band lies below the bipolaron band. Curve *1* corresponds to $u=2\times10^4\,\mathrm{cm\cdot s}^{-1}$ and $M=30m_e$; curve *2*, to $u=2\times10^4\,\mathrm{cm\cdot s}^{-1}$ and $M=40m_e$; curve *3*, to $u=10^5\,\mathrm{cm\cdot s}^{-1}$ and $M=30m_e$; and curve *4*, to $u=10^5\,\mathrm{cm\cdot s}^{-1}$ and $M=40m_e$.

By way of example, Fig. 6 depicts a (T_0,n) phase diagram, where T_0 is the temperature of Bose condensation, of a system of bipolarons for two values of the parameter u ($u=2\times10^4\,\mathrm{cm\cdot s}^{-1}$ and $10^5\,\mathrm{cm\cdot s}^{-1}$) and bipolaron mass ($M=30m_e$ and $40m_e$; the value of M is estimated in accordance with Refs. 4 and 5). Clearly, the high-temperature section of the curve $T_0(n)$ reaches $T_0=100$ K at carrier densities even below the usual values for high-T_c superconductors. The possibility of the condensate disappearing as the temperature drops to very low values is due to the condensation of carriers in polaron states, which are advantageous in energy.

*)E-mail: rochal@phys.rnd.runnet.ru

[1] L. D. Landau and S. I. Pekar, Zh. Éksp. Teor. Fiz. **18**, 419 (1948); S. I. Pekar, *Studies in Electron Crystal Theory*, Rept. AEC-tr-5575, U. S. Atomic Energy Commission (1963).
[2] A. É. Myasnikova, Phys. Rev. B **52**, 10 457 (1995).
[3] A. É. Myasnikova and É. N. Myasnikov, Zh. Éksp. Teor. Fiz. **112**, 278 (1997) [JETP **85**, 152 (1997)].
[4] A. É. Myasnikova and E. N. Myasnikov, Phys. Rev. B **56**, 5316 (1997); Zh. Éksp. Teor. Fiz. **115**, 180 (1999) [JETP **88**, 101 (1999)].
[5] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3rd ed., Pergamon Press, Oxford (1980).
[6] A. E. Myasnikova, Ferroelectrics **133**, 247 (1992).
[7] G. Iadonisi, G. Capone, V. Cataudella, and G. De Filippis, Phys. Rev. B **53**, 13 497 (1996).

[8] C. J. Stevens, D. C. Smith, C. Chen, J. F. Ryan, B. Podobnik, D. Mihailovic, G. A. Wagner, and J. E. Evetts, Phys. Rev. Lett. **78**, 2212 (1997); D. Mihailovic, C. Stevens, J. Demsar, B. Podobnik, D. C. Smith, and J. F. Ryan, J. Supercond. **10**, 337 (1997).

[9] S. G. Suprun and B. Ya. Moĭzhes, Fiz. Tverd. Tela (Leningrad) **24**, 1571 (1982) [Sov. Phys. Solid State **24**, 903 (1982)].

Translated by Eugene Yankovsky

Ergodic theorem for an impurity spin subsystem in a paramagnet

F. S. Dzheparov*)

Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia
(Submitted 23 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1398–1418 (October 1999)

A model for verifying and developing the fundamental ideas underlying the ergodic hypothesis is proposed. The model describes the dynamics of the spin subsystem formed by impurity charges with spin I and a small g factor in a crystal immersed in a strong constant external magnetic field under conditions where the spin system of the nuclei in the crystal is isolated from the other degrees of freedom. The additive integral of motion is the projection of the total spin of the subsystem onto the external field. Attention is focused mainly on the case of $I=1/2$. It is shown that the ergodic hypothesis holds if the correlation radius is finite in the initial state and that the ergodic hypothesis is violated if the initial state is sharply localized or has global correlation. The nonergodicity of the ^{8}Li–^{6}Li spin subsystem, which is a convenient object for experimental investigations of spin dynamics, is revealed. An estimate is obtained for the time for transition from a sharply localized disturbance of the canonical distribution to a quasiequilibrium state. © *1999 American Institute of Physics.* [S1063-7761(99)02010-7]

1. INTRODUCTION

The ergodic hypothesis is usually interpreted as follows. The mean value of any observable over a long time interval is equal to its mean over any of the Gibbs distributions:

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t d\tau\, f(\tau)=\langle f\rangle_G. \tag{1}$$

This hypothesis is one of the most important elements of modern statistical physics. Originally created to substantiate equilibrium theory, this hypothesis subsequently made its way into physical kinetics through the conception of local equilibrium ensembles.[1]

One of the most important applications of the ergodic hypothesis stems from the fact that, usually, in deriving kinetic equations either it is directly required that their equilibrium solutions be consistent with the Gibbs distribution,[2] or strong hypotheses involve decoupling of the many-particle correlations, which satisfy this requirement, are introduced.[1,3–6]

The standard derivation of the Gibbs distribution[7] is based on the hypothesis that macroscopic subsystems of a large system are quasi-independent. The literature devoted to attempts to directly prove the ergodic hypothesis is vast and frequently very complicated. Opposite views of the degree to which this hypothesis holds (see, for example, Refs. 8–13) and the very possibility of constructing a theory of irreversible processes on the basis of pure (quantum) mechanics within known interactions[13,14] have coexisted for more than 30 years. The terminological barriers separating physics from mathematics here are extremely high. Thus, it would be useful to find a simple but inclusive physical process for verifying and developing at least some central ideas of ergodic theory.

For this purpose, in this paper we examine the dynamics of impurity nuclear spins in the case where the external constant magnetic field is significantly stronger than the local fields induced by the spins at one another, the complete spin system of the host and impurity nuclei is isolated from the other interactions in the crystal, the magnetic moments of the impurity nuclei are smaller, and the distances between them are greater than in the host. When these conditions are satisfied, the influence of the impurity on the evolution of the host spins can be neglected, and the fluctuations of the local fields induced by the host spins at the impurity spins are faster than the processes unfolding in the impurity subsystem. A realistic example is provided here by the spin system of ^{107}Ag nuclei (the impurity, with spin $I=1/2$) and ^{19}F nuclei (the host) in an AgF crystal enriched with the ^{109}Ag isotope (the natural abundances of the ^{107}Ag and ^{109}Ag isotopes are roughly equal, and the interaction of the ^{109}Ag and ^{107}Ag spins can be neglected). The ^{109}Ag nuclei could be selected as the impurity in AgF in a completely analogous manner. Another example (for which the principal idealizations discussed below hold somewhat less well) is the ^{8}Li–^{6}Li spin subsystem in LiF.[15] It can be assumed that in this system there is initially only one polarized nucleus, viz., ^{8}Li ($I=2$) and that with the passage of time its polarization is spread among the surrounding ^{6}Li ($I=1$) nuclei because the g factors of the ^{8}Li and ^{6}Li nuclei are essentially equal. In this case the observable quantity is the polarization of the β-active ^{8}Li nucleus. This system is convenient for experimental studies of many important processes in spin dynamics and statistical physics, since it offers a possibility for observing local properties of the processes owing to the presence of the β-active nuclei in it, in contrast to standard NMR methods, where all the measured quantities are essentially global.[15]

An analysis of this model would be timely, because it gives a graphic picture of the properties of a ''normal'' system, while indications of ''anomalies'' (nonergodicity[16–18] and a nondiffusive long-time asymptote[19]) in the behavior of

some $1d$ and $2d$ spin models have been found in several studies.

We note that the advantages of spin systems for treating problems in equilibrium statistical physics were exquisitely demonstrated by Kittel in Ref. 20.

This paper is organized in the following manner. In Sec. 2 the basic model is formulated, the exact master equation for the diagonal part of the density matrix ρ_D is derived, it is used to obtain a simpler Markov equation for ρ_D for the case of a weak interaction within the subsystem, and the relationship between the off-diagonal part ρ_N and ρ_D is determined. In Sec. 3 it is shown that the Markov equation can be broken up into independent equations that describe the evolution of m-particle correlations. The stationary solutions of these equations and their correspondence to the canonical Gibbs distribution are analyzed in Sec. 4. It is demonstrated in Sec. 5 that the stationary solutions of the exact master equation and the Markov equation coincide for any finite ratio between the strengths of the interaction within the subsystem and the interaction with the thermostat. These results are extended to the case of an exact secular dipole–dipole interaction in Sec. 6. The nonergodicity of the evolution of the ^{8}Li–^{6}Li system is revealed in Sec. 7. A different, perfectly general view of the role of the initial conditions in exact quantum-statistical evolution in connection with the ergodic problem is presented in Sec. 8. Some general properties of the evolution of the off-diagonal part of the density matrix are considered in Sec. 9, and an estimate is obtained in Sec. 10 for the ''thermalization'' time, during which a subsystem starting from a state containing a sharply localized disturbance of the canonical distribution approaches a quasiequilibrium state. To complete the presentation and close off some loose ends, the reversal of evolution is briefly examined in Sec. 12 for both the model studied and the real spin systems represented by it, and the derivation of the secular Hamiltonian of the dipole–dipole interaction controlling the processes considered is presented in the Appendix.

2. FORMULATION OF THE MODEL AND DERIVATION OF THE KINETIC EQUATION

If a spin system is immersed in a strong constant magnetic field aligned parallel to the z axis, the subsystem of the impurity spins I_k, $k=1,\ldots,N$, is not closed, and its energy (the sum of the Zeeman and dipolar energies) is not conserved. However, the subsystem has another additive integral of motion, viz., the z component of the total spin angular momentum:

$$I_z=\sum_{k=1}^{N} I_k^z. \tag{2}$$

Hence it can be expected that under typical conditions, where the dipolar temperature and the specific heat of the thermostat (the host spins) are infinite, a subsystem of finite, but fairly large volume can start from any physically realizable distribution and, with the passage of time, reach an equilibrium state described by the canonical distribution

$$\rho_G=\exp(\Phi-\xi I_z). \tag{3}$$

Here, as usual, Φ and ξ are determined from the conditions $\mathrm{Tr}\,\rho=1$ and $\mathrm{Tr}\,I_z\rho=\langle I_z\rangle$.

Accordingly, as the time t increases, the quantum-statistical mean $f(t)=\mathrm{Tr}(f\rho(t))$ tends to the stationary value $\langle f\rangle_G=\mathrm{Tr}(f\rho_G)$, as a result of which relation (1) is satisfied. The interaction with the thermostat gives rise to phase relaxation, and for this reason the density matrix for a sample of finite volume (with the ''correct'' choice of the initial conditions, see below) tends to the Gibbs distribution (3) as the time increases. In this section the evolution of the system under consideration differs significantly from the behavior of isolated finite objects, in which (according to the classical Poincaré recurrence theorem and its quantum-mechanical analog; see, for example Refs. 6 and 13) the initial state is reproduced repeatedly and to any accuracy with the passage of time.

The evolution of our system can be described using the theory developed in Ref. 21.

The mathematical model underlying it employs the Hamiltonian

$$H=H_0+H_1,\quad H_0=\sum_j \omega_j(t)I_j^z,\quad H_1=\frac{1}{2}\sum_{jk} a_{jk}I_j^+I_k^-, \tag{4}$$

where $I_j^+=I_j^x+iI_j^y$, $I_j^-=(I_j^+)^+$, $a_{j\neq k}=a_0r_0^3r_{jk}^{-3}(1-3\cos^2\vartheta_{jk})$, and $a_{jj}=0$. Here $\mathbf{r}_{jk}=\mathbf{r}_j-\mathbf{r}_k$ is the vector joining spins j and k, and ϑ_{jk} is the angle between $\mathbf{r}_{jk}$ and the external constant field aligned along the z axis. To fix ideas, in the treatment of the long-range part of the interaction (if the distribution of the impurities is three-dimensional) we shall assume that the sample has a spherical shape and that the macroscopic field from the nuclei is therefore equal to zero. The principal model assumption is that the local field $\omega_j(t)$ induced by the thermostat spins at the j impurity spin is a δ-correlated normal stationary random process, i.e.,

$$\langle\omega_j(t)\omega_k(\tau)\rangle_n=\frac{2}{T_2}\delta(t-\tau)\delta_{jk}, \tag{5}$$

$$U(t,t_0,[\alpha])=\left\langle \exp\left(i\sum_{j=1}^{N}\int_{t_0}^{t} d\tau\,\alpha_j(\tau)\omega_j(\tau)\right)\right\rangle_n$$

$$=\exp\left[-\sum_{j=1}^{N}\left|\int_{t_0}^{t}\frac{d\tau}{T_2}\alpha_j^2(\tau)\right|\right].$$

Here $\langle\ldots\rangle_n$ denotes averaging over the distribution of local fields (the noise), and $\alpha_j(t)$ is a fairly arbitrary real function. A discussion of the ways and results of generalizing these formulas to take into account more realistic (smooth over time and correlated in space) fluctuations of the local fields can be found in Refs. 15 and 21.

It follows from the δ-correlated nature of the process that

$$U(t,t_0,[\alpha])=U(t,t_1,[\alpha])U(t_1,t_0,[\alpha]), \tag{6}$$

if $t\geqslant t_1\geqslant t_0$.

We bring the equation of motion for the density matrix $\rho(t)$

$$\dot\rho=-i(L_0(t)+L_1)\rho,\quad L_0(t)\rho=[H_0(t),\rho],$$

$$L_1\rho=[H_1,\rho]$$

into the integral form

$$\rho(t)=V(t,0)\rho_0-i\int_0^t d\tau\, V(t,\tau)L_1\rho(\tau),$$

$$V(t,t_0)=\exp\left(-i\int_{t_0}^t d\tau\, L_0(\tau)\right). \tag{7}$$

Here $\rho_0=\rho(t=0)$, and the superoperators L_0, L_1, and V act, as usual, in a Liouville space, whose vectors are operators of the quantum-mechanical Hilbert space.

The operators I_j^z commute with one another; therefore, it follows from (5) and (6) that

$$\langle V(t,\tau)\rangle_n=\exp\left(-\frac{|t-\tau|}{T_2}\sum_{j=1}^N S_j\right),\quad S_j f=[I_j^z,[I_j^z,f]], \tag{8}$$

$$\langle V(t,t_0)\rangle_n=\langle V(t,\tau)\rangle_n\langle V(\tau,t_0)\rangle_n,$$

where f is an arbitrary operator and $t\geqslant\tau\geqslant t_0$. We average (7) over the fluctuations of the local fields under the assumption that ρ_0 does not depend on them, and we take into account, that by virtue of the last of the relations (8),

$$\langle V(t,\tau)L_1\rho(\tau)\rangle_n=\langle V(t,\tau)\rangle_n L_1\langle\rho(\tau)\rangle_n. \tag{9}$$

The validity of this equality can be demonstrated, for example, by replacing $\rho(\tau)$ by an iterative solution of Eq. (7). Thus, we have

$$\langle\rho(t)\rangle_n=\langle V(t,0)\rangle_n\rho_0-i\int_0^t d\tau\langle V(t,\tau)\rangle_n L_1\langle\rho(\tau)\rangle_n,$$

or, in differential form,

$$\langle\dot\rho\rangle_n=-(R+iL_1)\langle\rho\rangle_n,\quad R=\frac{1}{T_2}\sum_{j=1}^N S_j,\quad t\geqslant 0. \tag{10}$$

Let the projector π isolate the diagonal part from any operator, and let $\rho_D=\pi\langle\rho\rangle_n$. In this case ρ_D is diagonal in the representation of the eigenvectors of the operators I_j^z and their products. Standard transformations[22] yield a closed equation for ρ_D:

$$\lambda\rho_D=\rho_0-M(\lambda)\rho_D,\quad M(\lambda)=\pi L_1\overline{\pi}\frac{1}{\lambda+R+i\overline{L_1}}\,\overline{\pi}L_1\pi, \tag{11}$$

where $\overline{\pi}=1-\pi$,

$$\overline{L_1}=\overline{\pi}L_1\overline{\pi},\quad \rho_D(\lambda)=\int_0^\infty dt\,\rho_D(t)\exp(-\lambda t)$$

is the Laplace transform for $\rho_D(t)$. Here it has been taken into account that $\overline{\pi}\rho_0=0$, $\pi R=R\pi=0$, and $\pi L_1\pi=0$. In the generally accepted representation for spin matrices[2] the operator ρ_D and the superoperators L_1 and R are real. Hence for real λ the operator $M(\lambda)$ is also real, i.e.,

$$M(\lambda)=\mathrm{Re}\left(\pi L_1\overline{\pi}\frac{1}{\lambda+R+i\overline{L_1}}\,\overline{\pi}L_1\pi\right).$$

Accordingly, for all λ

$$M(\lambda)=\pi L_1\overline{\pi}\frac{1}{\lambda+R+\overline{L_1}\dfrac{1}{\lambda+R}\,\overline{L_1}}\,\overline{\pi}L_1\pi. \tag{12}$$

This relation was obtained from the following transformation of the resolvent:

$$\frac{1}{\lambda+R+i\Lambda}=\frac{1}{(1+i\Lambda G_0)(\lambda+R)}$$

$$=G_0\frac{1}{1+i\Lambda G_0}=G_0\frac{1}{1+(\Lambda G_0)^2}(1-i\Lambda G_0)$$

$$=\frac{1}{\lambda+R+\Lambda G_0\Lambda}(1-i\Lambda G_0), \tag{12a}$$

where $G_0=(\lambda+R)^{-1}$.

To leading order in L_1^2

$$M(\lambda)=\pi L_1\overline{\pi}\frac{1}{\lambda+R}\,\overline{\pi}L_1\pi=\frac{1}{\lambda+2T_2^{-1}}\pi L_1^2\pi,$$

where it has been taken into account that the operator $L_1\pi\rho$ is an eigenvector of the superoperator R. To the same accuracy, ρ_D satisfies the following simpler equation[21] in the limit $t\gg T_2$:

$$\dot\rho_D=-\frac{1}{2}\sum_{jk} w_{jk}[I_j^+I_k^-,[I_j^-I_k^+,\rho_D]]=-\hat A\rho_D, \tag{13}$$

where $w_{j\neq k}=a_{jk}^2T_2=w_{kj}=\nu_0(1-3\cos\vartheta_{jk})^2r_0^6/r_{jk}^6$ and $w_{kk}=0$.

An analysis of the matrix elements $M(\lambda)$ (in the basis set of operator products introduced below) which are discarded in going from (12) to (13) shows that they are small, if phase relaxation is faster than transport among impurity spins, i.e., if $\epsilon\sim\Sigma_j w_{jk}T_2\ll 1$ (Ref. 21).

It follows directly from Eq. (10) that to leading order in ϵ and for $t\gg T_2$ the off-diagonal part of the density matrix equals

$$\rho_N(t)=\overline{\pi}\rho(t)=R^{-1}L_1\rho_D(t)=2T_2[H_1,\rho_D(t)]. \tag{13a}$$

3. ISOLATION OF INVARIANT SUBSPACES

We assume that all $I_k=1/2$. We use the term ''operator product of rank m'' for the operator

$$C_m(j_1,\ldots,j_m)=\prod_{\alpha=1}^m{}' I_{j_\alpha}^z,$$

where the prime indicates that all the j_α are different. The set of all independent operator products (for example, only those in which $j_1<\ldots<j_m$) forms a complete orthogonal basis set in the subspace of diagonal density matrices.

The set of operator products of a fixed rank m forms the mth invariant subspace of the operator $\hat A$, since

$$\hat{A}C_m(j_1,\ldots,j_m)=\sum_{j_i\notin[m]}\sum_{l=1}^{m} w_{j_ij_l}(C_m(j_1,\ldots,j_m) - C_m(j_1,\ldots,j_m|j_l\to j_i)). \quad (14)$$

Here $[m]$ is the set of numbers $j_1,\ldots,j_m$, and $C_m(j_1,\ldots,j_m|j_l\to j_i)$ is the operator product $C_m(j_1,\ldots,j_m)$ in which spin j_l is replaced by spin j_i.

We note that in calculating

$$\hat{A}C(j_1,\ldots,j_m)=\frac{1}{2}\sum_{kl} w_{kl}[I_k^+I_l^-,[I_k^-I_l^+,C(j_1,\ldots,j_m)]]$$

it is sufficient to consider the situations in which only one of the indices k or l belongs to the group $j_1,\ldots,j_m$, and then a direct calculation of the commutators leads to the relation (14). In fact, if both k and l belong to this group, the corresponding contribution equals zero, since $I_k^zI_l^z=(I_k^z+I_l^z)^2/2 - 1/4$ and $[I_k^+I_l^-,I_k^z+I_l^z]=0$.

The coefficients $K_m(j_1,\ldots,j_m,t)$ in the expansion

$$\rho_D(t)=\frac{1}{\mathrm{Tr}\,1}\left\{1+\sum_{m\geqslant 1}\sum_{j_1<j_2<\ldots<j_m} z_mC_m(j_1,\ldots,j_m) \times K_m(j_1,\ldots,j_m,t)\right\}, \quad (15)$$

where

$$K_m(j_1,\ldots,j_m,t)=\langle C_m(j_1,\ldots,j_m)\rangle = \mathrm{Tr}\,C_m(j_1,\ldots,j_m)\rho_D(t),$$

$$z_m^{-1}=\mathrm{Tr}(C_m(j_1,\ldots,j_m))^2/\mathrm{Tr}\,1,$$

satisfy the equation

$$\dot{K}_m(j_1,\ldots,j_m,t)=-A_mK_m(j_1,\ldots,j_m,t) = -\sum_{j_i\notin[m]}\sum_{l=1}^{m} w_{j_ij_l}(K_m(j_1,\ldots,j_m,t) - K_m(j_1,\ldots,j_m,t|j_l\to j_i)), \quad (16)$$

which follows directly from (14) and (15).

When $m=1$, Eq. (16) coincides with the familiar equation for random walks:

$$\dot{K}_1(j)=-\sum_{i(\neq j)} w_{ji}(K_1(j)-K_1(i)). \quad (17)$$

We note that if the impurity spins form a regular d-dimensional sublattice, the long-time asymptote of the solution of Eq. (17) will be diffusive for any d. A comparison with the existing, exactly solvable one-dimensional[23] and multidimensional[24] models, together with the results of numerical simulation for $d=3$ (Ref. 25), provides a firm foundation for a similar claim for a random impurity distribution. The long-time asymptote of the solutions of Eqs. (16) has been studied to a significantly lesser extent. It is natural to expect that it is also diffusive in a dm-dimensional space.

4. ERGODIC PROPERTIES OF SOLUTIONS OF THE KINETIC EQUATION

Let us turn to the ergodic hypothesis. The single stationary state of Eq. (14) corresponds to uniform spreading of the initial condition to all $N_m=N!/(m!(N-m)!)$ independent operator products of rank m. In fact, after multiplying Eq. (16) by $K_m(j_1,\ldots,j_m,t)$ and performing some standard transformations, it is found that

$$\frac{d}{dt}\sum_{j_1<\ldots<j_m}K_m^2(j_1,\ldots,j_m,t) = -\sum_{j_1<\ldots<j_m}\sum_{j_i\notin[m]}\sum_{l=1}^{m} w_{j_ij_l}(K_m(j_1,\ldots,j_m,t) - K_m(j_1,\ldots,j_m,t|j_l\to j_i))^2.$$

Hence it follows that if $K_m(j_1,\ldots,j_m,t)$ with different j_α are different, the state is not stationary. If they are identical, the state is stationary, as is clear from (16).

Thus,

$$K_m(j_1,\ldots,j_m,t\to\infty)=Q_m,$$

$$Q_m=\frac{1}{N_m}\sum_{j_1<\ldots<j_m}K_m(j_1,\ldots,j_m,t=0). \quad (18)$$

When the Gibbs state (3) is represented in the form (15), we obtain

$$\rho_G=\frac{1}{\mathrm{Tr}\,1}\prod_{j=1}^{N}(1+z_1pI_j^z),\quad p=-\frac{1}{2}\tanh\frac{\xi}{2},$$

$$z_1=\frac{3}{I_j(I_j+1)}=4,$$

$$K_m^G(j_1,\ldots,j_m)=\mathrm{Tr}(C_m(j_1,\ldots,j_m)\rho_G)=p^m. \quad (3a)$$

Thus, we still need to compare Q_m and K_m^G.

Let the initial state have the form

$$\rho(t=0)=\left(1+\frac{3p_1^0}{I_1(I_1+1)}I_1^z\right)\Big/\,\mathrm{Tr}\,1,$$

$$p_1^0=\mathrm{Tr}\,I_1^z\rho(t=0), \quad (19)$$

i.e., initially only one spin is polarized and, accordingly, all $K_m(t=0)=0$, $m>1$. Then in the limit $t\to\infty$ the density matrix of any large but finite system takes the form

$$\rho(t\to\infty)=\rho_1=(1+\overline{p}z_1I_z)/\mathrm{Tr}\,1,\quad \overline{p}=p_1/N, \quad (20)$$

since the initial polarization is uniformly distributed among all the spins. This distribution obviously does not coincide with the Gibbs distribution described by (3) and (3a). It is important that the states (19) exist in real experiments,[15] and it will be shown below that the evolution of, for example, the ^{8}Li–^{6}Li system in LiF crystals is likewise not ergodic in the part which is accessible to measurement.

We note that the initial state (19) can be regarded as a model of the initial state in the theory of correlation functions, since the mean $\langle I_j^z(t)\rangle$ is proportional to the one-spin correlation function:

$$\langle I_j^z(t)\rangle = \mathrm{Tr}(I_j^z\rho(t)) = p_1^0 z_1 \mathrm{Tr}(I_j^z(t) I_1^z)/\mathrm{Tr}\,1.$$

Let the initial state have the so-called local equilibrium form

$$\rho(t=0) = \exp\left(\Phi - \sum_{j=1}^{N} \xi_j I_j^z\right) = \frac{1}{\mathrm{Tr}\,1} \prod_{j=1}^{N} (1 + z_1 p_j^0 I_j^z),$$

$$p_j^0 = -\frac{1}{2}\tanh\frac{\xi_j}{2}, \tag{21}$$

which ideally corresponds to the model of quasi-independent subsystems. In this case

$$K_m(j_1,\ldots,j_m, t=0) = \prod_{\alpha=1}^{m} p_{j_\alpha}^0.$$

The Gibbs distribution described by (3) and (3a) is clearly a special case of the relation (21) for $\xi_j = \xi$ and $p_j^0 = p$, where $K_m(j_1,\ldots,j_m) = p^m$.

The state (21) is an analog of the familiar initial condition traditionally used in the theory of kinetic equations.[3,1,5]

We assume that

$$\overline{p} = \frac{1}{N}\sum p_j^0 \sim N^0. \tag{21a}$$

Then it follows from formulas (18) that the state (21) corresponds to

$$K_1(t\to\infty) = Q_1 = \overline{p}, \tag{22}$$

$$K_2(j_1,j_2,t\to\infty) = Q_2 = \overline{p^2} - \frac{1}{N-1}(\overline{p^2} - \overline{p}^2)$$

$$= \overline{p}^2 - \frac{1}{N-1}\Delta p^2,$$

and, in general, if $\overline{p} \sim N^0$, then

$$K_m(j_1,\ldots,j_m, t\to\infty) = \overline{p}^m + O((m-1)/N). \tag{23}$$

We ultimately obtain a distribution which is indistinguishable from the Gibbs distribution to within quantities $\sim O[(m-1)/N]$ in the first m operator products, i.e., the evolution of the system is ergodic in this case. We note that although the initial state (19) is a special case of Eq. (21), opposite conclusions were drawn regarding the ergodicity of the behavior, since the condition $\overline{p} \sim N^0$ does not hold in the case of (19).

We stress that the distribution (20) can formally be represented in the form

$$\rho_1 = \rho_{G1}(1 + O(1/N^2)), \quad \rho_{G1} = \exp(\Phi_1 - \overline{p}_1 z_1 I_z),$$

and the relations (21a)–(23) correspond to

$$\rho_2 = \rho_{G2}(1 + O(1/N)), \quad \rho_{G2} = \exp(\Phi_2 - \overline{p}_2 z_1 I_z),$$

where $\overline{p}_1 \sim 1/N$ and $\overline{p}_2 \sim N^0$. When the expressions are written in this manner, ρ_1 appears to be far closer to the Gibbs distribution than ρ_2. We drew the opposite conclusion because K_m for $m \geqslant 2$, calculated with the distribution ρ_{G2}, differs from the exact value (23) by an amount of order $1/N$, while the analogous deviation for ρ_{G1} is not small.

The initial distribution (21) is one of the simplest distributions in which the states of different spins are not correlated. If the initial states are correlated in a finite volume of radius R_c containing N_c spins, i.e., if, for example, $K_2(i,j,t=0) \neq p_i^0 p_j^0$, but $K_2(i,j,t=0) \to p_i^0 p_j^0$ as $R_c/r_{ij} \to 0$, in which case

$$Q_2 = \frac{1}{N_2}\sum_{i<j} K_2(i,j,t=0) = \overline{p}^2 + O\left(\frac{N_c}{N}\right), \quad \overline{p} \sim N^0, \tag{24}$$

and, in general,

$$Q_m = \overline{p}^m + O((m-1)N_c/N), \tag{25}$$

then the system is again ergodic to leading order in mN_c/N. However, if the correlations are global, i.e., if for $N\to\infty$

$$Q_m \neq Q_1^m = \overline{p}^m, \tag{26}$$

the ergodicity is violated.

5. NONPERTURBATIVE THEORY: SPECTRAL ANALYSIS OF THE PROPERTIES OF AN EXACT MEMORY OPERATOR

Our analysis is based on Eq. (13), which was obtained by expanding the memory operator M in powers of $\epsilon \sim \Sigma_j w_{jk} T_2 \ll 1$. Hence from the formal mathematical standpoint it is not rigorous for finding the asymptote in the limit $t\to\infty$.

For a more systematic analysis we consider the superoperator spectrum $M(\lambda)$ in Eq. (12). We represent it as

$$M(\lambda) = \pi L_1 \overline{\pi}\,\frac{1}{\lambda + R + M_1}\,\overline{\pi} L_1 \pi,$$

$$M_1 = \overline{L}_1 \frac{1}{\lambda + R}\,\overline{L}_1, \quad R \equiv \overline{\pi} R \overline{\pi}. \tag{27}$$

At real nonnegative values of λ the Hermitian superoperator $M_1(\lambda)$ is nonnegative, since for any ϕ

$$(\phi|M_1(\lambda)|\phi) = \sum_{\mu} |(\phi|\overline{L}_1|\mu)|^2 \frac{1}{\lambda + R_\mu} \geqslant 0, \tag{28}$$

where the scalar product in the Liouville space is defined, as usual, by the relation

$$(\phi|\chi) = \mathrm{Tr}(\phi^+ \chi),$$

and $|\mu)$ is an eigenvector for R. In addition, $R|\mu) = R_\mu|\mu)$ and $R_\mu \geqslant 1/T_2$. The eigenvectors of R can be chosen in the form

$$|\mu) = C_m(j_1,\ldots,j_m) \prod_{\alpha=1}^{n}{}' I_{l_\alpha}^{s_\alpha}, \quad n \geqslant 1. \tag{29}$$

Here, as in the original definition for $C_m(j_1,\ldots,j_m)$, all the subscripts $j_1,\ldots,j_m$ and $l_1,\ldots,l_n$ of the spins are different. The indices s_α take the values $+$ and $-$ and label the components of the spin operators that are orthogonal to the external field. The eigenvalues $R_\mu = n/T_2$ depend only on the element n of the multi-index μ. The smallest eigenvalue of the Hermitian superoperator $M_0 = M_1 + R$ is not smaller than $1/T_2$, since M_1 is nonnegative.

The stationary solutions of the exact kinetic equation

$$\dot{\rho}_D = -\int_0^t d\tau M(\tau)\rho_D(t-\tau) \tag{30}$$

coincide with the eigenvectors of the Hermitian superoperator $M(\lambda)$ [and $M(t)$] corresponding to zero eigenvalues. To find these eigenvectors, we consider an arbitrary diagonal element

$$(f|M(\lambda)|f) = \sum_\nu |(\nu|L_1|f)|^2(\lambda + M_{0\nu})^{-1},$$

where $\pi f = f$, and $|\nu)$ and $M_{0\nu}$ are eigenvectors and eigenvalues for M_0. This expression can vanish for $\lambda \geqslant 0$ only if $(\nu|L_1|f) = 0$. Because the set of $|\nu)$ is complete, this condition requires that $L_1|f) = 0$. We expand $|f)$ in the basis set $C_m(j_1,\ldots,j_m)$:

$$|f) = \sum_m \sum_{j_1<\ldots<j_m} f_m(j_1,\ldots,j_m)C_m(j_1,\ldots,j_m).$$

We also take into account that

$$L_1 I_k^z = \frac{1}{2}\sum_i a_{ik}(I_i^+ I_k^- - I_i^- I_k^+).$$

Thus, we obtain the equation

$$\begin{aligned} L_1|f) = \frac{1}{4}\sum_m \sum_{j_1<\ldots<j_m} \sum_{j_i\notin[m]} \sum_{l=1}^m a_{j_i j_l}(I_{j_i}^+ I_{j_l}^- - I_{j_i}^- I_{j_l}^+) \\ \times C_m(j_1,\ldots,j_m|\overline{j_l})(f_m(j_1,\ldots,j_m) \\ - f_m(j_1,\ldots,j_m|j_l\to j_i)) = 0, \end{aligned} \tag{31}$$

where

$$C_m(j_1,\ldots,j_m|\overline{j_l}) = \prod_{\alpha=1(\alpha\neq l)}^{m\,\prime} I_{j_\alpha}^z.$$

With consideration of the completeness of the basis set of operator products (29), we ultimately arrive at the redefined system of equations

$$\begin{aligned} &a_{j_i j_l}(f_m(j_1,\ldots,j_m) - f_m(j_1,\ldots,j_m|j_l\to j_i)) = 0, \\ &j_l \in [m], \quad j_i \notin [m] \end{aligned} \tag{32}$$

with the single nontrivial solution $f_m(j_1,\ldots,j_m) = f_m$, which does depend on the coordinates of the spins. These solutions clearly coincide with the solutions (18) obtained above.

Any operator $O_m = \Sigma_{j_1<\ldots<j_m} C_m(j_1,\ldots,j_m)$ can be represented as a polynomial of I_z of degree m, and, conversely, any power $(I_z)^m$ can be expanded in operator products of rank $k \leqslant m$ as

$$I_z^m = a_0 + \sum_{n=1}^m a_n O_n.$$

Hence all the stationary solutions ρ_{DS} are described by the general formula

$$\rho_{DS} = F(I_z), \tag{33}$$

where $F(x)$ is a fairly arbitrary function, which is constrained only by the conditions that ρ_{DS} be nonnegative, Hermitian, and normalized.

In this derivation [as well as in the discussion of the question of the uniqueness of the stationary solution of Eqs. (16) in Sec. 4] it is significant that a_{ij} is a long-range interaction. If the interaction vanishes outside a certain radius r_c, isolated clusters, which cannot be coupled by a nonzero interaction to the surrounding spins, can appear. Minimum clusters can be singled out among them. The stationary solutions $f_m(j_1,\ldots,j_m)$ can be classified according to the numbers of the minimum clusters among which the spins $j_1,\ldots,j_m$ are distributed. Importance is attached here not to the actual labels of the spins, but only to the fact that they belong (or do not belong) to specific minimum clusters. The solution presented above for long-range a_{ij} corresponds to the case where there is one minimum cluster—the entire system.

We can show for complex $\lambda = \lambda_1 + i\lambda_2$ that all the singularities of the resolvent $G(\lambda) = [\lambda + M(\lambda)]^{-1}\pi$ are located at $\lambda_1 \leqslant 0$ and that the singularities at $\lambda_1 = 0$ correspond to $\lambda = 0$.

For this purpose we consider the auxiliary resolvent

$$Z(\lambda) = \frac{1}{\lambda + \hat{a} + i\hat{b}} = Z_1 + iZ_2, \tag{34}$$

$$Z_1 = \frac{1}{\lambda_1 + \hat{a} + (\lambda_2 + \hat{b})\dfrac{1}{\lambda_1 + \hat{a}}(\lambda_2 + \hat{b})},$$

$$Z_2 = -Z_1(\lambda_2 + \hat{b})\frac{1}{\lambda_1 + \hat{a}}.$$

Here we have utilized (12a). Let the operators $\hat{a}$ and $\hat{b}$ be Hermitian, and let $\hat{a}$ be nonnegative. The operator coefficient Z_2 (as well as Z_1) is Hermitian, as can easily be verified on the basis of relations of the form $\hat{c}(1+\hat{d}\hat{c})^{-1} = (1+\hat{c}\hat{d})^{-1}\hat{c}$. It is clear that all the singularities of $Z(\lambda)$ satisfy $\lambda_1 \leqslant 0$.

It follows from (12) and (34) that

$$M_1(\lambda) = M_1^{(1)} + i\lambda_2 M_1^{(2)}, \tag{35}$$

$$M_1^{(1)} = \overline{L_1}(\lambda_1 + R)[(\lambda_1 + R)^2 + \lambda_2^2]^{-1}\overline{L_1},$$

$$M_1^{(2)} = -\overline{L_1}[(\lambda_1 + R)^2 + \lambda_2^2]^{-1}\overline{L_1},$$

$$M(\lambda) = M^{(1)} + i\lambda_2 M^{(2)}, \tag{36}$$

$$M^{(1)} = \pi L_1 Y L_1 \pi,$$

$$M^{(2)} = -\pi L_1 Y(M_1^{(2)} + 1)(\lambda_1 + R + M_1^{(1)})^{-1}\overline{\pi}L_1\pi,$$

$$Y = \overline{\pi}[\lambda_1 + R + M_1^{(1)} + \lambda_2^2(M_1^{(2)} + 1)(\lambda_1 + R + M_1^{(1)})^{-1}(M_1^{(2)} + 1)]^{-1}\overline{\pi},$$

$$G(\lambda) = G^{(1)} + i\lambda_2 G^{(2)}, \tag{37}$$

$$G^{(1)}=\pi[\lambda_1+M^{(1)}+\lambda_2^2(M^{(2)}+1)\times[\lambda_1+M^{(1)}]^{-1}(M^{(2)}+1)]^{-1},$$
$$G^{(2)}=-\pi G^{(1)}(M^{(2)}+1)(\lambda_1+M^{(1)})^{-1}.$$

It is clear from (36) on the basis of a treatment similar to the one given above in deriving (33) that all the eigenvectors of the operator $M^{(1)}$ with zero eigenvalues are described by formula (33) and are zero eigenvectors for $M^{(2)}$ as well. Hence it follows from (37) that the singularities which $G(\lambda)$ can have at $\lambda_1=0$ also correspond to $\lambda_2=0$. All the singularities are arranged symmetrically relative to the real axis, ensuring the realness of $\rho_D(t)$.

When the singularities are arranged in this way the solutions of the exact kinetic equation (30) approach the stationary values (18) with the passage of time under any initial condition $\rho_D(t=0)$ by virtue of the ordinary properties of the Laplace transformation. The limiting values $\rho_D(t\to\infty)$ can then either coincide or not coincide with the Gibbs values, in exact agreement with the analysis performed above in Sec. 4.

In discussing the present work, V. I. Oseledets noted that we could have arrived at formula (33) in another way, by starting out from Eq. (10) and using Evans' theorem (see Refs. 26 and 27, Sec. 2.4), which states that the algebra of invariant elements of the Liouvillian operator $\mathscr{L}$ defined by the relation

$$\mathscr{L}\rho=[h,\rho]-i\sum_{k=1}^{n}(v_k^+v_k\rho+\rho v_k^+v_k-v_k^+2\rho v_k),$$

where $h^+=h$, is a commutant of the smallest self-conjugate subalgebra containing h and all the v_k.

6. GENERALIZATION TO THE EXACT SECULAR HAMILTONIAN OF THE INTERIMPURITY INTERACTION

It is natural that the solutions of Eqs. (13) and (30) at finite values of t differ in proportion to the small parameter ϵ. Corrections of the same order of magnitude should also appear when allowance is made for the so-called $z-z$ terms of the dipolar Hamiltonian

$$H_2=-\sum_{jk}a_{jk}I_j^zI_k^z, \tag{38}$$

which were omitted in (4), but unquestionably exist in real spin dynamics[2,21] (see also the Appendix). To prove that the $z-z$ terms do not influence the ergodic properties of the system under consideration, it is sufficient to note that only M_1 varies in response to the replacement $H_1\to H_1+H_2$ and, accordingly, the replacement $L_1\to L=L_1+L_2$ in Eq. (27), since $\pi L_2=L_2\pi=0$. After this, the entire spectral analysis and conclusions from the preceding section remain valid.

7. NONERGODICITY OF THE EVOLUTION OF THE ^{8}Li-^{6}Li SPIN SYSTEM

As we have already noted, the initial state (19) is realized in present-day experiments. For example, in typical β-NMR experiments[15] a sample of LiF with $N_0\sim10^{23}$Li sites simultaneously contains $N_I\leqslant10^8$ β-active ^{8}Li nuclei, which initiate the process, and $N=cN_0\sim3\times10^{-2}N_0$ impurity ^{6}Li nuclei, at which transport occurs. The process can be traced experimentally only while the initial polarization of a single ^{8}Li nucleus is spread among $N_D\leqslant10^4$ ^{6}Li spins. Thus, the entire observable part of the process is such that the zones of influence of the ^{8}Li nuclei do not overlap and the initial condition (19) can be used. In this case the index p_1 corresponds to ^{8}Li, and $p_{j\neq1}$ corresponds to ^{6}Li. The fact that this process a) is not ergodic and b) cannot be approximated by the quasiequilibrium distribution (21) follows directly already from, e.g., the fact that all the means satisfy

$$\langle C_m(j_1,\ldots,j_m,t)\rangle=0 \tag{39}$$

for even $m=2n$. It is impossible to satisfy this property within the quasi-Gibbs distributions (21). The equality (39) is a direct consequence of the invariance of the total secular Hamiltonian of the spin–spin interactions (impurity+host) in a strong constant external field under rotations by π about the x or y axis. Of course, the equality (39) also holds in the model described by (4), (5), and (38) for arbitrary values of the spins I_k.

8. MORE ON THE ROLE OF THE INITIAL STATE

We also mention another opinion regarding some of the results obtained. The canonical distribution

$$\rho_G=\exp\left(\Phi-\sum_{k=1}^{K}\beta_k\hat{J}_k\right) \tag{40}$$

is given by the mean values of the K additive integrals of motion $J_k=\mathrm{Tr}(\hat{J}_k\rho_G)$ and the normalization condition $\mathrm{Tr}\rho_G=1$. At the same time, the dispersion operators $\hat{D}_{kl}^{(2)}=\Delta\hat{J}_k\Delta\hat{J}_l$, where $\Delta\hat{J}_k=\hat{J}_k-J_k$, have perfectly definite mean values, as do all the dispersion operators of higher ranks, for example, $\hat{D}_{kl,m}^{(3)}=\Delta\hat{D}_{kl}^{(2)}\Delta\hat{J}_m$, $\hat{D}_{kl,mn}^{(4)}=\Delta\hat{D}_{kl}^{(2)}\Delta\hat{D}_{mn}^{(2)}$, etc., and they are all integrals of motion. Hence if the mean values of these operators in the initial state are such that they do not correspond to any canonical distribution, they significantly restrict the accuracy to which the density matrix can be approximated by a quasi-Gibbs or Gibbs distribution as t increases. The existence of a global correlation means that at least one of these dispersions has an anomalously large value with respect to N.

9. EVOLUTION OF THE OFF-DIAGONAL PART OF THE DENSITY MATRIX

In the general case, where the density matrix has nonzero off-diagonal components in the initial state, the following representation, which is a direct generalization of (11) is convenient:

$$\hat{G}(\lambda)=(\lambda+R+iL)^{-1}=G(\lambda)+\pi\hat{G}(\lambda)\overline{\pi}+\overline{\pi}\hat{G}(\lambda)\pi+\overline{\pi}\hat{G}(\lambda)\overline{\pi}, \tag{41}$$
$$G(\lambda)=\pi\hat{G}(\lambda)\pi=\pi[\lambda+M(\lambda)]^{-1},$$

$$M(\lambda)=\pi L_1\overline{\pi}\,\frac{1}{\lambda+R+\overline{L}\dfrac{1}{\lambda+R}\,\overline{L}}\,\overline{\pi}L_1\pi, \tag{42}$$

$$\overline{\pi}\hat{G}(\lambda)\pi=-\overline{G}(\lambda)iL_1G(\lambda), \tag{43}$$

$$\pi\hat{G}(\lambda)\overline{\pi}=-G(\lambda)iL_1\overline{G}(\lambda), \tag{44}$$

$$\overline{\pi}\hat{G}(\lambda)\overline{\pi}=\overline{G}(\lambda)-\overline{G}(\lambda)L_1G(\lambda)L_1\overline{G}(\lambda), \tag{45}$$

$$\overline{G}(\lambda)=\overline{\pi}[\lambda+R+i\overline{L}]^{-1}, \tag{46}$$

where $L=L_1+L_2$ and $\overline{L}=\overline{\pi}L\overline{\pi}$. Formula (34) and the fact that none of the eigenvalues of R in the space of off-diagonal density matrices is smaller than $1/T_2$ clearly indicate that all the singularities of $\overline{G}(\lambda)$ are located at $\lambda_1\leqslant -1/T_2$.

It follows directly from (41)–(46) that the singularities of the off-diagonal part of the density matrix $\rho_N(\lambda)=\overline{\pi}\rho(\lambda)=\overline{\pi}G(\lambda)\rho_0$ are determined by the singularities of both $\overline{G}(\lambda)$ and $G(\lambda)$. In addition, all the contributions to the damping which are slower than $\exp(-t/T_2)$ are generated from $G(\lambda)$, while $\overline{G}(\lambda)$ determines only the amplitude of these terms. The representation (13a) illustrates this general situation in the case of small ϵ.

Unlike ρ_D, the off-diagonal part satisfies $\rho_N(t\to\infty)=0$. In fact, we have $\rho_N(t\to\infty)=\lim\lambda\rho_N(\lambda)$ when $\lambda\to+0$, and it follows from (44)–(46) that the latter expression can be nonzero only by virtue of the multiplier $\lim\lambda G(\lambda)$, which, according to the analysis in Sec. 5 (together with all the multipliers to the right of it) has the form $F(I_z)$. However, this multiplier appears in (44) and (45) only as the product $L_1F(I_z)\equiv 0$.

An important class of experimentally observable quantities associated with off-diagonal density matrices consists of quantities for which either the off-diagonal initial state or the off-diagonal operator itself is invariant to interchange (relabeling) of the spins. Both these properties are satisfied at once, for example, for the free-induction decay signal

$$\Gamma(t)=\mathrm{Tr}(I_+(t)I_-)/\mathrm{Tr}\,1=\mathrm{Tr}(I_+\overline{\pi}\hat{G}(t)\overline{\pi}I_-)/\mathrm{Tr}\,1,$$

which can be written in the form

$$\Gamma(t)=\frac{1}{h}\,\mathrm{Tr}(I_+\rho(t))$$

for $\rho_0=(1+hI_-)/\mathrm{Tr}\,1$, where h is a certain constant. In the class of observables under consideration, the contribution of $\overline{\pi}G\pi$ and $\pi G\overline{\pi}$ vanishes, and $\overline{\pi}G\overline{\pi}$ reduces to $\overline{G}$. Hence they decay with a rate no slower than $\exp(-t/T_2)$.

Let us prove this assertion for the case where the off-diagonal observable $\tilde{f}$ is invariant under interchange of the spins and the initial state ρ_0 is arbitrary. We introduce the (super)operator $\hat{S}$ for symmetrization with respect to interchange of the spins. The Liouvillian $\mathscr{L}=L_1-iR$, which, according to (10), determines the evolution of $\rho(t)$, is invariant under interchange of the spins, i.e., $[\hat{S},\mathscr{L}]=0$. Thus,

$$(\tilde{f}|\rho(t))=(\hat{S}\tilde{f}|\rho(t))=(\tilde{f}|\hat{S}|\rho(t))=(\tilde{f}|\hat{G}(t)\hat{S}|\rho_0), \tag{47}$$

and $\rho_0^s=\hat{S}\rho_0$ can be chosen as the initial state. We separate ρ_0^s into diagonal ($\rho_{0D}=\pi\rho_0^s$) and off-diagonal ($\rho_{0N}=\overline{\pi}\rho_0^s$) parts. Taking into account (43) and the fact that any diagonal operator which is invariant to interchange of the spins can be represented as $F(I_z)$, we find that the matrix element $(\tilde{f}|\hat{G}(\lambda)|\rho_{0D}^s)=0$, since $L_1F(I_z)=0$. According to completely analogous arguments $(\tilde{f}|\overline{G}(\lambda)L_1\{G(\lambda)L_1\overline{G}(\lambda)|\rho_{0N}^s)\}=0$. Here $\{\ldots\}$ is a certain function of I_z.

10. TIME FOR APPROACHING A LOCAL EQUILIBRIUM STATE

Local equilibrium states play an important role in physical kinetics, acting as the leading approximation to the exact statistical operator during a certain ''thermalization'' time τ_q (see, for example, Refs. 1 and 5). It is natural to assume that in our model the rate at which the quasiequilibrium (21) is established agrees in order of magnitude with $W_q\sim\Sigma_j w_{jk}$ and depends weakly on the initial state. However, τ_q depends significantly on the initial state.

We confine ourselves to a very simple (but experimentally feasible, at least in principle) case, where only one spin is initially displaced from equilibrium:

$$\rho(t=0)=\exp\left(\Phi-\xi_1I_1^z-\xi\sum_{j=2}^{N}I_j^z\right)$$
$$=\frac{1}{\mathrm{Tr}\,1}\left(1+\frac{3p_1^0}{I_1(I_1+1)}I_1^z\right)\prod_{j=2}^{N}\left(1+\frac{3p^0}{I_j(I_j+1)}I_j^z\right). \tag{48}$$

We represent the quasiequilibrium state for any moment in time t as

$$\rho_q(t)=\exp\left(\Phi(t)-\sum_{j=1}^{N}\xi_j(t)I_j^z\right)$$
$$=\frac{1}{\mathrm{Tr}\,1}\prod_{j=1}^{N}\left(1+\frac{3p_j(t)}{I_j(I_j+1)}I_j^z\right), \tag{49}$$

where $p_j(t)=K_1(j,t)$ are defined by Eq. (17). We set

$$p_j(t)=p^0+q_j(t),\quad q_j(t=0)=\delta_{j1}q_1^0=\delta_{j1}(p_1^0-p^0)$$

and compare the value of

$$K_2^q(i,j,t)=\mathrm{Tr}(C_2(i,j)\rho_q(t))$$
$$=p_i(t)p_j(t)=(p^0)^2+p^0(q_i(t)+q_j(t))+q_i(t)q_j(t) \tag{50}$$

with the exact value $K_2(i,j,t)$, which is specified by Eq. (16) with the initial condition

$$K_2(i,j,t=0)=(p^0)^2+p^0(q_i^0+q_j^0). \tag{51}$$

It follows from the linearity of Eq. (16) that the term q_iq_j in (50) is unconditionally superfluous. The remaining terms coincide approximately with the analogous terms in $K_2(i,j,t)$. Thus, τ_q must satisfy the condition

$$\max_j(q_j(\tau_q))\approx q_1(\tau_q)\ll p^0. \tag{52}$$

When this condition is satisfied, the superfluous (quadratic with respect to q_k) term in (50) is small compared with the terms which are linear with respect to q_k and convey the entire significant time dependence. In the limit $q_1^0 \gg p^0$ the inequality (52) is achieved in the diffusive stage of evolution, when $q_1(t) \sim (Wt)^{-d/2} q_1^0$. When the impurities have a random spatial distribution and we have $d=3$, the Forster constant $\beta = (256/243)\pi^3 \nu_0 (n r_0^3)^2$, where n is the impurity number density and the remaining parameters are defined in formula (13), should be employed[15,21] as W. It is clear that $\tau_q \to \infty$ as $p^0 \to 0$, and in this limit the system becomes nonergodic, in complete agreement with the results of Sec. 4.

We note one more important consequence of Eq. (16), which is that the development of multispin correlations is greatly slowed, if the spins $j_1, \ldots, j_m$ form a dense group. In fact, there are no intragroup interactions in (16); therefore, the entire evolution of a dense group is determined by its surface interactions, whose efficiency decreases with increasing m due to the decrease in the ratio of the number of spins on the surface of a group to the number in the bulk.

11. ERGODICITY AND REVERSIBILITY OF MOTION

Spin dynamics is exceptional in that highly impressive advances have been achieved in devising a systematic microscopic theory and in understanding macroscopic processes on its basis. For example, the main causes of the reversal of evolution (time reversal) of nuclear spin systems were found and realized experimentally back in 1970.[28,29] Spin diffusion was subsequently reversed in Ref. 30. It is significant that nuclear spin systems are thermodynamic in the same sense in which thermodynamics is applicable in general to magnets. In particular, paramagnetic, ferromagnetic, and antiferromagnetic states, as well as phase transitions between them, have been discovered in them, in complete agreement with Gibbs' equilibrium theory.[2]

We have shown above that the evolution in our model system gives rise to Gibbs equilibrium for an extensive list of natural initial conditions. Nevertheless, in accordance with the general principles of Hamiltonian dynamics, this evolution is reversible, as can be demonstrated experimentally by methods similar to those used in Refs. 28–30. They are based on the creation of conditions such that the effective Hamiltonian coincides with the original one multiplied by a negative numerical coefficient. This can be realized in homonuclear systems by applying a fairly strong variable external field with an amplitude ω_1 and a frequency $\tilde{\omega}$ close to the Zeeman value ω_Z. Let the effective field in the rotating coordinate frame $\omega_{\text{eff}} = \sqrt{\omega_1^2 + \Delta^2} \gg \omega_l$ be aligned at the angle $\theta = \arctan(\omega_1/\Delta)$ to the constant field $\mathscr{H} = (0,0,\mathscr{H})$. Here $\Delta = \tilde{\omega} - \omega_Z$, and ω_l is the so-called local frequency, which is determined in the standard manner.[2,29] It characterizes the root-mean-square value of the spin–spin interactions. Under these conditions the secular part of the homonuclear dipole–dipole interactions $H_{\text{sec}}(z) = H_1 + H_2$, which is defined by Eq. (A4) [or (4) and (38)], can be subjected to repeated secularization, which effectively reduces to the transformation

$$H_{\text{sec}}(z) \to H'_{\text{sec}}(z') = \frac{3\cos^2\theta - 1}{2} H_{\text{sec}}(z'). \tag{53}$$

Here $H_{\text{sec}}(z')$ is distinguished from $H_{\text{sec}}(z)$ by the replacement of the z axis by the z' axis, which is oriented along the effective field, in the operators $I_j^z I_k^z$. The difference between the directions of the z and z' axes is easily compensated by applying pulses of an oscillating field which rotate the spin system (but not the crystal lattice) as a whole. As a result, if $\cos^2\theta < 1/3$, many-particle evolution takes place under the influence of the Hamiltonian $H'_{\text{sec}}(z')$ in the direction opposite to the direction of evolution with $H_{\text{sec}}(z)$ and, obviously, at a slower rate. The value $\cos\theta = 0$ was used in Refs. 28–30.

The impurity system together with the thermostat contains two kinds of spins, viz., impurity spins (I) and thermostat spins (F). The total secular Hamiltonian of the dipole–dipole interactions is

$$H_{\text{sec}}(z) = H^I_{\text{sec}}(z) + H^F_{\text{sec}}(z) + H^{IF}_{\text{sec}}(z), \tag{54}$$

where the first two terms are defined, as in homonuclear systems, by formula (A4), and the last term is defined by (A5) and describes the interaction of the I and F spins with one another. To reverse the many-particle evolution, here we must apply two variable fields with the frequencies $\tilde{\omega}_I = \omega_{IZ} + \Delta_I$ and $\tilde{\omega}_F = \omega_{FZ} + \Delta_F$ [which are close to the resonant frequencies of the impurity (ω_{IZ}) and the thermostat (ω_{FZ})] and with the corresponding amplitudes ω_{1I} and ω_{1F} and direction angles θ_I and θ_F. The effective Hamiltonian for describing evolution in the presence of the interaction (54) and the variable fields indicated above is derived by standard methods[29,31,32] and has the form

$$\begin{aligned} H'_{\text{sec}}(z'_I, z'_F) = {} & \frac{3\cos^2\theta_I - 1}{2} H^I_{\text{sec}}(z'_I) \\ & + \frac{3\cos^2\theta_F - 1}{2} H^F_{\text{sec}}(z'_F) \\ & + \cos\theta_I \cos\theta_F H^{IF}_{\text{sec}}(z'_I, z'_F). \end{aligned} \tag{55}$$

Time reversal will clearly occur when

$$\frac{3\cos^2\theta_I - 1}{2} = \frac{3\cos^2\theta_F - 1}{2} = \cos\theta_I \cos\theta_F < 0, \tag{56}$$

i.e., when

$$\cos^2\theta_I = \cos^2\theta_F = \frac{1}{5}, \quad \cos\theta_I = -\cos\theta_F. \tag{57}$$

This method of reversing evolution can be implemented in any nuclear spin system containing two kinds of spins with significantly different gyromagnetic ratios. We note that time reversal has already been realized experimentally in heteronuclear systems,[33] but for a Hamiltonian for the interaction between different kinds of spins that is somewhat different from (A5).

If we confine ourselves to the formal framework of the model previously considered, in which the motion of the local fields is simulated by a normal random process, the evolution of the impurity spins is reversed when the sign of the spin Hamiltonian is reversed and the trajectories of the

local fields are simultaneously reversed. More precisely, let the motion of the system be described by the Hamiltonian $H=H_0(t)+H_1+H_2$ in the time interval $0\leqslant t\leqslant T$, and let reversal of its evolution be performed for $t>T$. Such reversal will be achieved if the effective Hamiltonian has the following form for $t>T$:

$$H=-\kappa[H_0(T-\kappa(t-T))+H_1+H_2],$$

where $\kappa>0$ is a numerical coefficient [we note that $\kappa<1$ is usually realized in the known methods for reversing spin evolution, that we have $\kappa=1/5$ in (55)–(57), and that in Ref. 33 we have $\kappa=1$, but the rate of forward evolution was slowed]. In particular, the initial state ρ_0 is reproduced at the time $t=(\kappa+1)T/\kappa$.

12. CONCLUSION

Our analysis has shown that the same open system can exhibit either ergodic or nonergodic behavior, depending on the initial condition. Hence exact kinetic master equations and their approximate forms can have solutions of both types, provided none of them are treated differently during the derivation of these equations.

In the example studied above ergodicity is observed, if in the initial state a) the correlations decay fairly rapidly with increasing distance between the particles and b) the additive integral of motion satisfies $\langle I_z\rangle\sim N^1$; ergodicity is violated, if in the initial state satisfies $\langle I_z\rangle\sim N^0$ or if there are global correlations. This classification is not general, but, as a whole, the results obtained permit determination of the final state for an extensive list of initial conditions.

We note that ergodicity is ensured in the system considered by phase relaxation, which stems from the interaction with the thermostat. Phase relaxation can be very effective even if the interaction causing it is weak, since, unlike longitudinal relaxation (population relaxation), it does not require a significant energy expenditure.

We stress once again that, according to our analysis, the quasiequilibrium density matrix (21) will not be a good approximation for the true density matrix $\rho_D(t)$ if the initial state is close to (19). Thus, a theory of correlation functions and a theory of kinetic equations should, generally speaking, be constructed on the basis of methods capable of taking into account this property.

I thank B. M. Gurevich, I. P Zvyagin, R. A. Minlos, V. I. Oseledets, Ya. G. Sinaĭ, E. B. Fel'dman, T. N. Khazanovich, V. E. Shestopal, the participants in the All-Moscow Seminar ''Problems in Magnetic Resonance,'' and the participants in the seminar of the Institute of Theoretical and Experimental Physics on spin dynamics for discussing this work.

APPENDIX A: RELATIONSHIP BETWEEN THE MODEL HAMILTONIAN (4) AND THE STANDARD DIPOLE–DIPOLE INTERACTION

The magnetic dipolar interaction of the nuclear spins S_1 and S_2, which are located at the points $\mathbf{r}_1$ and $\mathbf{r}_2$ and are immersed in the external field $\mathscr{H}=(0,0,\mathscr{H})$, is described by the Hamiltonian[2]

$$H_{12}=H_Z+H_{12}^d, \tag{A1}$$

$$H_Z=-(\mathbf{m}_1+\mathbf{m}_2).\mathscr{H}=-\sum_{j=1}^{2}\omega_{Zj}S_j^z,$$

$$H_{12}^d=B_{12}[\mathbf{S}_1\cdot\mathbf{S}_2-3(\mathbf{S}_1\cdot\mathbf{r}_{12})\cdot(\mathbf{S}_2\cdot\mathbf{r}_{12})r_{12}^{-2}],$$

where $\mathbf{m}_j=g_j\beta_n\mathbf{S}_j$ is the magnetic moment, g_j is the g factor, β_n is the nuclear magneton, $B_{12}=g_1g_2\beta_n^2/r_{12}^3$, and $\omega_{Zj}=g_j\beta_n\mathscr{H}$. In a fairly strong magnetic field we have $B_{12}\sim\varepsilon\omega_{Zj}$, where $\varepsilon\ll 1$ holds, and the Hamiltonian H_{12}^d can be simplified significantly. For this purpose, it is convenient to convert the equation of motion

$$\dot\rho=-i[H_{12},\rho] \tag{A2}$$

into the interaction representation using the unitary transformation $V(t)=\exp(iH_Zt)$:

$$\dot{\tilde\rho}(t)=-i[\tilde H_{12}^d(t),\tilde\rho(t)], \tag{A3}$$

$$\tilde\rho(t)=V(t)\rho(t)V^+(t),\quad \tilde H_{12}^d(t)=V(t)H_{12}^dV^+(t).$$

We use $V(t)S_j^+V^+(t)=S_j^+\exp(-i\omega_{Zj}t)$, $V(t)S_j^-V^+(t)=S_j^-\exp(i\omega_{Zj}t)$, and $V(t)S_j^zV^+(t)=S_j^z$, and we express the Cartesian components of the operators S_j^α from $H_{12}^d(t)$ in terms of $S_j^+=S_j^x+iS_j^y$ and $S_j^-=S_j^x-iS_j^y$. It is now clear that the operator $\tilde H_{12}^d(t)$ can be divided into time-independent (secular) and rapidly (in comparison to B_{12}) oscillating (nonsecular) parts and that, as is usually the case in the theory of rapidly oscillating interactions, the influence of the nonsecular part is small compared with the contribution of the secular interaction if $B_{12}t\sim 1$ (see Refs. 2, 29, 31–34)

If $\omega_{Z1}=\omega_{Z2}$, the secular part is

$$H_{12}^{\rm sec}=\overline{B}_{12}(3S_1^zS_2^z-\mathbf{S}_1\cdot\mathbf{S}_2)$$

$$=\overline{B}_{12}\left(2S_1^zS_2^z-\frac{1}{2}(S_1^+S_2^-+S_1^-S_2^+)\right), \tag{A4}$$

$$\overline{B}_{12}=\frac{1}{2}B_{12}(1-3\cos^2\vartheta_{12}),$$

which, according to formulas (4) and (38), coincides to within the notation with the two-spin interaction corresponding to the Hamiltonian H_1+H_2.

The Hamiltonian for the interaction of different spins with significantly different frequencies $\omega_{Z1}\neq\omega_{Z2}$ is

$$H_{12}^{\rm sec}=2\overline{B}_{12}S_1^zS_2^z. \tag{A5}$$

In the generally accepted representation for spin operators[2] the matrices S_j^x and S_j^z are real, and S_j^y is purely imaginary; hence the matrices of the Hamiltonians (A4) and (A5) are real. The operators (A4) and (A5) are clearly invariant to rotation through π about any axis orthogonal to the z axis. This property was utilized in Sec. 7.

In the main body of the text the interaction (A5) was replaced by a random external field acting on the spins of the subsystem being studied. This replacement, which is also known as the Anderson–Weiss–Kubo model, was introduced more than 40 years ago specifically for the case of fast fluctuations of the local fields at impurity spins, which was

considered in this paper, and no contraindications for its use in this region have been reported hitherto. Moreover, it was shown in Ref. 15 that it gives good results with experimental verification even in cases where the fluctuation rate of the local fields is comparable to $1/T_2$.

The terms of the dipolar Hamiltonian discarded in going from (A1) to (A4) describe multispin and multiple-quantum relaxation processes, which have been thoroughly studied (see, for example, Refs. 31, 34, and 35). The rates of these processes decrease with increasing values of the external field $\mathscr{H}$ no more slowly than $1/\mathscr{H}^2$, allowing us to neglect them over a broad range of values of $\mathscr{H}$ and the time t.

*)E-mail: dzheparav@vxitep.itep.ru

[1] D. N. Zubarev, *Nonequilibrium Statistical Thermodynamics*, Consultants Bureau, New York, 1974 [Russ. original, Nauka, Moscow (1971)].
[2] A. Abragam and M. Goldman, *Nuclear Magnetism: Order and Disorder*, Clarendon Press, Oxford (1982) [Russ. transl., Mir, Moscow (1984), Vols. 1 and 2].
[3] N. N. Bogoliubov, ''Problems of a dynamical theory in statistical physics,'' in *Studies in Statistical Mechanics, Vol. 1*, J. de Boer and G. E. Uhlenbeck (Eds.), Wiley, New York (1962), pp. 5–118 [Russ. original, Gostekhizdat, Moscow–Leningrad (1946)].
[4] E. M. Lifshitz, L. P. Pitaevskii, and L. D. Landau, *Physical Kinetics*, Pergamon Press, Oxford (1981) [Russ. original, Nauka, Moscow (1979)].
[5] R. L. Liboff, *Introduction to the Theory of Kinetic Equations*, Wiley, New York (1969) [Russ. transl., Mir, Moscow (1974)].
[6] Yu. L. Klimontovich, *Statistical Theory of Open Systems, Vol. 1: A Unified Approach to Kinetic Description of Processes in Active Systems*, Kluwer Academic Publishers, Dordrecht–Boston (1995) [Russ. original, Yanus, Moscow (1995)].
[7] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Vol. 1*, 3rd. ed., Pergamon Press, Oxford–New York (1980) [Newer ed. of Russ. original, Nauka, Moscow (1995), Chap. 1].
[8] R. Kubo, *Statistical Mechanics: an Advanced Course with Problems and Solutions*, 1st ed., North-Holland, Amsterdam (1965) [Russ. transl., Mir, Moscow (1967)].
[9] R. L. Dobrushin and Yu. M. Sukhov, in *Achievements of Science and Technology: Current Problems in Mathematics Series, Vol. 14* [in Russian], VINITI, Moscow (1979).
[10] B. M. Gurevich, Teor. Mat. Fiz. **90**, 424 (1992).
[11] G. A. Martynov, Usp. Fiz. Nauk **166**, 1105 (1996) [Phys. Usp. **39**, 1045 (1996)].
[12] Ya. G. Sinaĭ, *Introduction to Ergodic Theory*, Princeton University Press, Princeton, NJ (1976) [Newer ed. of Russ. original, Fazis, Moscow (1996)].
[13] I. Prigogine, *From Being to Becoming, Time and Complexity in the Physical Sciences*, W. H. Freeman, San Francisco (1980) [Russ. transl., Nauka, Moscow (1985)].
[14] G. Nicolis and I. Prigogine, *Exploring Complexity: an Introduction*, W. H. Freeman, New York (1989) [Russ. transl., Mir, Moscow (1990)].
[15] Yu. G. Abov, A. D. Gul'ko, F. S. Dzheparov, S. V. Stepanov, and S. S. Trostin, Fiz. Élem. Chastits At. Yadra **26**, 1654 (1995) [Phys. Part. Nuclei **26**, 692 (1995)].
[16] A. A. Lundin and V. E. Zobov, Physica A **115**, 185, 200 (1982).
[17] R. Bruschweiler and R. R. Ernst, Chem. Phys. Lett. **264**, 393 (1997).
[18] E. B. Fel'dman and S. Lacelle, J. Chem. Phys. **108**, 4709 (1998).
[19] D. K. Sodickson and J. S. Waugh, Phys. Rev. B **52**, 6467 (1995).
[20] C. Kittel, *Thermal Physics*, Wiley, New York (1973) [Russ. transl., Nauka, Moscow (1977)].
[21] F. S. Dzheparov, Zh. Éksp. Teor. Fiz. **99**, 982 (1991) [Sov. Phys. JETP **72**, 546 (1991)].
[22] D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions*, W. A. Benjamin, Reading, MA (1975) [Russ. transl., Atomizdat, Moscow (1980)].
[23] S. Alexander, J. Bernasconi, W. Schneider, and R. Orbach, Rev. Mod. Phys. **53**, 175 (1981).
[24] F. S. Dzheparov and V. E. Shestopal, JETP Lett. **60**, 182 (1994).
[25] F. S. Dzheparov, D. V. L'vov, and V. E. Shestopal, Zh. Éksp. Teor. Fiz. **114**, 2166 (1998) [JETP **87**, 1179 (1998)].
[26] D. E. Evans, Commun. Math. Phys. **54**, 293 (1977).
[27] V. I. Oseledets, in *Achievements of Science and Technology: Probability Theory, Mathematical Statistics, and Theoretical Cybernetics Series, Vol. 20* [in Russian], VINITI, Moscow (1983), p. 52.
[28] W.-K. Rim, A. Pines, and J. S. Waugh, Phys. Rev. Lett. **25**, 220 (1970).
[29] *Advances in Magnetic Resonance, Vol. 8*, J. S. Waugh (Ed.), Academic Press, New York (1976) [Russ. transl., Mir, Moscow (1978)].
[30] S. Zhang, B. H. Meier, and R. R. Ernst, Phys. Rev. Lett. **69**, 2149 (1992).
[31] F. S. Dzheparov and S. V. Stepanov, Preprint of the Institute of Theoretical and Experimental Physics, No. 139 (1982).
[32] R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford (1987) [Russ. transl., Mir, Moscow (1990)].
[33] M. Ernst, B. H. Meier, M. Tomaselli, and A. Pines, J. Chem. Phys. **108**, 9611 (1998).
[34] M. Goldman, *Spin Temperature and Nuclear Magnetic Resonance in Solids*, Clarendon Press, Oxford (1970) [Russ. transl., Mir, Moscow (1972)].
[35] Yu. G. Abov, M. I. Bulgakov, A. D. Gul'ko *et al.*, JETP Lett. **35**, 424 (1982).

Translated by P. Shelnitz

Combined polaron states in magnetooptical effects in a quantum well

L. I. Korovin and I. G. Lang

A. F. Ioffe Physico-Technical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

S. T. Pavlov*)

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 30 November 1998)
Zh. Éksp. Teor. Fiz. **116**, 1419–1439 (October 1999)

Combined polaron states in a rectangular quantum well in a strong magnetic field perpendicular to the well plane are discussed. These states are due to interaction between two discrete electron levels with different Landau quantum numbers (n and n_1) and different size-quantization quantum numbers (m and m_1) on the one hand and a confined LO phonon on the other under conditions of low temperature when the energy difference between the electronic levelsis equal or close to the energy of the confined LO phonon. The expression for the resonant magnetic field $H_{\rm res}$ at which a combined polaron is formed contains the energy difference between size-quantized levels, so it is a function of quantum well parameters. The separation $\Delta E_{\rm res}$ between branches in the energy spectrum of a combined polaron and $H_{\rm res}$ has been calculated as a function of the quantum well width d. The resonant field $H_{\rm res}$ can be reduced dramatically in comparison with the case $m=m_1$. The case of size-quantization with $n=n_1$ has been analyzed. The energy difference $\Delta E_{\rm res}$ is in the range $(1–5)\cdot 10^{-3}$ eV. The damping of combined polaron states due to the effect of anharmonicity on the LO phonon has been studied. Interband absorption and features in the reflection spectrum due to interband transitions have been calculated for an arbitrary ratio between the radiative and ''phonon'' lifetime of a combined polaron have been investigated. © *1999 American Institute of Physics.* [S1063-7761(99)02110-1]

1. INTRODUCTION

The polaron shift due to weak interaction between electrons and LO phonons in a strong magnetic field is calculated using perturbation theory. However, when the resonant condition

$$\omega_{\rm L1}=j\Omega_e, \quad j=1,2,3,\ldots, \tag{1}$$

holds, where $\omega_{\rm L1}$ is the highest frequency of the LO phonon and Ω_e is the electron cyclotron frequency, the coupling between Landau levels becomes resonant, since an electron on the upper Landau level can emit a real LO phonon and go to the lower level characterized by the quantum number $n-j$. When it is on the level $n-j$ the electron can absorb the previously emitted phonon and transition to the upper level, then again emit the LO phonon, and so on. All these processes have a considerable effect on the formation of the polaron state (1) and should be taken into account. The spectrum of the unperturbed electron–phonon system consists of two levels, namely, the electron on the nth Landau level and the electron on the level $n-j=n_1$ plus the LO phonon, whose curves plotted versus the electron cyclotron frequency Ω_e cross at the point where $\Omega_e=\omega_{\rm L1}/j$. Determination of the polaron state in terms of unperturbed states of the electron–phonon system when condition (1) holds involves summation of the series in the perturbation theory using the electron–phonon coupling constant, which is equivalent to taking into account multiple processes of LO phonon emission and absorption. When this sum is calculated, the degeneracy at the crossing point $\Omega_e=\omega_{\rm L1}/j$ is lifted, the polaron spectrum has the shape of two noncrossing branches, and the separation between them at point $\Omega_e=\omega_{\rm L1}/j$ is determined by the electron–phonon coupling constant. Such a polaron state was first detected in interband magnetoabsorption spectra of bulk InSb.[1]

Polaron states can be formed in a strong magnetic field in both three-dimensional (3D) and quasi-two-dimensional (2D) structures. Quasi-two-dimensional systems include quantum wells and inversion layers in metal–dielectric–semiconductor structures. Polaron states have considerable effect on magnetooptical spectra of both 3D and 2D systems, such as spectra of interband light absorption (see, for example, the reviews in Refs. 2–4). The difference between these two kinds of system is in the electron and hole spectra: in the case of three dimensions, these are one-dimensional Landau bands, in 2D structures these are discrete levels. This difference affects the splitting in the polaron spectrum at the point $\Omega_e=\omega_{\rm L1}/j$, which increases as the system dimensionality is reduced: in a 3D structure is it proportional to $\alpha^{2/3}$ (Ref. 5), and in the 2D case to $\alpha^{1/2}$ (Refs. 6–14), where $\alpha\ll 1$ is Fröhlich's nondimensional constant of coupling between electrons or holes and LO phonons.

In the case of a quantum well, which is treated in this paper as an example of a quasi-two-dimensional system, the polaron state discussed above is called a double magnetopolaron (in accordance with the number of crossing levels of

the unperturbed electron–phonon system). More complex polaron states, such as a triple magnetopolaron, a quateron, etc.,[14–17] are also possible. All these states are generated on the same size-quantized level characterized by the quantum number m. Thus, the resonant condition (1) is independent of the position of the size-quantization level and hence of the quantum well width.

The energies of the electron and hole levels, however, are functions of two integers in a quantum well; i.e., the Landau quantum number n and the size-quantization number m. Therefore, in addition to condition (1), resonances of a different type can occur in this system of discrete levels, namely, when the electron–phonon coupling links two electron (hole) levels with both n and m being different. In this case, the quantum well width d, hence the gaps between size-quantized levels, affect resonant magnetic fields. This paper is dedicated to studying these combined polaron states and their role in magnetooptical spectra.

2. RESONANT CONDITION FOR COMBINED POLARONS AND THEIR CLASSIFICATION

In this paper we consider a type I rectangular quantum well with a gap width E_g, with barriers of heights ΔE_e and ΔE_h for electrons and holes, respectively. The magnetic field vector $\mathbf{H}$ is perpendicular to the quantum well plane (it is aligned, as is usual, with the z-axis). The vector potential $\mathbf{A}_0$ is expressed in the Landau gauge:

$$\mathbf{A}_0=\mathbf{A}_0(-yH,0,0). \tag{2}$$

The energies of electrons, $E^e_{m,n}$, and holes, $E^h_{m,n}$, measured with respect to the bottom of the electron quantum well are expressed by

$$E^e_{m,n}=\varepsilon^e_m+(n+1/2)\,\hbar\Omega_e, \tag{3}$$

$$E^h_{m,n}=E_g+\varepsilon^h_m+(n+1/2)\,\hbar\Omega_h,$$

$$\Omega_{e(h)}=\frac{|e|H}{m_{c(v)}c}, \tag{4}$$

where e is the electron charge, c is the speed of light in vacuum, $m_{c(v)}$ is the electron (hole) effective mass, and $\varepsilon^{e(h)}_m=\hbar\omega^{e(h)}_m$ is the size-quantization energy of the electrons (holes).

The resonant condition for the combined magnetopolaron is the condition that the energies of two levels of the electron–phonon system coincide, namely, that of an electron on the level $E^e_{m,n}$ and of an electron on level $E^e_{m_1,n_1}$ plus an LO phonon. It is generally accepted that only electrons can interact resonantly with phonons, and it is impossible in the case of holes owing to the difference between the electron and hole effective masses. The condition that two levels of the electron–phonon system with different quantum numbers n and m coincide takes the form

$$E^e_{m,n}=E^e_{m_1,n_1}+\hbar\omega_{\mathrm{L1}}. \tag{5}$$

By substituting in Eq. (5) the expression for $E^e_{m,n}$ from Eq. (3), we obtain an expression for the resonant cyclotron frequency in the case of a combined magnetopolaron:

$$\Omega^{\mathrm{res}}_e=\frac{\omega_{\mathrm{L1}}-(\omega^e_m-\omega^e_{m_1})}{n-n_1}. \tag{6}$$

The quantum numbers m and n characterize the purely electronic state of the electron–phonon system, whereas m_1 and n_1 refer to an electronic state to which one LO phonon is added.

It is known that energies of levels in a quantum well of a finite depth are determined by the equations

$$\cot t=\frac{t}{\sqrt{\beta_e^2-t^2}},\quad m=1,3,5,\ldots,$$

$$\tan t=-\frac{t}{\sqrt{\beta_e^2-t^2}},\quad m=2,4,6,\ldots, \tag{7}$$

where

$$t=\frac{k_m d}{2},\quad \beta_e=\frac{Q_e d}{2},$$

$$Q_e=\sqrt{\frac{2m_c\Delta E_e}{\hbar^2}},\quad k_m=\sqrt{\frac{2m_c\varepsilon^e_m}{\hbar^2}}, \tag{8}$$

and ΔE_e is the barrier height for the electron quantum well. In the limit $\Delta E_e\to\infty$ (an infinitely deep well), we have $k_m d\to m\pi$. One can see in Eq. (6) that there are three different configurations with different relations between the quantum numbers m,m_1 and n,n_1. First, $m>m_1$ and $n>n_1$, second, $m>m_1$ and $n<n_1$, and, finally, $m<m_1$ and $n>n_1$. The configuration with $m<m_1$ and $n<n_1$ is out of question because it requires that Ω^{res}_e is negative. Let us discuss these three configurations in detail.

The first configuration requires that

$$\hbar\omega_{\mathrm{L1}}\geqslant\varepsilon^e_m(d)-\varepsilon^e_{m_1}(d). \tag{9}$$

Suppose that the equality in condition (9) takes place when $d=d_{\min}(m)>d'_{\min}(m)$ holds, where

$$d'_{\min}(m)=(m-1)\pi\sqrt{\hbar^2/2m_c\Delta E_e} \tag{10}$$

is the value of d at which the upper level m is ejected from the well. It can be derived from Eq. (7). Since the difference $\varepsilon^e_m(d)-\varepsilon^e_{m_1}(d)$ decreases monotonically with the quantum well width d from $\Delta E_e-\varepsilon^e_{m_1}[d'_{\min}(m)]$ to zero, the inequality (9) holds in the interval

$$\infty>d>d_{\min}(m). \tag{11}$$

For $\hbar\omega_{\mathrm{L1}}>\Delta E_e-\varepsilon^e_{m_1}[d'_{\min}(m)]$ (a shallow well) condition (9) also holds at $d<d_{\min}(m)$, i.e., throughout the interval of d in which there is an upper level in the well.

In the second configuration, the following condition should be satisfied:

$$\hbar\omega_{\mathrm{L1}}\leqslant\varepsilon^e_m(d)-\varepsilon^e_{m_1}(d), \tag{12}$$

which holds in the interval

$$d'_{\min}(m)<d<d_{\min}(m), \tag{13}$$

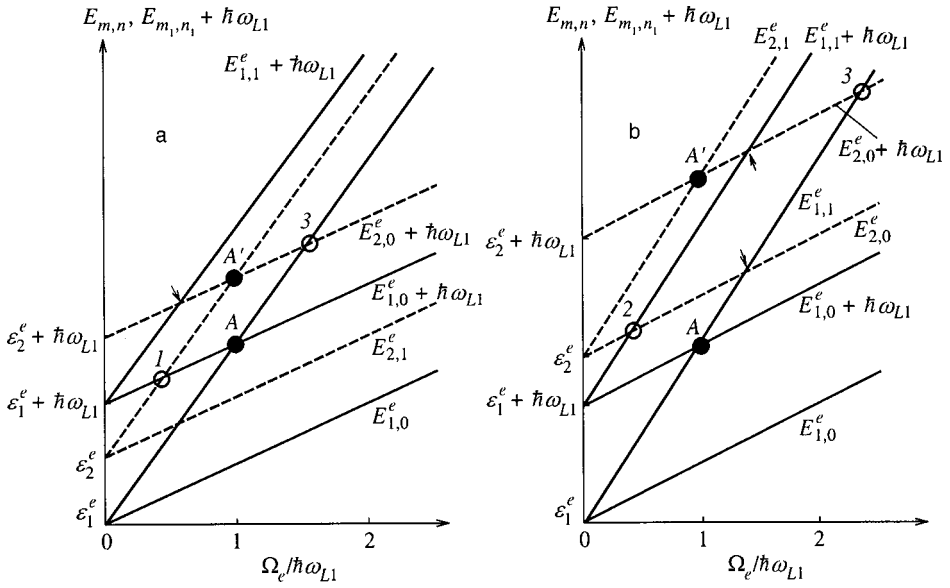


FIG. 1. Energy levels of an unperturbed electron–phonon system in a quantum well as functions of the electron cyclotron frequency Ω_e; ω_{L1} is limiting frequency of the confined phonon. The solid lines plot curves for the quantum number $m=1$, the dashed lines for $m=2$; A and A' denote simple polarons; *1*, *2*, and *3* (open circles) show data for combined polarons. The arrows mark weak polarons; (a) $\hbar\omega_{L1}>\varepsilon^e_2-\varepsilon^e_1$; (b) $\hbar\omega_{L1}<\varepsilon^e_2-\varepsilon^e_1$; n is the Landau quantum number.

since for $d>d_{\min}(m)$ the second configuration does not contain a polaron, and for $d<d'_{\min}(m)$ the mth level is ejected from the well. Note that the first and second configurations cannot be unified because different levels of the electron–phonon system cross one another in different configurations.

In the third configuration the right-hand side of Eq. (6) is positive, so it can be realized in the interval $\infty>d\geqslant d'_{\min}(m_1)$. Figure 1 shows energy levels of the unperturbed electron–phonon system as functions of the cyclotron frequency Ω_e for the two lower size-quantized levels and two Landau levels characterized by quantum numbers $n=0$ and $n=1$. The solid lines represent levels with $m=1$ and the dashed lines represent levels with $m=2$. The crossing points correspond to different magnetopolaron states both in the presence and in the absence of resonant coupling between levels. The filled circles (A and A') denote ordinary double magnetopolarons, in which levels of the electron–phonon system characterized by the same size-quantization number are involved.[14,15,18] Combined magnetopolarons are marked by open circles.[14,15,18] They correspond to different configurations of the quantum numbers m,m_1 and n,n_1. The numberss *1*, *2*, and *3* label combined magnetopolarons in the first, second, and third configurations, respectively. Figure 1 shows that in the first configuration levels $E^e_{2,1}$ and $E^e_{1,0}+\hbar\omega_{L1}$ cross one another (Fig. 1a), and in the second configuration (Fig. 1b) these are the levels $E^e_{2,0}$ and $E^e_{1,1}+\hbar\omega_{L1}$.

Conditions (5) and (6) are radically different from condition (1) for the double magnetopolaron, in which the resonant magnetic field is independent of the quantum well parameters. In the case of the combined magnetopolaron, $\Omega^{\rm res}_e$ is a function of both the depth and width of the well; in other words, there is a specific resonant magnetic field for each quantum well. In Fig. 1 there are crossings between levels of the electron–phonon system with equal numbers N of phonons. These polaron states are similar to weak polarons,[18] when the difference ΔN between phonon numbers in two states of the electron–phonon system does not equal unity. In this case, resonant transitions between levels with emission of one phonon are forbidden. In order to calculate anticrossing splittings, one has to take into account transitions via virtual states. In this situation, the degeneracy at crossing points is lifted, but the separation between the splitted levels should be of an order in α higher than $\alpha^{1/2}$. In what follows, such weak polarons will not be discussed.

To conclude this section, let us consider the case of resonant coupling between levels of an electron–phonon system at all values of magnetic field. This requires that the following condition be satisfied:

$$\hbar\omega_{L1}=\varepsilon^e_m-\varepsilon^e_{m_1}. \tag{14}$$

In this case, we have resonant coupling of two size-quantized levels with equal Landau numbers, i.e., $n=n_1$. Condition (6) can be satisfied over a wide range of quantum well parameters, since the difference of the energy difference between size-quantized levels from $\hbar\omega_{L1}$ can be compensated for by varying the magnetic field, whereas condition (14) can be satisfied only at certain quantum well parameters, and the magnetic field is needed only to form Landau levels, so it can be relatively weak.

3. THE SPECTRUM OF A COMBINED MAGNETOPOLARON

The energy spectrum of a combined magnetopolaron is determined, as is well known, by the poles of the one-particle electron Green's function. The corresponding equation has the form

$$\varepsilon-E^e_{m,n}-\Sigma(m,n,\varepsilon)=0, \tag{15}$$

where $E^e_{m,n}$ is given by Eq. (3) and $\Sigma(m,n,\varepsilon)$ is the mass operator. In single quantum wells in such structures as AlAs/GaAs/AlAs and AlSb/InSb/AlSb for quantum well widths $d>200$ Å, it suffices to take account of only confined phonons, since the interaction with interface phonons and half-space phonons can be neglected.[8,18] The Hamiltonian of interaction between an electron and phonons confined in the quantum well, which occupies the region $0\leqslant z\leqslant d$, can be expressed as[19]

$$H_c=\sum_{\mathbf{q}}\left[\sum_{p=1,3,\ldots} C_{\mathbf{q},p}\cos\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right](a_{\mathbf{q},p}+a^{+}_{-\mathbf{q},p})\right.$$
$$\left.+\sum_{p=2,4,\ldots} C_{\mathbf{q},p}\sin\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right](a_{\mathbf{q},p}+a^{+}_{-\mathbf{q},p})\right]e^{i\mathbf{q}\cdot\mathbf{r}},$$
$$0\leqslant z\leqslant d, \tag{16}$$
$$H_c=0,\quad z<0,\quad H_c=0,\quad z>d.$$

The parameter $C_{\mathbf{q},p}$ is given by

$$C_{\mathbf{q},p}=-\hbar\omega_{\mathrm{L}1}\frac{\sqrt{8\pi\alpha l^3/S_0 d}}{l\sqrt{q^2+(p\pi/d)^2}}, \tag{17}$$

where the nondimensional constant of coupling between the electron and confined phonon is defined as

$$\alpha=\frac{e^2}{2l\hbar\omega_{\mathrm{L}1}}(\varepsilon_{\infty 1}^{-1}-\varepsilon_{01}^{-1}),\quad l=\sqrt{\frac{\hbar}{2m_c\omega_{\mathrm{L}1}}}, \tag{18}$$

$\omega_{\mathrm{L}1}$ is the frequency of the confined phonon, whose dispersion is neglected, $\varepsilon_{01}(\varepsilon_{\infty 1})$ is the static (high-frequency) dielectric susceptibility of the quantum well material, $\mathbf{r}=\mathbf{r}(x,y)$ and $\mathbf{q}=\mathbf{q}(q_x,q_y)$ are the two-dimensional radius-vector of the electron and the phonon wave vector, respectively, $a^{+}_{\mathbf{q},p}(a_{\mathbf{q},p})$ is the operator creating (annihilating) a confined phonon with wave vector $\mathbf{q}$ and quantum number p, which is an analogue of the wave vector component q_z in the 3D configuration, and S_0 is the normalizing area.

If two levels of an electron–phonon system cross, it is sufficient to take into account the simplest Feynman diagram (two vertices connected by electron and phonon lines) in the mass operator. The contributions of diagrams with larger numbers of vertices are of higher orders in the coupling constant α, which is assumed to be small. Using the standard techniques, we obtain the following expression for the mass operator:

$$\Sigma(m,n,\varepsilon)=\sum_{m_1,n_1}\sum_{\mathbf{q},p}\frac{|\Phi_p(m,m_1,n,n_1,\mathbf{q})|}{\varepsilon-E^{e}_{m_1,n_1}-\hbar\omega_{\mathrm{L}1}+i\delta},$$
$$\delta\to 0,\quad p=1,2,3,\ldots, \tag{19}$$

$$\Phi_p(m,m_1,n,n_1,\mathbf{q})\equiv\Phi_p(\mathbf{q})=\frac{1}{2}C_{\mathbf{q},p}\{M^{(1)}_{m,m_1}(p)\times[1+(-1)^{p+1}]+M^{(2)}_{m,m_1}(p)\times[1+(-1)^{p}]\}I_{n,n_1}(\mathbf{q}).$$

In deriving Eq. (19), we have assumed that the temperature is low enough so that optical phonon states are unpopulated and the resonant interaction is due to emission of phonons. The following notation has been introduced in Eq. (19):

$$I_{n,n_1}(\mathbf{q})=\frac{1}{L_x}\int d\mathbf{r}\exp\{i(k-k_1)x+i\mathbf{q}\cdot\mathbf{r}\}\times\phi_n(y-y_k)\phi_{n_1}(y-y_{k_1}), \tag{20}$$

$$M^{(1)}_{m,m_1}(p)=\int_0^d dz\chi_m(z)\chi_{m_1}(z)\cos\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right],$$
$$p=1,3,\ldots,$$
$$M^{(2)}_{m,m_1}(p)=\int_0^d dz\chi_m(z)\chi_{m_1}(z)\sin\left[\frac{p\pi}{d}\left(z-\frac{d}{2}\right)\right], \tag{21}$$
$$p=2,4,\ldots$$

The wave function $\phi_n(y-y_k)$ describes the motion of an electron in the quantum well plane:

$$\phi_n(y)=\frac{1}{\sqrt{2^n n!\sqrt{\pi}R_0}}\exp\left(-\frac{y^2}{2R_0^2}\right)H_n\left(\frac{y}{R_0}\right), \tag{22}$$

where

$$R_0^2=c\hbar/|e|H,\quad y_k=-c\hbar k/eH, \tag{23}$$

k is the x component of the electron wave vector, $H_n(y)$ is the Hermitian polynomial. The electron motion in the perpendicular direction is described by the function $\chi_m(z)$:

$$\chi_m(z)=C_m(-1)^{(m-1)/2}\times\begin{cases}\cos(k_m d/2)\exp(\kappa_m z), & z\leqslant 0,\\ \cos[k_m(z-d/2)], & 0\leqslant z\leqslant d,\\ \cos(k_m d/2)\exp[-\kappa_m(z-d)], & z\geqslant 0\end{cases} \tag{24}$$

for $m=1,3,5,\ldots$ and

$$\chi_m(z)=C_m(-1)^{m/2}\times\begin{cases}-\sin(k_m d/2)\exp(\kappa_m z), & z\leqslant 0,\\ \sin[k_m(z-d/2)], & 0\leqslant z\leqslant d,\\ \sin(k_m d/2)\exp[-\kappa_m(z-d)], & z\geqslant 0\end{cases} \tag{25}$$

for $m=2,4,6,\ldots$ The normalizing constant is

$$C_m=\sqrt{\frac{2K_m}{1+K_m d\pm\cos(k_m d)\pm(K_m/k_m)\sin(k_m d)}}, \tag{26}$$

$$K_m=\sqrt{Q_e^2-k_m^2}, \tag{27}$$

the upper signs are applied to even m, the lower signs to odd m. The quantity k_m, and hence the energy ε^e_m of a level in the quantum well, are determined by Eqs. (7) and (8).

In the limit $k_m d\to m\pi$ of an infinitely deep well, $K_m\to\infty$, $C_m\to\sqrt{2/d}$, $\chi_m(z)\to\sqrt{2/d}\sin(m\pi z/d)$ if $0\leqslant z\leqslant d$, and $\chi_m(z)\to 0$ for $z\leqslant 0$ and $z\geqslant d$. By substituting wave functions (22) in Eq. (20) for $I_{n,n_1}(\mathbf{q})$, we obtain

$$|I_{n,n_1}(\mathbf{q})|^2=\frac{\min(n!,n_1!)}{\max(n!,n_1!)}u^{|n-n_1|}e^{-u}[L^{|n-n_1|}_{\min(n,n_1)}(u)]^2, \tag{28}$$

where

$$u=l_H^2q^2,\quad l_H^2=\frac{c\hbar}{2|e|H}=\frac{R_0^2}{2},$$

$L^s_n(u)$ is the adjoint Laguerre polynomial.

Similarly, using the functions $\chi_m(z)$ from Eqs. (24) and (25), we obtain the expressions for $M^i_{m,m_1}(p)$ $(i=1,2)$

$$M^{(1)}_{m,m_1}(p)=C_mC_{m_1}\pi pd(-1)^{(p-1)/2}\times\left[\pm\frac{\cos(q_1/2)}{p^2\pi^2-q_1^2}+\frac{\cos(q_2/2)}{p^2\pi^2-q_2^2}\right], \quad (29)$$

$p=1,3,5,\ldots$, the plus sign applies to odd m and m_1 and the minus sign to even m and m_1,

$$M^{(2)}_{m,m_1}(p)=C_mC_{m_1}\pi pd(-1)^{p/2}\times\left[-\frac{\sin(q_1/2)}{p^2\pi^2-q_1^2}\pm\frac{\sin(q_2/2)}{p^2\pi^2-q_2^2}\right], \quad (30)$$

$p=2,4,6,\ldots$, the plus sign applies to odd m and even m_1, the minus sign to even m and odd m_1. In Eqs. (29) and (30) we have introduced the notation

$$q_1=(k_m+k_{m_1})d, \quad q_2=(k_m-k_{m_1})d, \quad (31)$$

where k_m and k_{m_1} are solutions of either the first or the second equations in (7), depending on the parities of m and m_1. For an infinitely deep quantum well

$$M^{(1)}_{m,m_1}(p)=\pm\frac{2p}{\pi}(-1)^{(p+m+m_1-1)/2}\times\left[\frac{1}{p^2-(m+m_1)^2}-\frac{1}{p^2-(m-m_1)^2}\right] \quad (32)$$

(the plus is used for odd m and m_1, the minus for even m and m_1), and

$$M^{(2)}_{m,m_1}(p)=\frac{2p}{\pi}(-1)^{(p+m+m_1-1)/2}\times\left[-\frac{1}{p^2-(m+m_1)^2}+\frac{1}{p^2-(m-m_1)^2}\right] \quad (33)$$

for m and m_1 with different parities.

Using Eqs. (7) and (28), it is convenient to express the mass operator as

$$\Sigma(m,m_1,n,n_1,\varepsilon)=\sum_{m_1,n_1}\frac{w_c(m,m_1,n,n_1)}{\varepsilon-E_{m_1,n_1}-\hbar\omega_{\mathrm{L1}}+i\delta}, \quad (34)$$

where the function $w_c(m,m_1,n,n_1)$ is given by the expression

$$w_c(m,m_1,n,n_1)=\frac{\alpha}{2}(\hbar\omega_{\mathrm{L1}})^2\sqrt{\frac{\Omega_c}{\omega_{\mathrm{L1}}}\frac{\min(n!,n_1!)}{\max(n!,n_1!)}}\times\left[\int_0^\infty duu^{|n-n_1|-1/2}e^{-u}[L^{|n-n_1|}_{\min(n,n_1)}(u)]^2\times F^c_{m,m_1}(\beta_0\sqrt{u})\right]^{1/2}, \quad \beta_0=\frac{d}{l_H}, \quad (35)$$

$$F^c_{m,m_1}(x)=\sum_{p=1,3,\ldots}\frac{[M^{(1)}_{m,m_1}(p)]^2}{x^2+p^2\pi^2}+\sum_{p=2,4,\ldots}\frac{[M^{(2)}_{m,m_1}(p)]^2}{x^2+p^2\pi^2}. \quad (36)$$

The mass operator (34) is a sum over the quantum numbers m and n_1. If the resonant condition (6) is satisfied, it contains one large term in which the denominator $\varepsilon-E^e_{m_1,n_1}-\hbar\omega_{\mathrm{L1}}$ is small. This term corresponds to an electronic transition from level m,n to the level with specified m_1 and n_1 through emission of an LO phonon. The rest of the terms are small, since they are proportional to the coupling constant $\alpha\ll 1$, and their denominators are far from zero. Thus, the mass operator is expressed as

$$\Sigma(m,n,\varepsilon)\equiv\Sigma(m,m_1,n,n_1,\varepsilon)=\frac{w_c(m,m_1,n,n_1)}{\varepsilon-E^e_{m,n}+\lambda+i\delta}, \quad (37)$$

where

$$\lambda=\varepsilon^e_m-\varepsilon^e_{m_1}+\hbar\Omega_e(n-n_1)-\hbar\omega_{\mathrm{L1}} \quad (38)$$

has the sense of the difference between the resonant magnetic field at fixed d and the actual field. By substituting Eq. (37) in (5), which describes the spectrum, we obtain a quadratic equation for the variable $\varepsilon-E^e_{m,n}$, whose solution determines two branches of the combined magnetopolaron spectrum. The separation between the spectral branches is given by the formula

$$\Delta E(\lambda)=\sqrt{\lambda^2+4w_c(m,m_1,n,n_1)}. \quad (39)$$

At the point of the exact resonance $\lambda=0$ and

$$\Delta E_{\mathrm{res}}=2\sqrt{w_c(m,m_1,n,n_1)}. \quad (40)$$

Alongside ΔE_{res}, it is interesting to calculate the resonant magnetic field H_{res} as a function of the well width d. The expression for H_{res} is derived from Eqs. (6) and (8), and for the polaron in the first configuration it has the form

$$H^{(1)}_{\mathrm{res}}=H^{(0)}_{\mathrm{res}}\left\{1-\frac{l^2}{d^2}[(k_md)^2-(k_{m_1}d)^2]\right\},$$

where l is determined by Eq. (18), k_m and k_{m_1} by the solution of Eq. (7), and

$$H^{(0)}_{\mathrm{res}}=\frac{m_cc\omega_{\mathrm{L1}}}{|n-n_1||e|}$$

is the resonant magnetic field determined by condition (1) for the double polaron.

For the polaron in the third configuration,

$$H^{(3)}_{\mathrm{res}}=H^{(0)}_{\mathrm{res}}\left\{1+\frac{l^2}{d^2}[(k_{m_1}d)^2-(k_md)^2]\right\},$$

$H^{(1)}_{\mathrm{res}}$ and $H^{(3)}_{\mathrm{res}}\to H^{(0)}_{\mathrm{res}}$ in the limiting case of large d. When the well width d is diminished, $H^{(1)}_{\mathrm{res}}$ drops and $H^{(3)}_{\mathrm{res}}$ increases.

For the polaron in the second configuration,

$$H^{(2)}_{\mathrm{res}}=H^{(0)}_{\mathrm{res}}\left\{\frac{l^2}{d^2}[(k_md)^2-(k_{m_1}d)^2]-1\right\}, \quad n_1>n.$$

Here the field $H^{(2)}_{\mathrm{res}}$ is low in the region close to $d_{\min}(m)$, at which point the polaron disappears, and becomes higher as d approaches $d'_{\min}(m)$. In GaAs, the resonant field for transition $n=1\to n_1=0$ is $H^{(0)}_{\mathrm{res}}=20.8$ T. It can be reduced severalfold if one takes a larger difference $n-n_1$.

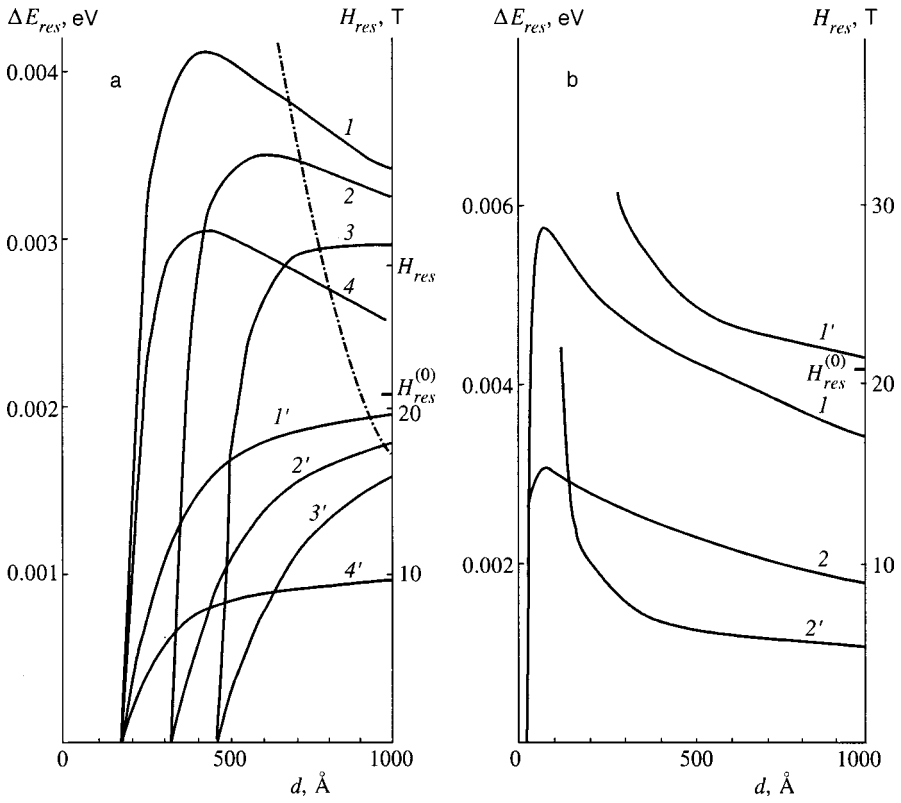


FIG. 2. Separations $\Delta E_{\rm res}$ between branches of the spectrum of combined magnetopolaron versus the quantum well width d (curves *1–4*) for an $Al_{0.32}Ga_{0.68}As/GaAs/Al_{0.32}Ga_{0.68}As$ structure ($\Delta E_{\rm res}=0.35$ eV, $\hbar\omega_{\rm L1}=0.036$ eV, $m_c/m_0=0.067$, $\alpha=0.07$). Curves $1'-4'$ plot the resonant magnetic field $H_{\rm res}$ calculated by Eq. (6) (the right-hand scale) for curves *1–4*, respectively; (a) polaron of the first configuration; (*1*) transition $m=2$, $n=1\to m=1$, $n=0$; (*2*) $m=3$, $n=1\to m=1$, $n=0$; (*3*) $m=4$, $n=1\to m=1$, $n=0$; (*4*) $m=2$, $n=2\to m=1$, $n=0$; the dash-dotted line plots $\varepsilon_2^e(d)-\varepsilon_1^e(d)$; (b) polaron in the third configuration: (*1*) $m=1$, $n=1\to m=2$, $n=0$; (*2*) $m=1$, $n=4\to m=2$, $n=0$.

Curves of $\Delta E_{\rm res}$ as a function of the quantum well width d calculated for the structures $Al_{0.32}Ga_{0.68}As/GaAs/Al_{0.32}Ga_{0.68}As$ are plotted in Fig. 2. Each $\Delta E_{\rm res}$ curve has a corresponding curve of the resonant magnetic field $H_{\rm res}$.

Figure 2a (the case of $\hbar\omega_{\rm L1}\geqslant\varepsilon_m^e-\varepsilon_{m_1}^e$) shows curves for the polaron in the first configuration ($m>m_1$, $n>n_1$). The curves of $\Delta E_{\rm res}$ rapidly drop to zero around the point $d_{\min}(m)$, and the polaron does not exist there. As the quantum number m increases, the point $d_{\min}(m)$ shifts to the side of larger d. So, we have $d_{\min}(2)\simeq 180$ Å for curve *1* in Fig. 2a (this curve corresponds to the combined polaron *1* in Fig. 1a), $d_{\min}(3)\simeq 320$ Å for curve *2*, and $d_{\min}(4)\simeq 460$ Å for curve *3*. Since $H_{\rm res}\propto 1/(n-n_1)$, the increase in the number n of the initial Landau level n at fixed n_1 leads to a considerable decrease in $H_{\rm res}$ (curves $1'$ and $4'$ in Fig. 2a and curves $1'$ and $2'$ in Fig. 2b).

Figure 2b shows curves for the polaron of the third configuration, when the domain of the curve of $\Delta E_{\rm res}$ is limited to $d'_{\min}(m)$, where the upper level is ejected from the quantum well. On curve *1*, which corresponds to polaron *3* in Fig. 1, $d'_{\min}(2)\simeq 40$ Å. The polaron of the second configuration exists in a relatively narrow range of d determined by Eq. (13). In the case of the transition $m=2$, $n=0\to m_1=1$, $n_1=1$, this range extends from $d'_{\min}(2)\simeq 40$ Å to $d_{\min}(2)\simeq 180$ Å. This range broadens with the number m. So, $d'_{\min}(3)\simeq 80$ Å, $d_{\min}(3)\simeq 320$ Å for $m=3$ and $d'_{\min}(4)\simeq 120$ Å, $d_{\min}(4)\simeq 460$ Å for $m=4$. The field $H_{\rm res}$ is high near the lower limit (30.1 T for $d=200$ Å, $m=3$, $n_1=1\to m_1=1$, $n=0$) and drops rapidly as d approaches $d_{\min}(m)$ (3.84 T for $d=300$ Å). Here $\Delta E_{\rm res}$ is of the same order as the value given in Fig. 2.

In the case when the resonant condition (14) holds, in the mass operator (19) we set $n=n_1$. As a result, we have the expression

$$\Sigma(m,n,\varepsilon)=\sum_{m_1}\frac{w_c(m,m_1,n)}{\varepsilon-E^e_{m_1,n}-\hbar\omega_{\rm L1}+i\delta}, \tag{41}$$

where we can retain only the largest term corresponding to the small denominator on the right-hand side. The function $w_c(m,m_1,n)$ in this case equals

$$w_c(m,m_1,n)=\frac{\alpha}{2}(\hbar\omega_{\rm L1})^2\frac{ld}{l_H^2}\times\int_0^\infty du e^{-u}[L_n(u)]^2F'_{m,m_1}(\beta_0\sqrt{u}), \tag{42}$$

where $L_n(u)$ is the Laguerre polynomial and $F'_{m,m_1}(\beta_0\sqrt{u})=F^c_{m,m_1}/\beta_0$ [see Eq. (36)]. The gap between the two branches of the combined magnetopolaron spectrum in this case is

$$\Delta E(\lambda')=\sqrt{\lambda'^2+4w_c(m,m_1,n)}, \tag{43}$$

where

$$\lambda'=\hbar(\omega^e_m-\omega^e_{m_1})-\hbar\omega_{\rm L1} \tag{44}$$

determines the detuning from the resonance due to the deviation of d from the value determined by condition (14). In the case of fine tuning

$$\Delta E_{\rm res}=2\sqrt{w_c(m,m_1,n)}. \tag{45}$$

It follows from Eq. (45) that in the structure $Al_{0.32}Ga_{0.68}As/GaAs/Al_{0.32}Ga_{0.68}As$ a splitting energy of about 2×10^{-3} eV is achieved at magnetic field $H\simeq3.86$ T. The value $d_{\rm res}$ that satisfies condition (14) depends sensitively on the quantum number m. For example, we have $d_{\rm res}=191$ Å for the transition $m=2\to m_1=1$, $d_{\rm res}=329$ Å for $m=3\to m_1=1$, and $d_{\rm res}=460$ Å for $m=4\to m_1=1$. The splitting $\Delta E_{\rm res}$ is considerably smaller at Landau numbers $n>0$ owing to the oscillating nature of the function $\phi_n(y)$ [Eq. (22)]. The shape of $\Delta E_{\rm res}$ as a function of magnetic field is controlled by the function $w_c(m,m_1,n)$, whose magnetic field dependence, as one can see in Eq. (42), is determined by the parameters β_0 and l/l_H.

The magnetopolaron spectrum is described by the simple formula

$$\varepsilon^{\pm}=E^e_{m,n}-\lambda/2\pm\sqrt{(\lambda/2)^2+w_c}, \tag{46}$$

where the plus sign applies to the upper branch and the minus to the lower branch. Here w_c is determined by either the function $w_c(m,m_1,n,n_1)$ in Eq. (35) or the function $w_c(m,m_1,n)$ in Eq. (42). The form of λ is determined according to (38) or (44).

Consider first the case when the resonant condition (5) is satisfied. In a real system, there is always an interaction which makes a state unsteady and adds an imaginary component to the energy. The cause of instability of a polaron state can be the instability of the electronic state or the phonon. The mechanism for instability of the electronic state can be two-phonon scattering, when the electron emits two acoustic phonons and makes a transit from level $n=1$ to $n=0$. The phonon instability is due to the decay of an LO phonon into two acoustic phonons caused by the lattice anharmonism. Let us discuss the phonon instability mechanism in detail. As was noted above, it suffices to take into account the simplest diagram in calculating the mass operator. The LO phonon decay is taken into account by renormalizing the zero-order Green's function of the LO phonon, i.e., a vertex describing the LO phonon decay into two acoustic phonons should be introduced (three-phonon interaction). Acoustic phonons in a quantum well are characterized by their two-dimensional wave vectors $\mathbf{q}$ in the quantum well plane and a discrete quantum number p_1, which is an analog of the wave vector component q_z, as in the case of a confined LO phonon. Let us suppose that the three-phonon interaction is small in comparison with the interaction between electrons and LO phonons. In this case, the correction to the LO phonon Green's function is calculated by the perturbation theory. The equation for the polaron spectrum takes the form

$$\varepsilon-E^e_{m,n}-\frac{w_c}{\varepsilon-E^e_{m,n}+\lambda}+\frac{i\hbar\gamma(\varepsilon)w_c}{(\varepsilon-E^e_{m,n}+\lambda)^2}=0, \tag{47}$$

and the last term on the left-hand side is assumed to be small in comparison with the other two with respect to the parameter $\hbar\gamma(\varepsilon)/\sqrt{w_c}$. The quantity $\hbar\gamma(\varepsilon)$ determines the instability of the magnetopolaron spectrum. It can be conveniently expressed as

$$\hbar\gamma(\varepsilon)=\frac{\pi}{w_c}\sum_{\mathbf{q},p_1,p_2}\Phi_p(\mathbf{q})\Phi^*_{p'}(\mathbf{q})\,\hbar\gamma_{\rm ph}(\mathbf{q}). \tag{48}$$

The function $\gamma_{\rm ph}(\mathbf{q})$ in the sum in Eq. (48) is the inverse LO phonon lifetime due to its decay into two acoustic phonons:

$$\hbar\gamma_{\rm ph}(\mathbf{q})=\sum_{\mathbf{q},p_1,p_2}C_{p,p_1,p_2}(\mathbf{q},\mathbf{q}_1)C^*_{p',p_1,p_2}(\mathbf{q},\mathbf{q}_1)\times\delta(\varepsilon-E^e_{m_1,n_1}-\hbar\omega_{p_1}(\mathbf{q}_1)-\hbar\omega_{p_2}(\mathbf{q}-\mathbf{q}_1)). \tag{49}$$

Here $C_{p,p_1,p_2}(\mathbf{q},\mathbf{q}_1)$ is the three-phonon vertex and $\omega_p(\mathbf{q})$ is the frequency of the confined acoustic phonon. In deriving Eqs. (48) and (49) we have taken into account the conservation of the two-dimensional momentum $\mathbf{q}$, whereas no conservation law applies to discrete quantum numbers p. This explains why the number of sums over p is twice the number of sums over $\mathbf{q}$. Together with the level broadening, the three-phonon interaction brings about their shift, which is neglected here owing to its smallness. Equation (47) is solved by the iteration method. As a result, we obtain a formula for the two magnetopolaron branches:

$$\varepsilon^{(\pm)}=E^e_{m,n}-\frac{\lambda}{2}\pm\sqrt{\left(\frac{\lambda}{2}\right)^2+w_c}-\frac{i\hbar\gamma^{(\pm)}}{2}, \tag{50}$$

where

$$\gamma^{(\pm)}=\gamma\left[1\mp\frac{\lambda/2}{\sqrt{(\lambda/2)^2+w_c}}\right] \tag{51}$$

is the inverse lifetime of the polaron spectrum branches due to the LO phonon decay (the upper signs are for the upper branch, and the lower signs are for the lower one). The parameter γ in Eq. (51) is

$$\gamma(\varepsilon)=-\lambda/2\pm\sqrt{(\lambda/2)^2+w_c},$$

and, strictly speaking, it should be different for different spectrum branches. As one can see in Eqs. (48) and (49), $\gamma(\varepsilon)$ is a function of ε owing to the function $\gamma_{\rm ph}(\varepsilon)$, where ε is included in the argument of the δ-function. In the zero-order approximation in ε, which should be used in this case,

$$\varepsilon-E^e_{m_1,n_1}=\lambda/2\pm\sqrt{(\lambda/2)^2+w_c}+\hbar\omega_{\rm L1}\simeq\hbar\omega_{\rm L1}, \tag{52}$$

since $\sqrt{w_c}\ll\hbar\omega_{\rm L1}$. Thus, $\gamma_{\rm ph}(\mathbf{q})$ is almost constant for all electronic states, and $\gamma(\varepsilon)$ is approximately the same for all branches of the polaron spectrum.

The parameters $\gamma^{(\pm)}$ sensitively depend on both the magnetic field and electron quantum numbers. At the resonance points $\lambda=0$ and $\gamma^{(+)}=\gamma^{(-)}=\gamma$. For $\lambda>0$ and $\lambda\gg\sqrt{w_c}$, $\gamma^{(+)}\to0$, since in this limiting case the upper branch of the magnetopolaron spectrum is a purely electronic state characterized by quantum numbers m and n, which cannot be affected by the phonon lifetime. The lower level in this case contains an electron with quantum numbers m_1 and n_1, and its decay rate is controlled by the phonon lifetime, soin this limit we have $\gamma^{(-)}\to\gamma$. With the opposite sign of λ, the situation is reversed: the damping of the lower branch vanishes, and the decay rate of the upper branch is determined by γ, the quantum numbers m and n, and the phonon lifetime, therefore we have in this limit $\gamma^{(-)}\to\gamma$.

4. INTERBAND ABSORPTION AND REFLECTION SPECTRA OF THE QUANTUM WELL

Since the electron and hole motion in type I quantum wells is quantized in the direction perpendicular to the quantum well plane, the spectrum of interband absorption of light in a strong magnetic field due to transitions between discrete levels consists of narrow lines. In a a strong magnetic field in III–V materials, which are considered as a specific example in this paper, each level of hole size-quantization is associated with four sets of Landau levels, two of which are related to the heavy holes and two to the light ones. The wave functions of the heavy holes, which are considered here, are known to have at the center of the Brillouin zone the form

$$\Psi_{1(2)}=(X\pm iY)\Psi_{\uparrow}+(X-iY)\Psi_{\downarrow}. \tag{53}$$

In the conduction band, there are also two wave functions with different spin orientations, namely, $u_c\Psi_{\uparrow}$ and $u_c\Psi_{\downarrow}$. Here $\Psi_{\uparrow}$ and $\Psi_{\downarrow}$ are the spin-dependent wave functions and u_c is the fast oscillating Bloch function of the conduction band, X and Y are the Bloch functions of the valence band. The momentum matrix elements relating states $\Psi_{1(2)}$ and $u_c\Psi_{\uparrow(\downarrow)}$ are equal and expressed by

$$\mathbf{p}_{cv}=\frac{1}{\sqrt{2}}p_{cv}(\mathbf{e}_x+i\mathbf{e}_y), \tag{54}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors aligned with the x and y axes, and the constant

$$p_{cv}=-i\hbar\int_{\text{cell}}d\mathbf{r}X^*\frac{d}{dx}u_c=-i\hbar\int_{\text{cell}}d\mathbf{r}Y^*\frac{d}{dy}u_c$$

can be chosen to be real.

The electron and hole discrete levels are broadened by various interactions in the quantum well. The broadening of absorption lines is determined by both homogeneous and inhomogeneous mechanisms, which are beyond the scope of this paper. Therefore, our discussion is limited to the mechanism of phonon damping of magnetopolaron states described in Sec. 3. The convenience of this mechanism (from the theoretical viewpoint) is that it broadens either electron or hole states because the resonant condition (5) or (14), which determine the region where the magnetopolaron exists, cannot be satisfied for both electrons and holes at the same time. When an electron–hole pair is generated, there is always a probability of its annihilation, and this process leads to radiative damping, which is determined by γ_r. If the inequality

$$\gamma_r\ll\gamma \tag{55}$$

holds, the absorption can be calculated in the approximation linear in the intensity of interaction with the electromagnetic wave.

Since in the accepted model hole levels are not broadened, the absorption at low temperatures is known to be proportional to $i\mathrm{Re}G_{\gamma}(\Gamma)$, where $G_{\gamma}(\Gamma)$ is the retarded one-particle electron Green's function, which is expressed as

$$G_r(\Gamma)=\hbar\left[\Gamma-\frac{w_c}{\Gamma+\lambda}+\frac{iw_c\,\hbar\gamma}{(\Gamma+\lambda)^2}\right]^{-1}, \tag{56}$$

where

$$\Gamma=\hbar\omega_l-E^h_{m_h,n}-E^e_{m_e,n}, \tag{57}$$

ω_l is the frequency of the incident electromagnetic wave, and γ is determined by Eq. (48). It is convenient to characterize the light absorption by the fraction A of absorbed energy, which is a nondimensional parameter equal to the absolute value of energy absorbed in the quantum well material divided by the incident flux of energy in the barrier. In the dipole approximation and given that condition (55) holds, A is simply related to the Green's function:

$$A=\gamma_r\,\mathrm{Re}[iG_r(\Gamma)], \tag{58}$$

where

$$\gamma_r=\frac{2e^2}{n_0\hbar c}\frac{p_{cv}^2}{m_0^2}\frac{|e|H}{\hbar c}\frac{\langle m_h|m_e\rangle}{\omega_g+\omega^h_{m_h,n}+\omega^e_{m_e,n}}. \tag{59}$$

Here $\langle m_h|m_e\rangle$ is the overlapping integral between the electron wave function χ_{m_e} and the similar hole wave function χ_{m_h}, $\omega_g+\omega^h_{m_h,n}+\omega^e_{m_e,n}$ is the frequency of the interband transition between levels m_h,n and m_e,n; n_0 is the refraction index of the barrier material. In this section we are using the quantum numbers m_e and m_h for electrons and holes instead of the quantum number m. In deriving Eq. (58), we have neglected the difference between the refraction indices of the barrier and quantum well and assumed that the barrier occupies the entire space except the region $0\leqslant z\leqslant d$. The light wave is also assumed to be incident normally on the quantum well plane and circularly polarized.

The function $\mathrm{Re}[iG_r(\Gamma)]$ describes two peaks of the interband absorption due to the transition from the valence band level with quantum numbers m_h,n to the two branches of the magnetopolaron spectrum in the conduction band. Consider the shape of the absorption spectrum in the same approximation that was used in calculating the spectrum (48) of the damped magnetopolaron. The quantity $\mathrm{Re}[iG_r(\Gamma)]$ can be expressed as a sum of two terms, each of which describes a Lorentzian. The expression for the fraction A of absorbed energy is

$$A(\Gamma)=2[\hbar\gamma_r^{(+)}\Delta^{(+)}+\hbar\gamma_r^{(-)}\Delta^{(-)}], \tag{60}$$

$$\Delta^{(\pm)}=\mathrm{Re}\frac{i}{\Gamma-\Gamma^{(\pm)}+i\hbar\gamma^{(\pm)}/2}$$

$$=\frac{\hbar\gamma^{(\pm)}/2}{(\Gamma-\Gamma^{(\pm)})^2+(\hbar\gamma^{(\pm)}/2)^2}, \tag{61}$$

$$\Gamma^{(\pm)}=(-\lambda\pm\sqrt{\lambda^2+4w_c})/2. \tag{62}$$

The parameter $\gamma_r^{(\pm)}$ is the radiative decay rate γ_r renormalized by the magnetopolaron state:

$$\gamma_r^{(\pm)}=\gamma_r\left[1\pm\frac{\lambda}{\sqrt{\lambda^2+4w_c}}\right]. \tag{63}$$

For $\lambda=0$ (accurate tuning to the resonance)

$$\Gamma^{(\pm)}=\pm\sqrt{w_c},\quad \gamma^{(\pm)}=\gamma,\quad \Delta^{(\pm)}=\frac{\gamma/2}{(\Gamma\mp\sqrt{w_c})^2+(\gamma/2)^2},$$

i.e., we have two Lorentzians with peaks at points $\Gamma=\pm\sqrt{w_c}$. The separation between the peak tops equals that between the magnetopolaron branches at point $\lambda=0$. For $\lambda>0$ and $\lambda\gg 2\sqrt{w_c}$ we have $\gamma_r^{(+)}\to\gamma_r$ and $\gamma_r^{(-)}\to 0$. The cause of this behavior of $\gamma_r^{(+)}$ is that the state on the upper branch transforms to the $n=1$ state at larger λ. At the same time, $\gamma_r^{(-)}$ drops, since the $n=0$ state becomes dominant on the lower branch of the spectrum and the interband transition to this branch is forbidden. The situation is reversed in the region $\lambda<0$: $\gamma_r^{(+)}\to 0$ and $\gamma_r^{(-)}\to\gamma_r$. The ratio

$$\frac{A(\Gamma^{(+)})}{A(\Gamma^{(-)})}=\frac{(\lambda+\sqrt{\lambda^2+4w_c})^2}{(-\lambda+\sqrt{\lambda^2+4w_c})^2} \tag{64}$$

increases as we move into the region $\lambda>0$, i.e., if the magnetic field is higher than the resonant value, the peak corresponding to $\Gamma=\Gamma^{(+)}$ dominates. In magnetic fields below the resonance ($\lambda<0$), the peak corresponding to $\Gamma=\Gamma^{(-)}$ dominates.

The condition $\lambda<0$ signifies that the linear approximation fails at large $|\lambda|$. If the resonant condition (14) is satisfied, the absorption is also determined by Eqs. (60)–(63) and (51) with w_c expressed by Eq. (42) and λ' from Eq. (44) substituted for λ. The replacement of λ with λ' radically changes the magnetic field dependence of A. Specifically, since λ' does not vary with the magnetic field, varying the magnetic field at a fixed well width d should not increase one peak amplitude and decrease the other, but they will change similarly.

Over recent years, the interest in studies of optical reflection spectra of low-dimensional systems has increased, and this technique has proved an efficient tool for investigating their electronic properties.[20–23] This section considers reflection of a plane, monochromatic wave of the optical range normally incident on the quantum well plane from the side of negative z. The light wavelength is assumed to be much larger than the well width. The light reflection is an effect of higher order in γ_r, and the approximation linear in γ_r is insufficient for calculations. Let us introduce a nondimensional transmissivity T, which equals the absolute value of the energy flux after of the quantum well (on the right-hand side of the well) divided by the incident energy flux before the well (on the left-hand side of the well). The perturbation series in terms of the interaction with the optical wave reduces to a geometric progression, whose sum yields for T the formula

$$T=\frac{1}{(1+a)^2+b^2}, \tag{65}$$

where $a=4\pi\mathrm{Re}\chi(\omega_1)$, $b=4\pi\mathrm{Im}\chi(\omega_1)$, and $\chi(\omega_1)$ is the dielectric susceptibility of the well material. If there is a fixed resonant frequency ω_0, then $\chi(\omega_1)$ is expressed by the well-known formula $\chi(\omega_1)=1/[4\pi(\omega-\omega_0+i\gamma/2)]$ (we neglect the nonresonant component $1/[4\pi(\omega-\omega_0-i\gamma/2)]$). Then

$$a=a_0=\frac{\gamma_r\gamma/4}{(\omega_l-\omega_0)^2+(\gamma/2)^2},$$

$$b=b_0=\frac{(\gamma_r/2)(\omega_l-\omega_0)}{(\omega_l-\omega_0)^2+(\gamma/2)^2}. \tag{66}$$

The parameter a for the magnetopolaron state is determined by Eq. (60):

$$a=A(\Gamma)/2=\hbar\gamma_r^{(+)}\Delta^{(+)}+\hbar\gamma_r^{(-)}\Delta^{(-)}, \tag{67}$$

and the parameter b, by analogy with Eq. (66), is

$$b=\hbar\gamma_r^{(+)}\Delta_1^{(+)}+\hbar\gamma_r^{(-)}\Delta_1^{(-)}, \tag{68}$$

where

$$\Delta_1^{(\pm)}=\mathrm{Re}\,\frac{1}{\Gamma-\Gamma^{\pm}+i\gamma^{(\pm)}}=\frac{\Gamma-\Gamma^{(\pm)}}{(\Gamma-\Gamma^{(\pm)})^2+(\hbar\gamma^{(\pm)}/2)^2}. \tag{69}$$

Let us consider, along with A and T, the fraction of energy reflected back to the region of negative z. These quantities are related by the obvious equation

$$A+R+T=1. \tag{70}$$

If we reject condition (55), the absorption A, after summing the perturbation series in terms of interaction with light, takes the form

$$A=\frac{2a}{(1+a^2)^2+b^2}, \tag{71}$$

and, according to Eqs. (65) and (70),

$$R=\frac{a^2+b^2}{(1+a^2)^2+b^2}. \tag{72}$$

By substituting a from Eq. (67) and b from Eq. (68) in (71) and (72), we obtain exact expressions, which hold for all ratios between γ_r and γ:

$$A=\frac{2}{Z}\left\{\frac{\hbar\gamma_r^{(+)}}{2}\frac{\hbar\gamma^{(+)}}{2}\left[(\Gamma-\Gamma^{(-)})^2+\left(\frac{\hbar\gamma^{(-)}}{2}\right)^2\right]+\frac{\hbar\gamma_r^{(-)}}{2}\frac{\hbar\gamma^{(-)}}{2}\left[(\Gamma-\Gamma^{(+)})^2+\left(\frac{\hbar\gamma^{(+)}}{2}\right)^2\right]\right\}, \tag{73}$$

$$R=\frac{1}{Z}\left\{\left[\frac{\hbar\gamma_r^{(+)}}{2}(\Gamma-\Gamma^{(-)})+\frac{\hbar\gamma_r^{(-)}}{2}(\Gamma-\Gamma^{(+)})\right]^2+\left[\frac{\hbar\gamma_r^{(+)}}{2}\frac{\hbar\gamma^{(-)}}{2}+\frac{\hbar\gamma_r^{(-)}}{2}\frac{\hbar\gamma^{(+)}}{2}\right]^2\right\}, \tag{74}$$

where

$$Z=[(\Gamma-\Gamma^{(+)})^2+(\hbar\gamma^{(+)}+\hbar\gamma_r^{(+)})^2/4]\times[(\Gamma-\Gamma^{(-)})^2+(\hbar\gamma^{(-)}+\hbar\gamma_r^{(-)})^2/4]. \tag{75}$$

In the limit $\gamma^{(\pm)}\gg\gamma_r^{(\pm)}$, expression (73) for A transforms to (60) and R is quadratic in γ_r. The ratio J between the peak intensity $\Gamma=\Gamma^{(+)}$ and the peak intensity $\Gamma=\Gamma^{(-)}$ in the approximation $\gamma,\gamma_r<2\sqrt{w_c}$ (when the peak widths are smaller than the separation between them) is the same for A and R:

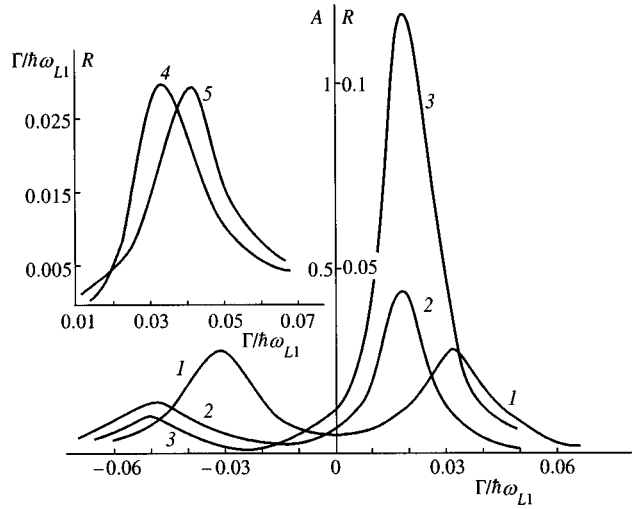


FIG. 3. Intensities of interband absorption, A (curves *1* and *2*), and reflection, R (curves *3–5*), of light at two values of λ' in Eq. (14) in the case of $m=2$, $m_1=1$, $n=0$ for the same structure as in Fig. 2. Curves *1*, *2*, and *4* plot the data for $H=4.8$ T, $w_c(2,1,0)=2.26\times10^{-3}$ eV, $\lambda'=0$; curve *3* for $\lambda'/2\sqrt{w_c(2,1,0)}=0.5$; and curve *5* for $\lambda'=0$, $H=9.6$ T, $w_c(2,1,0)=3.18\times10^{-3}$ eV.

$$J=\frac{[\lambda+\sqrt{\lambda^2+4w_c}+(\gamma_r/\gamma)(-\lambda+\sqrt{\lambda^2+4w_c})]^2}{[-\lambda+\sqrt{\lambda^2+4w_c}+(\gamma_r/\gamma)(\lambda+\sqrt{\lambda^2+4w_c})]^2}. \quad (76)$$

In the case of a resonance described by condition (5), the detuning λ is determined by Eq. (38), and $w_c=w_c(m,m_1,n,n_1)$ [Eq. (35)]. In the case of condition (14), λ is replaced by λ' and $w_c=w_c(m,m_1,n)$ [Eq. (42)]. If $\gamma=\gamma_r$, then $J=1$ holds for all λ or λ'. If $\gamma\neq\gamma_r$, then $J=1$ holds only at the point $\lambda(\lambda')=0$.

The magnetic field dependence of J is determined by the resonance condition. First consider the detuning from condition (5). If we have $\gamma>\gamma_r$, then in the region $\lambda>2\sqrt{w_c}(\lambda>0)$ we have $J\to(\gamma/\gamma_r)^2>1$, i.e., the right-hand peak $\Gamma=\Gamma^{(+)}$ dominates in the spectrum. For $\lambda<0$ we have $|\lambda|>2\sqrt{w_c}$, $J\to(\gamma_r/\gamma)^2$ and the left-hand peak $\Gamma=\Gamma^{(-)}$ is dominant. If $\gamma<\gamma_r$ holds, then, on the contrary, for $\lambda>0$ the left-hand peak dominates, and for $\lambda<0$ the right-hand peak is higher. When the frequency is detuned from condition (14), the limiting cases for J are the same, but the parameters λ' and $w_c=w_c(m,m_1,n)$ vary with d at fixed magnetic field H. If d is fixed, the dependence of J on H is weaker and is controlled by the H-dependence of function $w_c=w_c(m,m_1,n)$ and constant γ_r given by Eq. (59).

Figure 3 shows curves of A and R versus $\Gamma/\hbar\omega_{L1}$ for two values of λ' in condition (14). Curves *1* and *2* plot parameters of absorption spectra, curves *3–5* refer to reflection spectra. In the case of an exact resonance, $\lambda'=0$, curves *1*, *4*, and *5* are symmetrical about the point $\Gamma=0$ (the inset shows peaks corresponding to $\Gamma>0$). If $\lambda'>0$, the right-hand peak dominates (curves *2* and *3*). In the case of resonance condition (14), the parameters A and R are slower functions of magnetic field than in the case of condition (5). This is illustrated by the inset to Fig. 3, where curve *5* corresponds to $H=6$ T, i.e., the magnetic field twice as high as for curve *4*. Curves *1–4* can be used for describing A and R versus frequency in the case of condition (5) if only we substitute $w_c=w_c(m,m_1,n)$ for $w_c=w_c(m,m_1,n,n_1)$ and λ' for λ.

5. DISCUSSION

The main conclusion that follows from the results described above is that the separation between the two branches of the combined magnetopolaron is within 2×10^{-3}–4×10^{-3} eV, which is slightly less than the respective difference for the double polaron, 3×10^{-3}–6×10^{-3} eV,[18] but can be detected in experiments.

In the case of a combined polaron one can considerably reduce the resonant magnetic field $H_{\rm res}$ in comparison with that for the double polaron. Whereas $H_{\rm res}$ drops with $j=n-n_1$ in the latter case [Eq. (1)], it follows from condition (6) that in the combined polaron $\hbar\Omega_e$ drops, first with j, second, when the difference $\hbar\omega_{\rm L1}-(\varepsilon^e_m-\varepsilon^e_{m_1})$ is lowered. This is the case of the first version of condition (6). For example, for transition $m=2$, $n=2\to m_1=1$, $n=0$, the level splitting of 2.8×10^{-3} eV (Fig. 2a) takes place at $H_{\rm res}=2.80$ T, and in the case of the double polaron (transition $n=2\to n_1=0$) $H_{\rm res}=10.4$ T. On the other hand, in the case of the third version of condition (6), $H_{\rm res}$ becomes very high (Fig. 2b), since $\hbar\omega_{\rm L1}-(\varepsilon^e_m-\varepsilon^e_{m_1})$ rises.

At the lowest level of size-quantization $m=1$, a double polaron can exist at all widths of the quantum well. Since the combined polaron is related to two levels of size-quantization, it cannot exist in the range of well width $d<d'(m)$, where $d'(m)$ is the well width at which the upper level of size-quantization is ejected. The width $d'(m)$ grows with m. The lower boundary of the region where a combined polaron of the first configuration can exist also moves upwards. Figure 2 shows that at $m_1=1$, when an electron goes to the lowest size-quantized level, $d(m=2)=180$ Å, $d(m=3)=320$ Å, and $d(m=4)=460$ Å. At smaller d the polaron of the first configuration cannot exist, and the polaron of the second configuration turns up, which persists in the region $d'(m)<d<d(m)$. At d close to $d(m)$, $H_{\rm res}$ is low. For example, if $d=300$ Å (the transition $m=3\to m_1=1$), $H_{\rm res}=3.84$ T.

The calculations of the level splitting $\Delta E_{\rm res}$ given above are valid only if the crossing point of two selected levels is not superposed by a crossing point between another pair of levels. For example, the gap between points *1* and *A* or *3* and *A* (Fig. 1) should be larger than the splitting at points *1* and *A* or *3* and *A*. This condition is expressed by the inequality[18]

$$\Delta E_{\rm res}(d)\ll\varepsilon^e_2(d)-\varepsilon^e_1(d). \quad (77)$$

Figure 2a plots the difference $\varepsilon_2(d)-\varepsilon_1(d)$ as a function of the quantum well width (dash-dotted line). One can see that there is a fairly wide region of d where level splittings at crossing points can be treated independently of other crossing points.

This paper takes account of only confined phonons, although interaction with interface and barrier phonons also takes place. Interaction between electrons and holes on the one side and interface phonons on the other can be neglected when the penetration depth of interface phonons is negligible

in comparison with the quantum well width d. Since the penetration depth is of order l_H, the application domain of Eqs. (35) and (42), which take into account only confined phonons, is defined by the condition $l_H \ll d$. Since $l_H \to \infty$ as $H \to 0$, the results do not apply to the region of low magnetic fields (Fig. 2). By comparing them with previously reported calculations,[18] we find that the theory neglecting interface phonons does not apply to the region on the left of the peaks on curves of $\Delta E_{\rm res}(d)$. Since the phonon frequency in the barrier is notably different from $\omega_{\rm L1}$, interaction with phonons in the barrier is nonresonant and can be neglected.[18]

In calculating the polaron spectra, we have assumed that the conduction band is parabolic and have not taken account of excitonic effects. Since only two electron levels are involved in the formation of a combined polaron, the nonparabolicity can only vary the resonance condition (5). The Coulomb interaction between electrons and holes in a strong magnetic field in a relatively narrow quantum well can be treated as a weak perturbation.[24] The small correction to the energy of an electron–hole pair due to excitonic effects has been calculated.[24] This correction is a function of the vertical component of the electron–hole pair quasi-momentum $\mathbf{K}$, i.e., the Coulomb interaction breaks the condition that energy levels of an electron–hole pair are discrete. Even so, $\mathbf{K}=0$ when an electron–hole pair is generated by light with a wave vector normal to the quantum well plane, i.e., the Coulomb interaction does not lead to a radical modification either, but only shifts a discrete level, thus changing $H_{\rm res}$.[24]

The work was supported by the Russian Fund for Fundamental Research and the *Physics of Solid-State Nanostructures* R&D program sponsored by the Ministry of Science and Technology of Russia (Project 79-1049).

*)E-mail: pavlov@sci.lpi.ac.ru

[1] E. J. Johnson and D. M. Larsen, Phys. Rev. Lett. **16**, 655 (1966).
[2] A. Petron and B. D. McCombe, in *Landau Level Spectroscopy*, ed. by G. Landwer and E. I. Rashba, *Modern Problems in Condensed Matter Sciences* (1988), Vol. 27.
[3] R. J. Nicholas, D. J. Barnes, D. R. Seadley, C. J. Langerak, J. Singleton, P. J. van der Wel, J. A. A. J. Perenboom, J. J. Harris, and C. T. Foxon, in *Spectroscopy of Semiconductors Microstructures*, *NATO ASI, Serie B: Physics*, ed. by G. Fasol, A. Fasolino, and P. Lugli, Plenum, New York (1980), Vol. 206, p. 451.
[4] R. J. Nicholas, in *Handbook of Semiconductors*, 2nd ed., ed. by M. Balkanski, North Holland, Amsterdam (1994), Vol. 2.
[5] L. I. Korovin and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **53**, 1708 (1967) [Sov. Phys. JETP **26**, 979 (1967)]; JETP Lett. **6**, 50 (1967).
[6] L. I. Korovin, S. T. Pavlov, and B. E. Eshpulatov, Fiz. Tverd. Tela **20**, 3594 (1978) [Sov. Phys. Solid State **20**, 2077 (1978)].
[7] Das Sarma and O. Madhukar, Phys. Rev. B **22**, 2823 (1980).
[8] Das Sarma, Phys. Rev. Lett. **52**, 859 (1984).
[9] G. O. Hai, F. M. Peeters, and J. T. Devreese, in *Phonons in Semiconductor Nanostructures*, *NATO ASI Serie E: Applied Sciences*, ed. by I. P. Leburston, I. Pascual, and C. Sotomayor Torres, Kluver Academic Publ., Dordrecht, Boston, London (1993), Vol. 236, p. 509.
[10] A. O. Govorov, Solid State Commun. **92**, 977 (1994).
[11] R. J. Nicholas, S. Sasaki, N. Niura, F. M. Peeters, J. M. Shi, C. O. Hai, J. T. Devreese, M. I. Lawless, D. E. Ashenlord, and B. Lunn, Phys. Rev. B **50**, 7596 (1994).
[12] J. M. Shi, F. M. Peeters, and J. T. Devreese, Phys. Rev. B **50**, 15182 (1994).
[13] L. I. Korovin, S. T. Pavlov, and B. E. Eshpulatov, Fiz. Tverd. Tela **35**, 1562 (1993) [Phys. Solid State **35**, 788 (1993)].
[14] I. G. Lang, V. I. Belitsky, A. Cantarero, L. I. Korovin, S. T. Pavlov, and M. Cardona, Phys. Rev. B **54**, 17768 (1996).
[15] L. I. Korovin, I. G. Lang, and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **111**, 2194 (1997) [JETP **84**, 1197 (1997)].
[16] I. G. Lang, V. I. Belitsky, A. Cantarero, L. I. Korovin, S. T. Pavlov, and M. Cardona, Phys. Rev. B **56**, 16880 (1997).
[17] L. I. Korovin, I. G. Lang, and S. T. Pavlov, JETP Lett. **65**, 532 (1997).
[18] L. I. Korovin, I. G. Lang, and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **115**, 187 (1999) [JETP **88**, 105 (1999)].
[19] N. Mori and T. Ando, Phys. Rev. B **40**, 6175 (1988).
[20] L. C. Andreani, F. Tassone, and F. Bassani, Solid State Commun. **77**, 641 (1991).
[21] E. L. Ivchenko, Fiz. Tverd. Tela **33**, 2388 (1991) [Sov. Phys. Solid State **33**, 1344 (1991)].
[22] F. Tassone, F. Bassani, and L. C. Andreani, Phys. Rev. B **45**, 6023 (1992).
[23] L. C. Andreani, in *Confined Electrons and Photons*, ed. by E. Burstein and C. Weisbuch, Plenum Press, New York (1995), p. 57.
[24] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)].

Translation provided by the Russian Editorial office.

Kinetic properties of a system of spatially separated excitons and electrons in the presence of a Bose condensate of excitons

Yu. E. Lozovik*) and M. V. Nikitkov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 4 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1440–1449 (October 1999)

A system of interacting, spatially separated excitons and electrons is examined in the presence of a Bose condensate of excitons. The kinetic properties of the system that are governed by the interaction of excitations in the exciton subsystem with electrons are investigated. It is shown that a nonequilibrium distribution of excitations in the exciton subsystem gives rise to an induced electron current. Experimental observation of the kinetic phenomena described can provide new information on the exciton phase state. © *1999 American Institute of Physics.* [S1063-7761(99)02210-6]

1. INTRODUCTION

The transition of an exciton system into the superfluid phase was considered in the three-dimensional case in Refs. 1–5 and in quasi-two-dimensional structures in Refs. 6 and 7. Interesting experiments, in which the transport properties, statistics, and photoluminescence of excitons were studied, have been carried out to detect the condensation and superfluid properties of excitons.[8–14] The experimental data may attest to realization of the Bose–Einstein condensation and superfluidity of excitons in Cu_2O as the temperature is lowered,[11,13] although this is a subject of debate.[12–14]

A magnetic field has a significant influence on the phase diagram[15–17] and spectra[18] of excitons in isolated and coupled quantum wells. The study of the properties of excitons in strong magnetic fields in coupled quantum wells has shown that there are dramatic changes in the photoluminescence intensity and decay time, as well as an increase in the magnetodiffusion of the excitons, which possibly attests to their superfluidity.[19] The availability of alternative methods for detecting the Bose condensation and superfluidity of excitons would be of great interest in connection with the debate regarding the interpretation of the experimental data indicated above (see, for example, Ref. 20). One interesting method for investigating the kinetic properties of a system of spatially separated excitons and electrons, which can provide additional information regarding the phase state of the exciton subsystem and phase transitions in it, is based on drag effects. One special feature of the system under consideration is the fact that the phase state of the exciton subsystem can be studied by performing a simpler investigation of the response of the electron subsystem. In other words, the transport properties of excitons and the changes in them upon phase transitions can be investigated by measuring the current or voltage in the electron subsystem. Another property of systems of spatially separated quasiparticles interacting with one another is the possibility of controlling the motion of the quasiparticles of one subsystem by altering the parameters of state of the quasiparticles in the other subsystem (for example, controlling the motion of electrons using a flow of excitons).

The kinetic properties of a two-layer system of electrons and excitons at temperatures above the temperature T_{KT} of the Kosterlitz–Thouless transition, at which there is no local condensate or superfluidity in the exciton system, were considered in Ref. 21. In the present paper we examine the kinetic properties of this system under conditions allowing the existence of a Bose condensate of excitons. The treatment is performed for two- and three-dimensional systems of spatially separated electrons and excitons.

We assume that the exciton system is dilute, i.e., $N_1 a_0^d \ll 1$, where $a_0=(d-1)\epsilon\hbar^2/2me^2$ is the exciton radius and d is the dimensionality of the system ($d=2,3$). In this case the excitons can be considered Bose particles in an approximation. In fact, the characteristic value for the exciton radius is $a_0 \sim 10-50$ Å. Therefore, the condition $N_1 a_0^d \ll 1$ holds at values of the exciton density up to $N_1 \simeq 10^{12}\,\mathrm{cm}^{-2}$ for $d=2$ and $N_1 \sim 10^{18}\,\mathrm{cm}^{-3}$ for $d=3$.[1)] We shall consider a nonequilibrium system of excitons, in which there is a directed flow of excitons. One method for obtaining a system of excitons with nonequilibrium characteristics is to create excitons in one region of the sample using laser radiation. This gives rise to chemical potential and temperature gradients in the sample, which cause excitons to flow. The interaction between spatially separated excitons and electrons in the presence of an exciton flux causes the appearance of an induced electron current.[21] The analogous drag effects in electron and electron–hole systems were considered in Refs. 25–46. We shall calculate the current in the electron subsystem in the case where the appearance of this current is associated with the scattering of quasiparticle excitations in the exciton subsystem on electrons. The kinetic properties of such a system will be investigated using the kinetic equations for the electrons and quasiparticles in an exciton Bose gas.

2. INTERACTION OF QUASIPARTICLE EXCITATIONS WITH ELECTRONS

Let us first consider the interaction of an isolated exciton and an isolated electron. For simplicity we assume that the

exciton and electron are separated by a spatial barrier of thickness D, such that $D>a_0$. Then the polarization energy for the interaction of the exciton and electron can be written in an approximation in the form

$$W(r,D)=-\frac{\gamma}{(r^2+D^2)^2}, \tag{1}$$

where $\gamma=(21/32)e^2a_0^3/\epsilon_0$ for $d=2$ and $\gamma=(9/4)e^2a_0^3/\epsilon_0$ for $d=3$, and ϵ_0 is the static dielectric constant of the medium. In the momentum representation Eq. (1) takes the form (here and below we shall use the system of units in which $\hbar=k_B=1$)

$$W(q,D)=-\gamma(\pi q/D)K_1(qD), \quad d=2,$$
$$W(q,D)=-2\pi\gamma qK_0(qD), \quad d=3. \tag{2}$$

Here $K_\nu(z)$ is the modified Bessel function of the second kind.

In a many-particle system of excitons and electrons the bare interaction $W(q,D)$ should be replaced by the effective interaction $W_{\rm eff}(q,D)$. For a two-component system of interacting particles the effective interaction has the form

$$W_{\rm eff}(q,D)=\frac{W(q,D)}{1+\Pi_1V_1+\Pi_2V_2+\Pi_1\Pi_2(V_1V_2-W^2)}. \tag{3}$$

Here V_i is the energy of the interaction between particles of species i, and Π_i is the polarization operator for particles of species i $(i=1,2)$. In our case $i=1$ corresponds to excitons, and $i=2$ corresponds to electrons. When the condition $N_1a_0^d\ll1$ is satisfied, the contribution of the electrons to the effective interaction will dominate, providing a basis to take into account only this contribution in (3):

$$W_{\rm eff}(q,D)=\frac{W(q,D)}{1+\Pi_2V_2}. \tag{4}$$

At a low exciton density, $N_1a_0^d\ll1$, the exciton-electron Hamiltonian can be written in the form

$$\hat H_{12}=\frac{1}{L^d}\sum_{\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_1',\mathbf{p}_2'}W(|\mathbf{p}_1-\mathbf{p}_1'|,D)a^+_{\mathbf{p}_1'}c^+_{\mathbf{p}_2'}c_{\mathbf{p}_2}a_{\mathbf{p}_1}, \tag{5}$$

where $a^+_{\mathbf{p}}$ and $a_{\mathbf{p}}$ are the exciton creation and annihilation operators, $c^+_{\mathbf{p}}$ and $c_{\mathbf{p}}$ are the electron creation and annihilation operators, and L is equal to the dimensions of the system. Starting out from (5), we can obtain the Hamiltonian for the interaction of quasiparticle excitations in a system of excitons with electrons. For this purpose, we use the Bogolyubov transformations[47–49]

$$a_{\mathbf{p}}=u_pb_{\mathbf{p}}+v_pb^+_{-\mathbf{p}}, \quad a^+_{\mathbf{p}}=u_pb^+_{\mathbf{p}}+v_pb_{-\mathbf{p}},$$
$$u_p=(1-L_p^2)^{-1/2}, \quad v_p=L_pu_p,$$
$$L_p=(\varepsilon_1(p)-\xi(p))/\mu_1, \quad \varepsilon_1(p)=(\xi^2(p)-\mu_1^2)^{1/2},$$
$$\xi(p)=p^2/2m_1+\mu_1, \tag{6}$$

where $b^+_{\mathbf{p}}$ and $b_{\mathbf{p}}$ are the creation and annihilation operators of the quasiparticle excitations, m_1 is the exciton effective mass, and μ_1 is the exciton chemical potential. Substituting $a^+_{\mathbf{p}}$ and $a_{\mathbf{p}}$ from (6) into (5), we obtain

$$\hat H_{12}=\frac{1}{L^d}\sum_{\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_1',\mathbf{p}_2'}W(|\mathbf{p}_1-\mathbf{p}_1'|,D)\{u_{p_1}u_{p_1'}+v_{p_1}v_{p_1'}\}b^+_{\mathbf{p}_1'}c^+_{\mathbf{p}_2'}c_{\mathbf{p}_2}b_{\mathbf{p}_1}. \tag{7}$$

In (7) we have taken into account only the terms which leave the number of quasiparticles unchanged (the terms which do not simultaneously satisfy the momentum and energy conservation laws have been omitted).

Let us now consider the transport relaxation time of the excitations $\tau_1(p)$ in the case of scattering on impurities. For the Hamiltonian of elastic interactions of excitons with impurities we have

$$\hat H_1=\frac{1}{L^d}\sum_{\mathbf{p},\mathbf{p}'}V(\mathbf{p},\mathbf{p}')a^+_{\mathbf{p}}a_{\mathbf{p}'}, \tag{8}$$

where $V(\mathbf{p},\mathbf{p}')$ is the matrix element for the interaction of an exciton with impurities. Replacing the exciton operators in (8) by the operators for quasiparticle excitations defined by formulas (6) and retaining only the terms which satisfy the requirement of elastic collisions, we obtain

$$\hat H_1=\frac{1}{L^d}\sum_{\mathbf{p},\mathbf{p}'}V(\mathbf{p},\mathbf{p}')\{u_p^2+v_p^2\}b^+_{\mathbf{p}}b_{\mathbf{p}'}. \tag{9}$$

Using the Hamiltonian (9), for the reciprocal of the transport relaxation time we write

$$\frac{1}{\tau_1(p)}=2\pi\int|V(\mathbf{p},\mathbf{p}')\{u_p^2+v_p^2\}|^2\delta(\varepsilon_1(p)-\varepsilon_1(p'))(1-\cos(\widehat{\mathbf{p},\mathbf{p}'}))\frac{d\mathbf{p}'}{(2\pi)^d}. \tag{10}$$

We transform the expression (10) using the equalities $(u_p^2+v_p^2)^2=(\xi(p)/\varepsilon_1(p))^2$ and $\delta(\varepsilon_1(p)-\varepsilon_1(p'))=(\varepsilon_1(p)/\xi(p))\delta(\varepsilon_1^0(p)-\varepsilon_1^0(p'))$, where $\varepsilon_1^0(p)=p^2/2m_1$. As a result, we find that the exciton relaxation time is related to the exciton relaxation time in the normal phase $\tau_n(p)$ by the expression

$$\tau_1(p)=\frac{\varepsilon_1(p)}{\xi(p)}\tau_n(p), \tag{11}$$

where

$$\frac{1}{\tau_n(p)}=2\pi\int|V(\mathbf{p},\mathbf{p}')|^2\delta(\varepsilon_1^0(p)-\varepsilon_1^0(p'))(1-\cos(\widehat{\mathbf{p},\mathbf{p}'}))\frac{d\mathbf{p}'}{(2\pi)^d}.$$

In the case of excitations with a small quasimomentum, where $\mu_1\gg\varepsilon_1^0$, the dispersion law will have an acoustic form: $\varepsilon_1(p)=cp$, where $c=(\mu_1/m_1)^{1/2}$. In this case, from (11) we have $\tau_1(p)=(p/m_1c)\tau_n(p)$.

3. KINETIC EQUATIONS

To describe the nonequilibrium processes in the structures under consideration, we utilize the system of kinetic equations for the quasiparticle-excitation distribution function n and the electron distribution function f. We consider the stationary regime, in which the functions $n(\mathbf{r},\mathbf{p})$ and $f(\mathbf{r},\mathbf{p})$ do not depend on time. The kinetic equations have the form[50,51]

$$\frac{\partial n}{\partial \mathbf{r}}\cdot\frac{\partial \tilde{\varepsilon}_1}{\partial \mathbf{p}}-\frac{\partial n}{\partial \mathbf{p}}\cdot\frac{\partial \tilde{\varepsilon}_1}{\partial \mathbf{r}}=I_1(n)+I_{12}(n,f), \tag{12}$$

$$\frac{\partial f}{\partial \mathbf{r}}\cdot\mathbf{v}+\frac{\partial f}{\partial \mathbf{p}}\cdot\dot{\mathbf{p}}=I_2(f)+I_{21}(f,n). \tag{13}$$

The terms on the right-hand sides of these equations have the following meanings: I_{12} and I_{21} are the collision integrals of the quasiparticles with electrons; I_1 I_2 are the collision integrals of quasiparticles and electrons with impurities (allowance for only the impurity scattering is justified at low temperatures, which we are considering here). The kinetic equation (12) contains the quasiparticle excitation energy $\tilde{\varepsilon}_1(\mathbf{p})=\varepsilon_1(p)+\mathbf{p}\cdot\mathbf{v}_s$, where $\mathbf{v}_s$ is the rate of superfluid motion of excitons. However, the influence of the superfluid motion of the excitons on the phenomena under consideration will be significant for $v_s\sim c$, which is unlikely under real conditions; therefore, the $\mathbf{p}\cdot\mathbf{v}_s$ term in the quasiparticle excitation energy will not be taken into account below.

Let us consider an exciton subsystem in which there is a temperature gradient. In this case directed motion of the quasiparticle excitations occurs and, in turn, gives rise to an induced current in the electron subsystem. Let $T_1(\mathbf{r})$ be the temperature of the exciton subsystem and T_2 be the temperature of the electron subsystem. We represent the distribution functions n and f in the following form:

$$n=n_0+n_0(1+n_0)g_1, \quad f=f_0+f_0(1-f_0)g_2,$$

where $n_0=(\exp(\varepsilon_1(p)/T_1)-1)^{-1}$, $f_0=(\exp((\varepsilon_2-\mu_2)/T_2)+1)^{-1}$, $\varepsilon_2=p^2/2m_2$, μ_2 is the electron chemical potential, and m_2 is the electron effective mass. We use the τ approximation for $I_1(n)$ and $I_2(f)$:

$$I_1(n)=(n_0-n)/\tau_1, \quad I_2(f)=(f_0-f)/\tau_2,$$

where $\tau_1(p)$ and $\tau_2(p)$ are the relaxation times of the quasiparticle excitations and electrons. Substituting n and f in the form just indicated into (12) and (13), we obtain linearized equations:

$$\frac{\varepsilon_1}{T_1^2}\frac{\partial \varepsilon_1}{\partial \mathbf{p}}\cdot\nabla T_1=-\frac{g_1}{\tau_1}, \tag{14}$$

$$I_{21}(g_2,g_1)=\frac{f_0(1-f_0)g_2}{\tau_2}. \tag{15}$$

Here the linearized collision integral I_{21} has the form

$$I_{21}(g_2,g_1)=\sum_{\sigma_2'}\int w(\mathbf{p}_1\mathbf{p}_2;\mathbf{p}_1'\mathbf{p}_2')n_0(1+n_0')f_0(1-f_0')\times(g_1'+g_2'-g_1-g_2)\delta(\varepsilon_1+\varepsilon_2-\varepsilon_1'-\varepsilon_2')\times\frac{d\mathbf{p}_1'}{(2\pi)^d}\frac{d\mathbf{p}_2'}{(2\pi)^d}. \tag{16}$$

In deriving Eq. (14) we neglected the term I_{12}, which is only a perturbation with respect to I_1.

The equality (14) yields an expression for g_1:

$$g_1(\mathbf{p})=-\frac{\tau_1(p)\varepsilon_1(p)}{T_1^2}\frac{\partial\varepsilon_1(p)}{\partial\mathbf{p}}\cdot\nabla T_1=-\frac{c^2\tau_1(p)}{T_1^2}\mathbf{p}\cdot\nabla T_1. \tag{17}$$

In a first approximation $I_{21}(g_2,g_1)$ can be replaced by $I_{21}(0,g_1)$. Then Eq. (15) is greatly simplified, and we at once obtain g_2:

$$f_0(1-f_0)g_2(\mathbf{p}_2)=\tau_2(p_2)\sum_{\sigma_2'}\int w(\mathbf{p}_1\mathbf{p}_2;\mathbf{p}_1'\mathbf{p}_2')\times n_0(1+n_0')f_0(1-f_0')(g_1'-g_1)\delta\times(\varepsilon_1+\varepsilon_2-\varepsilon_1'-\varepsilon_2')\frac{d\mathbf{p}_1'}{(2\pi)^d}\frac{d\mathbf{p}_2'}{(2\pi)^d}, \tag{18}$$

where $w(\mathbf{p}_1\mathbf{p}_2;\mathbf{p}_1'\mathbf{p}_2')$ is the probability of a collision between a quasiparticle and an electron. With consideration of the results in Sec. 1, we can write the following expression for the collision probability w in the Born approximation

$$w(\mathbf{p}_1\mathbf{p}_2;\mathbf{p}_1'\mathbf{p}_2')=2\pi|W_{\text{eff}}(q,D)\{u_{p_1}u_{p_1'}+v_{p_1}v_{p_1'}\}|^2, \tag{19}$$

where $\mathbf{q}=\mathbf{p}_1'-\mathbf{p}_1=\mathbf{p}_2-\mathbf{p}_2'$.

Using Eq. (17), we have the following expression for the electron current $\mathbf{j}_2$:

$$\mathbf{j}_2=-\frac{e}{m_2}\sum_{\sigma_2}\int\mathbf{p}_2f(\mathbf{p}_2)\frac{d\mathbf{p}_2}{(2\pi)^d}=-\frac{e}{m_2}\sum_{\sigma_2}\int\mathbf{p}_2f_0(1-f_0)g_2(\mathbf{p}_2)\frac{d\mathbf{p}_2}{(2\pi)^d}=-\beta_{21}\nabla T_1, \tag{20}$$

where

$$\beta_{21}=\frac{4\pi}{d}\frac{ec^2}{m_2T_1^2}\int W_{\text{eff}}^2(q,D)\{u(p_1)u(|\mathbf{p}_1+\mathbf{q}|)+v(p_1)v(|\mathbf{p}_1+\mathbf{q}|)\}^2n_0(p_1)(1+n_0(|\mathbf{p}_1+\mathbf{q}|))f_0(p_2)(1-f_0(|\mathbf{p}_2-\mathbf{q}|))\tau_2(p_2)\mathbf{p}_2\times[\tau_1(p_1)\mathbf{p}_1-\tau_1(|\mathbf{p}_1+\mathbf{q}|)(\mathbf{p}_1+\mathbf{q})]\delta(\varepsilon_1(p_1)+\varepsilon_2(p_2)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)-\varepsilon_2(|\mathbf{p}_2-\mathbf{q}|))\frac{d\mathbf{p}_1}{(2\pi)^d}\frac{d\mathbf{p}_2}{(2\pi)^d}\frac{d\mathbf{q}}{(2\pi)^d}. \tag{21}$$

We transform the expression (21) using the equalities

$$\delta(\varepsilon_1(p_1)+\varepsilon_2(p_2)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)-\varepsilon_2(|\mathbf{p}_2-\mathbf{q}|))$$
$$=\int_{-\infty}^{\infty} d\xi\,\delta(\varepsilon_1(p_1)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)+\xi)\,\delta(\varepsilon_2(p_2)$$
$$-\varepsilon_2(|\mathbf{p}_2-\mathbf{q}|)-\xi), \tag{22}$$

$$n_0(p_1)(1+n_0(|\mathbf{p}_1+\mathbf{q}|))=[n_0(|\mathbf{p}_1+\mathbf{q}|)-n_0(p_1)]$$
$$\times[\exp\{(\varepsilon_1(p_1)-\varepsilon_1(|\mathbf{p}_1$$
$$+\mathbf{q}|))/T_1\}-1]^{-1}, \tag{23}$$

$$f_0(p_2)(1-f_0(|\mathbf{p}_2-\mathbf{q}|))=[f_0(|\mathbf{p}_2-\mathbf{q}|)-f_0(p_2)]$$
$$\times[\exp\{(\varepsilon_2(p_2)-\varepsilon_2(|\mathbf{p}_2$$
$$-\mathbf{q}|))/T_2\}-1]^{-1}. \tag{24}$$

As a result, we obtain

$$\beta_{21}=\frac{\pi}{d}\frac{ec^2}{m_2T_1^2}\int\frac{d\mathbf{q}}{(2\pi)^d}W_{\mathrm{eff}}^2(q,D)$$
$$\times\int_{-\infty}^{\infty}d\xi\frac{\exp(\xi/2T_1-\xi/2T_2)}{\sinh(\xi/2T_1)\sinh(\xi/2T_2)}$$
$$\times\int\{u(p_1)u(|\mathbf{p}_1+\mathbf{q}|)+v(p_1)v(|\mathbf{p}_1+\mathbf{q}|)\}^2[n_0(|\mathbf{p}_1$$
$$+\mathbf{q}|)-n_0(p_1)](\tau_1(|\mathbf{p}_2+\mathbf{q}|)(\mathbf{p}_1+\mathbf{q})$$
$$-\tau_1(p_1)\mathbf{p}_1)\,\delta(\varepsilon_1(p_1)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)+\xi)\frac{d\mathbf{p}_1}{(2\pi)^d}$$
$$\times\int[f_0(|\mathbf{p}_2+\mathbf{q}|)-f_0(p_2)]\tau_2(p_2)\mathbf{p}_2\,\delta(\varepsilon_2(p_2)$$
$$-\varepsilon_2(|\mathbf{p}_2+\mathbf{q}|)-\xi)\frac{d\mathbf{p}_2}{(2\pi)^d}. \tag{25}$$

In the case of equality between the temperatures of the exciton and electron subsystems ($T_1=T_2=T$), the expression (25) can be written in a symmetric form. For this purpose we note that

$$\Phi(\mathbf{q},\xi)=\int\frac{d\mathbf{p}_1}{(2\pi)^d}\{u(p_1)u(|\mathbf{p}_1+\mathbf{q}|)+v(p_1)v(|\mathbf{p}_1$$
$$+\mathbf{q}|)\}^2[n_0(|\mathbf{p}_1+\mathbf{q}|)-n_0(p_1)](\tau_1(|\mathbf{p}_1+\mathbf{q}|)$$
$$\times(\mathbf{p}_1+\mathbf{q})-\tau_1(p_1)\mathbf{p}_1)\,\delta(\varepsilon_1(p_1)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)$$
$$+\xi)=-\Phi(\mathbf{q},-\xi), \tag{26}$$

and

$$\int[f_0(|\mathbf{p}_2+\mathbf{q}|)-f_0(p_2)]\tau_2(p_2)\mathbf{p}_2\,\delta(\varepsilon_2(p_2)-\varepsilon_2(|\mathbf{p}_2$$
$$+\mathbf{q}|)-\xi)\frac{d\mathbf{p}_2}{(2\pi)^d}=\frac{1}{2}\int[f_0(|\mathbf{p}_2+\mathbf{q}|)-f_0(p_2)]$$
$$\times\{\tau_2(p_2)\mathbf{p}_2\,\delta(\varepsilon_2(p_2)-\varepsilon_2(|\mathbf{p}_2+\mathbf{q}|)-\xi)+\tau_2(|\mathbf{p}_2$$
$$+\mathbf{q}|)(\mathbf{p}_2+\mathbf{q})\,\delta(\varepsilon_2(p_2)-\varepsilon_2(|\mathbf{p}_2+\mathbf{q}|)+\xi)\}\frac{d\mathbf{p}_2}{(2\pi)^d}. \tag{27}$$

Equations (26) and (27) are easily obtained via the following changes in the integration variables: $\mathbf{p}_1\to\mathbf{p}_1'-\mathbf{q}$ and $\mathbf{p}_1'\to-\mathbf{p}_1$, $\mathbf{p}_2\to\mathbf{p}_2'-\mathbf{q}$ and $\mathbf{p}_2'\to-\mathbf{p}_2$.

Now, performing the replacement $\xi\to-\xi$ in the second term in the curly brackets in Eq. (27) and taking into account that $\Phi(\mathbf{q},\xi)$ is an odd function with respect to ξ, we obtain the following expression for β_{21}:

$$\beta_{21}=\frac{\pi}{d}\frac{ec^2}{m_2T_2}\int\frac{d\mathbf{q}}{(2\pi)^d}W_{\mathrm{eff}}^2(q,D)$$
$$\times\int_0^{\infty}\frac{\Phi(\mathbf{q},\xi)\Psi(\mathbf{q},\xi)}{\sinh^2(\xi/2T)}d\xi, \tag{28}$$

where

$$\Psi(\mathbf{q},\xi)=\int[f_0(|\mathbf{p}_2+\mathbf{q}|)-f_0(p_2)](\tau_2(|\mathbf{p}_2+\mathbf{q}|)(\mathbf{p}_2+\mathbf{q})$$
$$-\tau_2(p_2)\mathbf{p}_2)\,\delta(\varepsilon_2(p_2)-\varepsilon_2(|\mathbf{p}_2+\mathbf{q}|)+\xi)\frac{d\mathbf{p}_2}{(2\pi)^d}. \tag{29}$$

Let us now examine the case where equilibrium between the exciton subsystem and the lattice is established after a time significantly shorter than the lifetime of the excitons, so that the lattice temperature and the temperature of the exciton subsystem are equal. In this case we shall consider the exciton subsystem at times greater than the time for the establishment of equilibrium with the lattice but smaller than the lifetime of the excitons. We shall obtain an expression for the induced electron current caused by the presence of a spatially inhomogeneous distribution of excitations.

We represent the excitation distribution function in the following form:

$$n=n_1+n_1(1+n_1)\phi_1, \tag{30}$$

where the first term of the expansion, i.e., the function n_1, already corresponds to a nonequilibrium state of the excitations:

$$n_1(\mu_{qp})=(\exp([\varepsilon_1(p)-\mu_{qp}(\mathbf{r})]/T_1)-1)^{-1}. \tag{31}$$

The parameter $\mu_{qp}=\mu_{qp}(\mathbf{r})$ is determined by the external conditions. The excitation density N_{qp} is a function of μ_{qp} and is defined by the expression

$$N_{qp}=\int n\frac{d\mathbf{p}}{(2\pi)^d}=\int n_1(\mu_{qp})\frac{d\mathbf{p}}{(2\pi)^d}. \tag{32}$$

It is shown below that ϕ_1 is an odd function of the momentum $\mathbf{p}$; hence in (32) the integral of $n_1(1+n_1)\phi_1$ with respect to the momentum vanishes.

We assume that there is a small gradient of μ_{qp} in the excitation subsystem. We find the induced electron current as the response to $\nabla\mu_{qp}$:

$$\mathbf{j}_2=-\alpha_{21}\nabla\mu_{qp}. \tag{33}$$

If we express the electron current in terms of the excitation concentration gradient in the form $\mathbf{j}_2=-\tilde{\alpha}_{21}\nabla N_{qp}$, we obtain the relation between the coefficients α_{21} and $\tilde{\alpha}_{21}$: $\alpha_{21}=(\partial N_{qp}/\partial\mu_{qp})\tilde{\alpha}_{21}$.

We linearize Eqs. (12) and (13) for the case under consideration in analogy to what was done above, except that now we define n by Eq. (30), we take f in the form $f=f_0+f_0(1-f_0)\phi_2$, and set $I_1=(n_1-n)/\tau_1$.

As a result of the linearization, we obtain the equations

$$\frac{1}{T_1}\frac{\partial\varepsilon_1(p)}{\partial\mathbf{p}}\cdot\nabla\mu_{qp}=-\frac{\phi_1}{\tau_1}, \tag{34}$$

$$I_{21}(\phi_2,\phi_1)=\frac{f_0(1-f_0)\phi_2}{\tau_2}. \tag{35}$$

From (34) we obtain an expression for ϕ_1:

$$\phi_1=-\frac{\tau_1(p)}{T_1}\frac{\partial\varepsilon_1(p)}{\partial\mathbf{p}}\cdot\nabla\mu_{qp}=-\frac{\tau_n(p)}{m_1T_1}\mathbf{p}\cdot\nabla\mu_{qp}. \tag{36}$$

After some transformations, which are analogous to the ones previously performed during the derivation of the expression for β_{21}, in the present case for α_{21} we obtain

$$\begin{aligned}\alpha_{21}=&\frac{\pi}{d}\frac{e}{m_1m_2T_1}\int\frac{d\mathbf{q}}{(2\pi)^d}W_{\text{eff}}^2(q,D)\\&\times\int_{-\infty}^{\infty}d\xi\frac{\exp(\xi/2T_1-\xi/2T_2)}{\sinh(\xi/2T_1)\sinh(\xi/2T_2)}\\&\times\int\{u(p_1)u(|\mathbf{p}_1+\mathbf{q}|)+v(p_1)v(|\mathbf{p}_1+\mathbf{q}|)\}^2[n_1(|\mathbf{p}_1\\&+\mathbf{q}|)-n_1(p_1)](\tau_n(|\mathbf{p}_1+\mathbf{q}|)(\mathbf{p}_1+\mathbf{q})\\&-\tau_n(p_1)\mathbf{p}_1)\delta(\varepsilon_1(p_1)-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)+\xi)\frac{d\mathbf{p}_1}{(2\pi)^d}\\&\times\int[f_0(|\mathbf{p}_2+\mathbf{q}|)-f_0(p_2)]\tau_2(p_2)\mathbf{p}_2\delta(\varepsilon_2(p_2)\\&-\varepsilon_2(|\mathbf{p}_2+\mathbf{q}|)-\xi)\frac{d\mathbf{p}_2}{(2\pi)^d}.\end{aligned} \tag{37}$$

In the case of $T_1=T_2=T$, for α_{21} we have

$$\begin{aligned}\alpha_{21}=&\frac{\pi}{d}\frac{e}{m_1m_2T}\int\frac{d\mathbf{q}}{(2\pi)^d}W_{\text{eff}}^2(q,D)\\&\times\int_0^{\infty}\frac{\tilde{\Phi}(\mathbf{q},\xi)\Psi(\mathbf{q},\xi)}{\sinh^2(\xi/2T)}d\xi,\end{aligned} \tag{38}$$

where

$$\begin{aligned}\tilde{\Phi}(\mathbf{q},\xi)=&\int\{u(p_1)u(|\mathbf{p}_1+\mathbf{q}|)+v(p_1)v(|\mathbf{p}_1\\&+\mathbf{q}|)\}^2[n_1(|\mathbf{p}_1+\mathbf{q}|)-n_1(p_1)](\tau_n(|\mathbf{p}_1+\mathbf{q}|)\\&\times(\mathbf{p}_1+\mathbf{q})-\tau_n(p_1)\mathbf{p}_1)\delta(\varepsilon_1(p_1)\\&-\varepsilon_1(|\mathbf{p}_1+\mathbf{q}|)+\xi)\frac{d\mathbf{p}_1}{(2\pi)^d}.\end{aligned} \tag{39}$$

Thus, we have obtained expressions for the linear response of the electron subsystem to a disturbance within the quasiparticle subsystem in which temperature and concentration gradients are created.

4. DISCUSSION OF RESULTS

We have considered a system of interacting excitons and electrons, in which there is a Bose condensate of excitons. Under conditions where the exciton system is in a nonequilibrium state, the mutual scattering of the quasiparticle excitations and electrons gives rise to an induced electron current. The value of the induced electron current should increase with the number of excitations. At low temperatures the principal excitations in the exciton system are excitons with an acoustic spectrum $\varepsilon_1(p)=cp$. The concentration of excitations of this type increases with temperature according to $N_{\text{qp}}\sim T_1^d$. Thus, an increase in the induced electron current with increasing exciton temperature should be expected.

At sufficiently low values of the temperatures T_1 and T_2 and small values of D, a bound state of an exciton and an electron can form.[52,53] The number of such bound pairs should rise as T_1 and T_2 are lowered. Thus, such a process can cause the induced electron current to increase with decreasing temperature.

An experimental investigation of the induced electron current caused by the nonequilibrium state of an exciton system can serve as an independent method for studying the phase state of the exciton system and the phase transitions in it.

This research was carried out with financial support from the Russian Fund for Fundamental Research, INTAS, and the ''Physics of Solid-State Nanostructures'' Program.

*)E-mail: lozovik@isan.troitsk.ru

1)The opposite limit, i.e., a dense electron–hole system with pairing, was considered in Refs. 22–24.

[1] S. A. Moskalenko, Fiz. Tverd. Tela (Leningrad) **4**, 276 (1962) [Sov. Phys. Solid State **4**, 199 (1962)].
[2] J. M. Blatt, K. W. Böer, and W. Brandt, Phys. Rev. **126**, 1691 (1962).
[3] R. S. Casella, J. Appl. Phys. **34**, 1703 (1963).
[4] V. A. Gergel', R. F. Kazarinov, and R. A. Suris, Zh. Éksp. Teor. Fiz. **53**, 544 (1967) [Sov. Phys. JETP **26**, 354 (1968)]; Zh. Éksp. Teor. Fiz. **54**, 298 (1968) [Sov. Phys. JETP **27**, 159 (1968)].
[5] L. V. Keldysh, A. N. Kozlov, JETP Lett. **5**, 190 (1967); Zh. Éksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys. JETP **27**, 521 (1968)].
[6] Yu. E. Lozovik and V. I. Yudson, JETP Lett. **22**, 274 (1975); Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)].
[7] Yu. E. Lozovik and O. L. Berman, JETP Lett. **64**, 573 (1996); Zh. Éksp. Teor. Fiz. **111**, 1879 (1997) [JETP **84**, 1027 (1997)]; Yu. E. Lozovik and O. L. Berman, Phys. Scr. **55**, 491 (1997).
[8] D. Hulin, A. Mysyrowicz, and S. Benoit à la Guillaume, Phys. Rev. Lett. **45**, 1970 (1980).
[9] V. V. Timofeev, V. D. Kulakovskii, and I. V. Kukushkin, Physica B (Amsterdam) **117–118**, 327 (1983).
[10] N. Peyghambarian, L. L. Chase, and A. Mysyrowicz, Phys. Rev. B **27**, 2325 (1983).
[11] D. W. Snoke, J. P. Wolfe, and A. Mysyrowicz, Phys. Rev. Lett. **64**, 2543 (1990).
[12] E. Fortin, S. Farad, and A. Mysyrowicz, Phys. Rev. Lett. **70**, 3951 (1993).
[13] J. L. Lin and J. P. Wolfe, Phys. Rev. Lett. **71**, 1222 (1993).

[14] E. Benson, E. Fortin, and A. Mysyrowicz, Solid State Commun. **101**, 313 (1997).
[15] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)]; Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)].
[16] A. B. Dzyubenko and Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **25**, 1519 (1983) [Sov. Phys. Solid State **25**, 874 (1983)]; Fiz. Tverd. Tela (Leningrad) **26**, 1540 (1984) [Sov. Phys. Solid State **26**, 938 (1984)]; J. Phys. A **24**, 415 (1991).
[17] Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, JETP Lett. **66**, 355 (1997); Phys. Rev. B **59**, 5627 (1999).
[18] Yu. E. Lozovik and A. M. Ruvinsky, Phys. Lett. A **227**, 271 (1997); Yu. E. Lozovik and A. M. Ruvinskiĭ, Zh. Éksp. Teor. Fiz. **112**, 1791 (1997) [JETP **85**, 797 (1997)].
[19] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Wiemann, Phys. Rev. Lett. **73**, 304 (1994).
[20] Yu. E. Lozovik and A. V. Poushnov, Solid State Commun. **105**, 527 (1998); Phys. Rev. B **58**, 6608 (1998); Zh. Éksp. Teor. Fiz. **115**, 1786 (1999) [JETP **88**, 747 (1999)].
[21] Yu. E. Lozovik and M. V. Nikitkov, Zh. Éksp. Teor. Fiz. **111**, 1107 (1997) [JETP **84**, 612 (1997)].
[22] L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela (Leningrad) **6**, 2791 (1964) [Sov. Phys. Solid State **6**, 2219 (1965)].
[23] A. N. Kozlov and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **48**, 1184 (1965) [Sov. Phys. JETP **21**, 790 (1965)].
[24] V. I. Halperin and T. M. Rice, Solid State Phys. **21**, 115 (1968).
[25] M. B. Pogrebinskiĭ, Fiz. Tekh. Poluprovodn. **11**, 637 (1977) [Sov. Phys. Semicond. **11**, 372 (1977)].
[26] P. J. Price, Physica B (Amsterdam) **117–118**, 750 (1983).
[27] P. M. Solomon, P. J. Price, D. J. Frank, and D. C. La Tulipe, Phys. Rev. Lett. **63**, 2508 (1989).
[28] U. Sivan, P. M. Solomon, and H. Shtrikman, Phys. Rev. Lett. **68**, 1196 (1992).
[29] T. J. Gramila, J. P. Eisenstein, A. H. MacDonald, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **66**, 1216 (1991); Phys. Rev. B **47**, 12957 (1993).
[30] B. Laikhtman and P. M. Solomon, Phys. Rev. B **41**, 9921 (1990).
[31] H. C. Tso, P. Vasilopoulos, and F. M. Peeters, Phys. Rev. Lett. **68**, 2516 (1992); Phys. Rev. Lett. **70**, 2146 (1993).
[32] Yu. M. Sirenko and P. Vasilopoulos, Phys. Rev. B **46**, 1611 (1992).
[33] A.-P. Jauho and H. Smith, Phys. Rev. B **47**, 4420 (1993).
[34] L. Zheng and A. H. MacDonald, Phys. Rev. B **48**, 8203 (1993).
[35] E. Shimshoni and S. L. Sondhi, Phys. Rev. B **49**, 11484 (1994).
[36] K. Flensberg and Ben Yu-Kvang Hu, Phys. Rev. Lett. **73**, 3572 (1994); Phys. Rev. B **52**, 14796 (1995).
[37] L. Swierkowski, J. Szymanski, and Z. W. Gortel, Phys. Rev. Lett. **74**, 3245 (1995).
[38] E. Shimshoni, Phys. Rev. B **51**, 9415 (1995).
[39] A. Kamenev and Y. Oreg, Phys. Rev. B **52**, 7516 (1995).
[40] K. Flensberg, B. Yu.-K. Hu, A.-P. Jauho, and J. M. Kinaret, Phys. Rev. B **52**, 14761 (1995).
[41] A. G. Rojo and G. D. Mahan, Phys. Rev. Lett. **68**, 2074 (1992).
[42] J. M. Duan and S. Yip, Phys. Rev. Lett. **70**, 3647 (1994).
[43] N. Giordano and J. D. Monnier, Phys. Rev. B **50**, 9363 (1994).
[44] X. Huang, G. Bazàn, and G. H. Bernstein, Phys. Rev. Lett. **74**, 4051 (1995).
[45] G. Vignale and A. H. MacDonald, Phys. Rev. Lett. **76**, 2786 (1996).
[46] H. Rubel, A. Fischer, W. Dietsche, K. von Klitzing, and K. Eberl, Phys. Rev. Lett. **78**, 1763 (1997).
[47] E. M. Lifshitz and L. P. Pitaevskii, *Statistical Physics*, *Vol. 2*, 2nd ed., Pergamon Press, Oxford–New York (1980) [Russ. original, Part 2, Nauka, Moscow (1978)].
[48] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood Cliffs, N.J. (1963) [Russ. original, Fizmatgiz, Moscow (1962)].
[49] E. Hanamura and H. Haug, Phys. Rep. C **33**, 209 (1977).
[50] E. M. Lifshitz, L. P. Pitaevskiĭ, and L. D. Landau, *Physical Kinetics*, Pergamon Press, Oxford (1981) [Russ. original, Nauka, Moscow (1979)].
[51] I. M. Khalatnikov, *Theory of Superfluidity* [in Russian], Nauka, Moscow (1971).
[52] V. S. Babichenko and M. N. Kiselev, J. Mosc. Phys. Soc. **2**, 311 (1992).
[53] V. S. Babichenko and M. N. Kiselev, JETP Lett. **57**, 179 (1993).

Translated by P. Shelnitz

Quantum glass transition in a periodic long-range Josephson array

D. M. Kagan*) and M. V. Feigel'man†)

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia

L. B. Ioffe

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia; Department of Physics, Rutgers University, Piscataway, NJ 08855, USA

(Submitted 18 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1450–1461 (October 1999)

We show that the ground state of a periodic long-range Josephson array frustrated by a magnetic field is a glass for sufficiently large Josephson energies despite the absence of quenched disorder. Like superconductors, this glass state has non-zero phase stiffness and Meissner response; for lower Josephson energies the glass ''melts'' and the ground state loses its phase stiffness and becomes insulating. We find the critical scaling behavior near this quantum phase transition: the excitation gap vanishes as $(J-J_c)^2$, and the frequency-dependent magnetic susceptibility behaves as $\chi(\omega)\propto\sqrt{\omega}\ln\omega$. © *1999 American Institute of Physics.* [S1063-7761(99)02310-0]

1. INTRODUCTION

Glass formation in the absence of intrinsic disorder is a long-standing problem, but recent years have witnessed rapid progress[1–7] in the qualitative understanding of this phenomenon. Mostly this progress is due to the solution of periodic models that assume a mapping between the periodic model and the appropriate random model.[1–3] The validity of this assumption is still an open question in the general case, but it was shown that at least one periodic model allows direct study of the phase transition[5] and non-ergodic behavior below the transition[7] without any reference to a disordered model. This model describes a long-range Josephson array in a magnetic field, and another reason for the interest in this model is that it can be realized experimentally (cf. Refs. 6 and 8 for a discussion of experimental conditions).

All these results were obtained in the framework of classical statistical mechanics; the glass formation in regular quantum systems has not been addressed. The goal of this paper is to fill this gap. The problem of glass formation in disordered quantum systems is discussed in a number of papers,[9–12] which studied critical behavior near the quantum vitrification transition[9,10] and the properties of the glassy phase itself[11] using the replica approach. They found that the glass phase transition at $T=0$ indeed exists; further, it strongly resembles the classical (high-T) phase transition in the same system. The main difference is in the critical exponent of the correlation function, which decays faster than at the classical critical point: $D(t)=\overline{\langle S_j(0)S_j(t)\rangle}\propto t^{-1}$ at $T=0$ (cf. $D(t)\propto t^{-1/2}$ at non-zero T). A surprising result obtained in Ref. 11 is that at zero temperature, no replica symmetry breaking (RSB) is needed for the description of the glassy state, i.e. the replica-symmetric solution is stable at $T=0$. Since RSB is usually believed to be a signature of non-ergodicity, this result means either absence of non-ergodic behavior at $T=0$ or violation of the usual relation between RSB and non-ergodicity. We believe that the latter case is more likely because of the following. The non-ergodicity affects the full thermodynamic averaging only if higher metastable states contribute to the partition sum. It is more likely that in a typical situation the gap between the ground state and the next state remains finite at $T\to 0$, so in this limit only the ground state contributes to the full thermodynamic average and the RSB does not occur, although the system retains its non-ergodic behavior; note that a similar phenomenon is believed to occur in the classical $T=0$ limit.[13] We feel that in order to clarify this important question, an approach that is free from the ambiguities of the replica method should be employed.

Understanding quantum glass formation in a system with a regular Hamiltonian is important for the general problem of quantum computation.[14] The reason is that a quantum computer is also a quantum system with an exponential number of states, and the process of computation can be viewed as an almost adiabatic change of the external parameters, resulting in a different state. The crucial issue is how to ensure that such a process does not lead to the collapse of the density matrix due to coupling to the environment. This issue is relevant to the spin glass system as well, and one can learn about decoherence in a generic large system with an exponential number of states by addressing it.

Here we study the quantum version of a long-range Josephson array in a frustrating magnetic field, as suggested in Refs. 4–7. We consider here only the problem of glass formation, approaching the glass from the ''liquid'' (i.e., insulating) side. We show that the quantum version of this problem can be described by the same dynamical equations as the quantum disordered p-spin model studied in Ref. 15. Thus, we explicitly prove that this frustrated quantum system can be mapped onto a quantum disordered system, in complete analogy with the situation for classical problems. Further, we provide a direct numerical proof that the transition in this

model is indeed continuous as conjectured in Ref. 15 and we calculate the anomaly of the diamagnetic response associated with this transition.

Another (more physical) justification of the model is the following. It is well established, both experimentally[16,17] and theoretically[18] that usual nearest-neighbor Josephson arrays made of small superconducting islands exhibit a zero-T superconductor—insulator transition as the ratio of the Josephson coupling E_J between superconductive islands to the Coulomb energy cost $E_C=(2e)^2/2C$ for the transfer of the Cooper pair between the islands decreases. At small values of $x=E_J/E_C$, the ground state is an insulator with nonzero Coulomb gap in the excitation spectrum. At nearly critical values of $x\approx x_{cr}$, the transition between insulating and superconducting states can be triggered by application of a weak magnetic field, producing frustration of the Josephson interaction. Moreover, this transition can be split[16] into a sequence of two different transitions: superconductor→metal→insulator. Although the main qualitative features of these phenomena are understood, there is still no quantitative theory that describes quantum phase transitions in two-dimensional short-range systems, especially in the presence of frustration. Therefore, in our attempt to study the origin of a quantum glass state, we have to turn to the simplest (theoretical) model of a Josephson array with long-range interaction, which consists of long superconducting wires (instead of islands), which will enable us to employ some version of mean-field theory and reduce the problem to a zero-dimensional quantum theory with an interaction that is non-local in time.

The system that we study is a stack of two mutually perpendicular sets of N parallel thin superconducting wires with Josephson junctions at each node, located in an external transverse field H. Macroscopic quantum variables of this array are the $2N$ superconducting phases associated with each wire (e.g., the phase of the superconducting order parameter at the center of each wire). We will always assume that excitations within individual wires can be neglected, so that the whole wire is characterized by a single phase ϕ_m. In the absence of an external field the phase differences would be zero at each junction, but this is not possible at finite H, so the phases are frustrated. Here we assume that the Josephson currents are sufficiently small so that the induced fields are negligible in comparison with H (this imposes an important constraint on experimental realization of this network).[6]

The array can therefore be described by the Hamiltonian[1)]

$$\mathscr{H}=\mathscr{H}_J+\mathscr{H}_C=-E_J\sum_{m,n}\cos\left(\phi_n-\phi_m-\frac{2e}{\hbar c}\int\mathbf{A}\cdot d\mathbf{l}\right)+\frac{(2e)^2}{2}\sum_{m,n}\hat{C}^{-1}_{m,n}\frac{\partial}{\partial\phi_m}\frac{\partial}{\partial\phi_n},\qquad(1)$$

where $\mathscr{H}_J$ and $\mathscr{H}_C$ represent the Josephson and Coulomb parts of the Hamiltonian, and $\hat{C}_{m,n}$ is the capacitance matrix. There are several different contributions to $\hat{C}$: self-capacitances of the wires C_l (with respect to the substrate), the junction capacitances C_J, and the mutual capacitances of wires C_{ll}. Below we assume that the self-capacitance is the largest of all: $C_l\gg C_{ll}$, NC_J (the factor N accounts for the fact that there are N junctions along each wire). These conditions enable us to neglect all mutual capacitances and consider the matrix $C_{m,n}$ to be diagonal, with eigenvalues C_l.

It is convenient to rewrite the Hamiltonian in terms of ''spin'' variables $s_m=e^{i\phi m}$. Choosing the Landau gauge for the vector potential and introducing J_0 via $E_J=J_0/\sqrt{N}$ so that the transition temperature remains constant in the limit $N\to\infty$ at fixed J_0, we obtain

$$\mathscr{H}=-\sum_{m,n}^{2N}s_m^*\mathscr{J}_{mn}s_n+\frac{E_C}{2}\sum_n Q_n^2,\qquad(2)$$

where $Q_n\equiv -i\partial/\partial\phi_n$ is the charge operator conjugate to the phase ϕ_n, $E_C=4e^2/C_l$, and $\mathscr{J}_{mn}$ is the coupling matrix

$$\hat{\mathscr{J}}=\begin{pmatrix}0 & \hat{J}\\ \hat{J}^\dagger & 0\end{pmatrix}\qquad(3)$$

with

$$J_{jk}=\frac{J_0}{\sqrt{N}}\exp(2\pi i\alpha jk/N),\quad 1\leqslant(j,k)\leqslant N,$$

where $j(k)$ is the index labeling the horizontal (vertical) wires; $s_m=e^{i\phi_m}$, where the ϕ_m are the superconducting phases of the $2N$ wires, and $\alpha=NHl^2/\Phi_0$ is the flux per unit strip (l is the inter-node spacing and Φ_0 is the flux quantum).

Because every horizontal (vertical) wire is coupled to every vertical (horizontal) wire, the connectivity in this model is high (N) and it is accessible to a mean-field treatment (its classical version was developed in Refs. 19 and 5). For $1/N\ll\alpha<1$ there exist many metastable solutions that minimize the Josephson (''potential'') part of the Hamiltonian (2); these minima are separated by barriers that scale[4] with N. A similar (classical) long-range network with disorder was previously found to display a spin glass transition[19] for $\alpha\gg 1/N$; in the absence of short-range phase coherence between wires ($\alpha\gg 1$), it is equivalent to the Sherrington–Kirkpatrick model.[20] Physically this glassy behavior occurs because the phase differences associated with the couplings, J_{jk}, acquire random values and fill the interval $(0, 2\pi)$ uniformly. For the periodic case, this condition is satisfied in the ''incommensurate window'' $1/N\ll\alpha\leqslant 1$, for which the magnetic unit cell is larger than the system size, so that the simple ''crystalline'' phase is inaccessible.[4]

There are thus no special field values for which there are only a few minima of the potential energy, in contrast to the situation for $\alpha>1$. Below we consider the case $1/N\ll\alpha\ll 1$ only. As follows from previous studies,[4–7] the characteristic energy scale related to the potential energy $\mathscr{H}_J$ is of the order of the glass transition temperature of the classical system, $T_G\approx J_0/\sqrt{\alpha}$. The zero-$T$ transition we study here is driven by the competition between Josephson and Coulomb energies, the scale of the latter being $E_C=4e^2/C_l$. Thus, we expect that the quantum transition occurs at $J_0/\sqrt{\alpha}\sim E_C$. Our goal is to show that such a (continuous) phase transition indeed occurs, and to study the critical behavior near the transition point. Below, in the main part of the paper, we

measure all energies in units of E_C, and return to physical units only in the final expression for the critical behavior of the ac diamagnetic susceptibility.

2. QUANTUM LOCATOR EXPANSION

We develop a diagram technique for the Hamiltonian that is (2) very similar to the one employed previously[5] for the classical Langevin dynamics of the same array. The idea is to treat the Coulomb part of the Hamiltonian as the zero-level approximation, and construct an expansion in powers of the Josephson coupling constant J_0, keeping all terms of lowest order in the coordination number $1/N$. Thus, our approach can be considered a quantum version of the Thouless–Anderson–Palmer[21] method.

The diagram technique for the Matsubara Green function

$$G_{m,n}(\tau)=-\langle T_\tau s_m(\tau)s_n^\dagger(0)\rangle, \quad s(\tau)=e^{-\tau\mathscr{H}}se^{\tau\mathscr{H}} \tag{4}$$

is closely related to the one developed in Ref. 5. The Dyson equation for the frequency-dependent matrix Green function reads (note that in our units $E_C=1$)

$$\mathbf{G}_\omega=\frac{1}{\tilde{\mathbf{G}}_\omega^{-1}-(\mathbf{J}\mathbf{J}^\dagger)\tilde{\mathbf{G}}_\omega}, \tag{5}$$

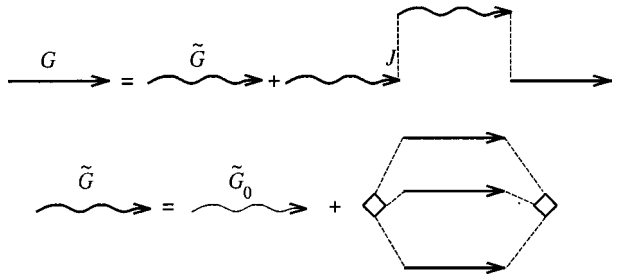
where we have introduced the local Green functions $\tilde{G}_\omega$ that is irreducible with respect to the J_{ij} lines. The matrix $(JJ^\dagger)_{ij}$ depends only on the ''distance'' $i-j$ and acquires a simple form in Fourier space:

$$(JJ^\dagger)_p=(J_0^2/\alpha)\,\theta(\alpha\pi-|p|);$$

therefore in this representation

$$G_\omega(p)=\frac{\theta(\alpha\pi-|p|)}{\tilde{G}_\omega^{-1}-\dfrac{J_0^2}{\alpha}\tilde{G}_\omega}+\frac{\theta(|p|-\alpha\pi)}{\tilde{G}_\omega^{-1}}. \tag{6}$$

Diagrammatically, Eq. (5) and the equation for the irreducible function $\tilde{G}_\omega$ are represented by the graphs shown below.



Note that the equation for $\tilde{G}$ is written in the lowest nontrivial order in α. Indeed, it is seen from Eq. (6) that the nontrivial part of the Green function, which contains critical slowing down, is of relatively small weight $\sim\alpha$. It is this long-time part of G_ω that enters into the three-line diagram and makes it proportional to α^3; more complicated diagrams either contain even higher powers of α, or are small and go as $1/N$. Since the second diagram contains single-site functions only, the whole system of equations can be written in the form

$$G(\omega)=(1-\alpha)\tilde{G}(\omega)+\hat{G}(\omega),$$

$$\hat{G}(\omega)=\frac{\alpha\tilde{G}(\omega)}{1-J_0^2\tilde{G}^2(\omega)/\alpha}, \tag{7}$$

$$\tilde{G}(\omega)=\tilde{G}_0(\omega)+\Sigma(\omega),$$

$$\Sigma(\omega)=\left(\frac{J_0^2}{\alpha}\right)^3\chi_3^2\int\hat{G}^3(t)\exp(i\omega t)dt. \tag{8}$$

Here $\chi_3\sim 1$, as in Ref. 5, is the static value of four-point vertex denoted by a square box in the diagram (we assume that, as in Ref. 5, the main critical anomaly is contained in the two-point Green function alone). Equations (7) and (8) must be solved with obvious initial condition

$$G(t=0)=\int\frac{d\omega}{2\pi}\,G(\omega)=1. \tag{9}$$

A similar normalization condition in the classical problem was sufficient to determine $\tilde{G}(\omega=0)$ exactly.[4] The same calculation is difficult in the present quantum problem, and we will not carry it out here. Instead, we use general properties of the function $\tilde{G}_0(\omega)$, namely: i) $\tilde{G}_0(0)\sim 1$, and ii) $\tilde{G}_0(\omega)$ is analytic at low ω, and has a characteristic frequency scale of the order of 1. In doing so, we do not determine the exact position of the phase transition (i.e., the critical value J_c of the coupling strength J_0), but we demonstrate the existence of a continuous transition and find the form of critical scaling.

We first analyze equations (7)–(9), omitting the term containing Σ, and using the simplest interpolation

$$\tilde{G}_0(\omega)=(\lambda+\omega^2)^{-1}.$$

Then initial condition (9) yields an equation for λ:

$$1=\frac{1}{2\sqrt{\lambda}}+\frac{\alpha}{4\sqrt{\lambda-g}}, \tag{10}$$

where $g=J_0/\sqrt{\alpha}$. Thus $\lambda\sim 1$ as long as $g\leqslant 1$. On the other hand, at $g\gg 1$ the solution is

$$\lambda-g\equiv a\approx(\alpha/4)^2.$$

The value of a determines the asymptotic decay rate of the Green function

$$G(t)=\frac{\alpha}{\sqrt{a}}\exp(-|t|\sqrt{a}) \tag{11}$$

with Σ being neglected. It will be seen below that $a\sim\alpha$ and thus $\lambda\sim 1$ near the phase transition point $g=g_c$ (we will also see that $\Sigma\sim\alpha$, and thus it is much smaller than the ω^2 term at high frequencies $\omega\gg\alpha^{1/2}$). This means that the parameter a can be considered a smooth function of g in the vicinity of g_c. Clearly, this conclusion does not depend on the model of $\tilde{G}_0(\omega)$ used in the above analysis.

Now we reintroduce $\Sigma(\omega)$ into the equations for $\hat{G}(\omega)$ and focus on its low-frequency behavior at $\omega\leqslant\sqrt{\alpha}$:

$$\hat{G}(\omega)=\frac{\alpha}{a-2\Sigma(\omega)+\omega^2}, \tag{12}$$

$$\Sigma(\omega)=\tilde{g}^6\int \hat{G}^3(t)\exp(i\omega t)dt, \tag{13}$$

where

$$\tilde{g}=g\chi_3^{1/3}\sim g.$$

Strictly speaking, Eqs. (12) and (13) do not form a complete set, since a should be determined with the use of Eq. (9) which contains high-frequency contributions. However, in this high-frequency region [which produces the main contribution to the normalization condition (9)] the contribution of $\Sigma(\omega)$ can be neglected and thus a can be treated as an external control parameter that governs the transition.

The Green function defined by Eqs. (12) and (13) acquires a singularity when $2\Sigma(0)=a$. To find the form of this singularity, we make use of the scaling ansatz $G(t)=qt^{-\nu}$ and neglect the ω^2 term in the denominator of Eq. (12). Then we find $\nu=1/2$ and $q\sim\tilde{g}^{-1}\alpha^{1/4}$. This critical-point solution matches the short-time asymptotic behavior (11) at $t\sim\alpha^{-1/2}$. The estimation of $\Sigma(0)$ that follows from the above scaling ansatz,

$$\Sigma(\omega=0)\approx\tilde{g}^6q^3\int_{\sqrt{a}}^{\infty}\frac{dt}{t^{3/2}}\approx\tilde{g}^3q^3a^{1/4}$$

yields $\Sigma(0)\approx a$ at $\tilde{g}\sim 1$ and $a\sim\alpha$, as expected. These estimates show that second-order phase transition with critical slowing down can indeed occur in the above range of parameters. In the next section we study the vicinity of the critical point in more detail.

3. GREEN FUNCTION NEAR THE $T=0$ TRANSITION POINT

To study the form of the critical singularity, it is convenient to define universal scaling functions $\mathscr{J}(\omega)$ and $\sigma(\omega)$ that do not contain the small parameter $\alpha\ll 1$, and a parameter b, which measures proximity to the critical point:

$$\hat{G}(\omega)=\mathscr{J}(\tilde{\omega}),\quad \alpha\Sigma(\omega)=\sigma(\tilde{\omega}),\quad \tilde{\omega}=\omega/\sqrt{\alpha},$$
$$b=(a-2\Sigma(0))/\alpha. \tag{14}$$

Equations (12) and (13) acquire then the following form:

$$\mathscr{J}(\tilde{\omega})=\frac{1}{b+2(\sigma(0)-\sigma(\tilde{\omega}))},$$

$$\sigma(\tilde{\omega})=\tilde{g}^6\int \mathscr{J}^3(\tilde{t})\exp(i\tilde{\omega}\tilde{t})d\tilde{t}. \tag{15}$$

Exactly at the critical point $b=0$, the solution of Eq. (15) is

$$\mathscr{J}(\tilde{\omega})=\left(\frac{\pi}{8}\right)^{1/4}\tilde{g}^{-3/2}|\tilde{\omega}|^{-1/2}. \tag{16}$$

Consider now the vicinity of the critical point, $0<b\ll 1$. It is clear from the form of the solution (16) that a similar result should be valid at $\tilde{\omega}\gg b^2$. Next we focus on the long-time, low-ω region, $\tilde{\omega}\ll b^2$, and look for the purely exponential solution

$$\mathscr{J}(\tilde{t})=\mathscr{J}_1\exp(-\tilde{t}/\tau_0). \tag{17}$$

This type of asymptotic solution is known to exist in the classical version of the same model (cf. Refs. 5 and 7). In the present problem one can show, considering the analytic structure of (15), that at $b>0$ the singularity of $\mathscr{J}(\tilde{\omega})$ closest to the real ω axis is necessarily a simple pole at some $\tilde{\omega}=i/\tau_0$; the next singularity may exist at $\tilde{\omega}\geqslant 3i/\tau_0$. Solving (15) with the ansatz (17) in the region $\tilde{t}\gg\tau_0$ determines parameters τ_0 and $\mathscr{J}_1$ as functions of b:

$$\tau_0=\frac{\sqrt{32/27}}{\tilde{g}^3}\frac{1}{b^2},\quad \mathscr{J}_1=\sqrt{\frac{27}{2}}\tilde{g}^3b. \tag{18}$$

This solution is similar to the one found in Ref. 5; however, an important difference is that in the present case the factor $\mathscr{J}_1$ scales to zero at the critical point $b=0$.

The full solution in the vicinity of the transition point should contain both (16) and (18) as asymptotic solutions, and can be written in the form

$$\mathscr{J}(\tilde{t})=\frac{1}{\sqrt{\tilde{t}}}f\left(\frac{\tilde{t}}{\tau_1}\right)+\mathscr{J}_1\exp\left(-\frac{\tilde{t}}{\tau_0}\right), \tag{19}$$

where $f(x)$ is some scaling function that approaches a constant at $x=0$ and decays rapidly as $x\to\infty$; $\tau_1\leqslant\tau_0/3$. To confirm an existence of this type of solution, we solved Eqs. (15) numerically for several values of $b\ll 1$. The results of this computation are shown in Fig. 1. Clearly, all three functions $\mathscr{J}(\omega)$ coincide in the high-ω region, where they are close to the square-root asymptotic behavior (16). The low-frequency parts (for $\omega\leqslant 0.08$) of these solutions can be made to coincide by a proper rescaling of their arguments, $\omega^*=\Lambda\omega$. Figure 2 demonstrates the linear relationship between b^{-2} and the scaling coefficient Λ, as suggested by Eqs. (18) and (19).

These results confirm the existence of $T=0$ critical behavior of the type of Eq. (19).

4. CRITICAL BEHAVIOR AT $T>0$

The above results refer to the zero-T phase transition controlled by the single parameter $g=J_0/\sqrt{\alpha}$. We found that this phase transition is a continuous one, and the corresponding critical behavior differs considerably from the behavior found in an analogous classical model.[5] In particular, at the $T=0$ critical point $g=g_c$, there is no ''plateau'' solution with approximately constant $G(t)$ at $t\to\infty$, which is known to be a distinctive property of regular classical glasses. Now we consider low but non-zero temperatures $T=\beta^{-1}$, and find out how ''classical'' critical scaling ''grows up'' from the ''quantum'' background; we also find the low-temperature shape of the phase transition curve in the (T,g) plane.

The Green function is now defined at discrete frequencies $\omega_n=2\pi nT$, and Eqs. (12) and (13) can be written in the form

$$\hat{G}(\omega_n)=\frac{\alpha}{a-2\Sigma(\omega_n)+\omega^2},$$

$$\Sigma_M(\omega_n)=\tilde{g}^6\int_0^{\beta}\hat{G}^3(t)\exp(i\omega_nt)dt. \tag{20}$$

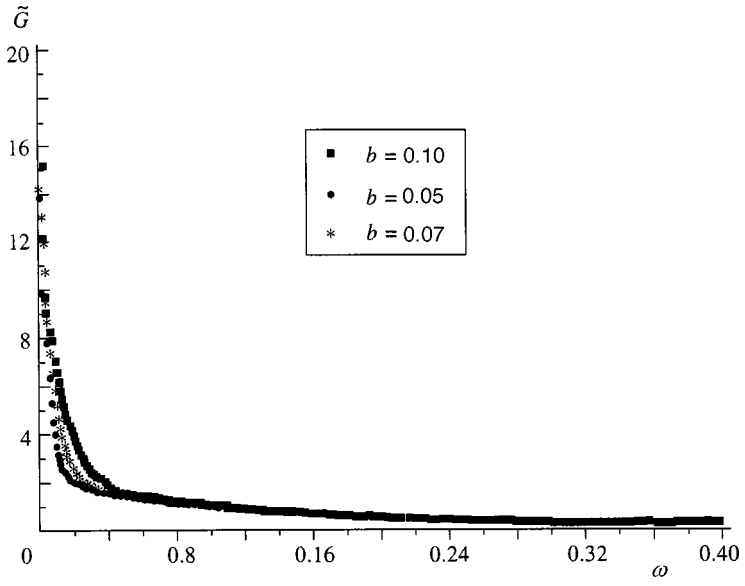


FIG. 1. Low-frequency asymptotic behavior of $\tilde{G}(\omega)$ for various b at $T=0$.

It will be convenient now to perform analytic continuation of Eqs. (20) and rewrite them in terms of real-time correlation function

$$D(t)=\langle [S(t),S(0)]_+ \rangle$$

and response function

$$\chi(t)=i[S(t),S(0)]_- \theta(t).$$

The functions $G(\omega_n)$, $D(\omega)$ and $\chi(\omega)$ are related as follows:

$$G(-i\omega+\eta)=\chi(\omega), \quad \eta\to+0,$$

$$D(\omega)=\mathrm{Im}\,\chi(\omega)\coth(\omega/2T). \tag{21}$$

After analytic continuation Eqs. (20) can be written

$$\chi(\omega)=\frac{\alpha}{\tilde{a}-2\Sigma(\omega)}, \quad \Sigma(\omega)=8\tilde{g}^6\int_0^\infty D^2(t)\chi(t)$$

$$\times(\exp(i\omega t)-1)dt, \quad \tilde{a}=a-2\Sigma(\omega=0), \tag{22}$$

where we omitted the ω^2 term, which is irrelevant in the vicinity of the critical point. Equations (22) form (together with the fluctuation-dissipation relation [second of Eqs. (21)]) a complete set that determines the critical singularity at $T>0$. Formally, Eqs. (22) coincide with the corresponding ''classical'' equations from Ref. 5, the only difference being the form of the fluctuation-dissipation relation.

We now consider the low-temperature region $T\ll\sqrt{\alpha}$. As long as we are interested in the long-time behavior $t\gg 1/T$, the correlation and response functions are related by the classical FDT: $D(\omega)=2T/\omega\,\mathrm{Im}\,\chi(\omega)$. Characteristic times relevant to (22) also belong to classical region $t\gg 1/T$. Therefore the correlation function at the transition point has the same critical behavior as in the classical case:

$$\lim_{t\to\infty} D(t)=q.$$

However, the parameter $a\equiv\lambda-g$ is determined by the ''quantum'' frequency range $\omega\gg T$, i.e., by Eq. (10). Substituting this expression into (22) yields

$$q\propto\alpha^{1/4}T^{1/2} \quad \tilde{a}\propto\alpha^{3/4}T^{1/2}. \tag{23}$$

In the short-time domain $t\ll T^{-1}$, the zero-T critical solution with $D(t)\propto\alpha^{1/4}t^{-1/2}$ is valid. Equation (23) demonstrates how the ''classical'' solution with nonzero $\lim_{t\to\infty}D(t)$ grows up with the temperature increase.

5. DIAMAGNETIC RESPONSE NEAR THE TRANSITION POINT

Correlation and response functions $D(t)$ and $\chi(t)$ are not directly measurable in our system, but they can be used to calculate a measurable physical quantity, the dynamic diamagnetic susceptibility $\chi_{\mathscr{M}}(\omega)$, as was done previously for the classical problem.[5] The total magnetic moment induced by a time-dependent external magnetic field is

$$\mathscr{M}=\frac{1}{2}\left(\frac{2e}{\hbar c}\right)l^2\sum_{mn} S_m^\dagger\,\tilde{\mathscr{J}}_{mn}S_n, \tag{24}$$

where $\tilde{\mathscr{J}}_{mn}=imn\mathscr{J}_{mn}$.[5] Then the magnetic susceptibility $\chi_{\mathscr{M}}$ can be found via the Kubo formula, $\chi_{\mathscr{M}}(t-t')=i[\mathscr{M}(t),\mathscr{M}^\dagger(t')]\theta(t)$, which leads to the expression

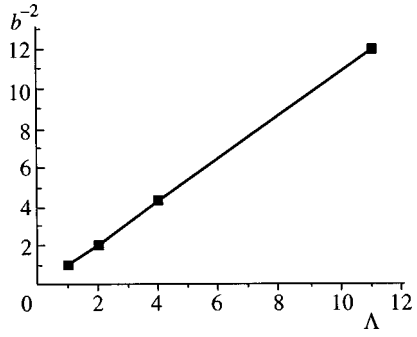


FIG. 2. The relation between scaling parameter Λ and proximity to the transition point b at $T=0$.

$$\chi_{\mathscr{M}}(\omega)=\left(\frac{2e}{\hbar c}\right)^2 l^2\int_0^\infty (e^{i\omega t}-1)\,\mathrm{Re\,Tr}\;\tilde{\mathscr{J}}\hat{\chi}(t)\tilde{\mathscr{J}}\hat{D}(t)dt. \quad (25)$$

Here we omit the term containing the irreducible four-spin correlator (of the order of $1/N$), and take into account that $\mathscr{M}(H=0)=0$. Note, that Eq. (25) formally coincides with the classical formula for magnetic response.[5] The matrix functions $\hat{D}(t)$ and $\hat{\chi}(t)$ contain elements (denoted by superscript$^{(0)}$) belonging to the same (horizontal or vertical) sublattice of our array, as well as ''off-diagonal'' elements (with superscript$^{(1)}$) that describe correlation of phases on wires of different type (horizontal/vertical). These functions are related by

$$\hat{\chi}(\omega)^{(1)}=J\tilde{G}(\omega)\hat{\chi}(\omega)^{(0)}.$$

Thus, the expression for magnetic susceptibility has the form

$$\chi_{\mathscr{M}}(\omega)=\left(\frac{2e}{\hbar c}\right)^2\left(\frac{l^2}{12}\right)^2 N^5\frac{J_0^2}{\alpha^2}\,I(\omega), \quad (26)$$

where

$$I(\omega)=\int\left(\delta(t-t_1)-\frac{J_0^2\tilde{G}^2(t-t_1)}{\alpha}\right)\times\chi(t_1)D(t_1)(e^{i\omega t}-1)\,\theta(t)dt\;dt_1. \quad (27)$$

Near the transition point, only the long-time parts of the functions $\chi(t)$ and $D(t)$ in (27) are relevant, and this expression can be reduced to the form

$$I(\omega)=(\Sigma(\omega)-\Sigma(0))\int\chi(t)D(t)e^{i\omega t}dt, \quad (28)$$

where the first factor came from the first brackets in (27); note that this vanishes in the limit $\omega\to 0$.

Using the solution (16), we obtain at the quantum critical point $J=J_c$

$$I(\omega)=\frac{\alpha}{2\pi}\left(\frac{\alpha\pi}{8\tilde{g}^6}\right)^{1/4}\sqrt{i\omega}\,\ln\omega. \quad (29)$$

Near the $T=0$ transition point at high enough frequencies

$$\omega\gg(J/J_c-1)^2\alpha^{-3/2},$$

Eq. (29) still holds. In the opposite case of low frequencies,

$$I(\omega)=\frac{8J_c^3\alpha^3}{81(J_c-J)^3}\,\frac{\omega^2\alpha^{1/2}}{g}. \quad (30)$$

Note that the parameter $\tilde{g}$ (which is known only up to factors of order 1) does not enter into the low-ω asymptotic behavior of $I(\omega)$.

Making use of Eqs. (26), (29) and (30), and returning to the original units of frequency, we finally obtain the ac diamagnetic susceptibility near the quantum transition point:

$$\chi_{\mathscr{M}}(\omega)\approx\left(\frac{2e}{\hbar c}\right)^2 l^4N^5\frac{(J_cC_l)^{1/2}}{2e}\,\sqrt{\frac{i\omega C_l}{e^2}}\ln\left(\frac{\omega C_l}{e^2}\right),$$

$$\omega\gg\frac{C_l(J-J_c)^2}{e^2\alpha^{5/2}}, \quad (31)$$

$$\chi_{\mathscr{M}}(\omega)=\left(\frac{2e}{\hbar c}\right)^2\left(\frac{l^2}{12}\right)^2 N^5\frac{2C_l\alpha^{7/2}}{81e^2J_c}\,\frac{J_c^3}{(J_c-J)^3}\,\omega^2,$$

$$\omega\ll\frac{C_l(J-J_c)^2}{e^2\alpha^{5/2}}. \quad (32)$$

These expressions are valid at frequencies $\omega\gg T/\hbar$, otherwise the ''classical'' asymptotic behavior of the Green functions should be used and will lead to frequency dependencies like those in Ref. 5.

6. CONCLUSIONS

We have shown that a regular frustrated long-range Josephson array has a quantum (zero-temperature) phase transition between the Coulomb-dominated insulator phase and a superconductive state. This transition happens when the nearest-neighbor Josephson coupling exceeds the critical value:

$$J_{ij}\sim N^{-1/2}\sqrt{\alpha}e^2/C_l,$$

where C_l is the self-capacitance of an individual wire.

We found that quantum critical behavior of the model at $J\to J_c$ is different from that of an analogous classical system:[5] at the quantum critical point

$$D(t)\propto t^{-1/2},$$

while at the classical critical point

$$q=\lim_{t\to\infty} D(t).$$

However, at any non-zero temperature a ''classical'' type of asymptotic behavior is recovered at the longest times, $t\gg\hbar/T$, leading to $q\propto T^{1/2}$. Near the $T=0$ critical point, the gap in the excitation spectrum decreases as $\tau_0^{-1}\propto(J_c-J)^2$.

Near the phase transition, the effective inductance $\mathscr{J}$ of the array, defined by

$$\mathscr{J}\propto\partial^2\chi_{\mathscr{M}}(\omega)/\partial\omega^2|_{\omega\to 0},$$

diverges as $(J_c-J)^{-3}$; this shows that the glass state has macroscopic phase rigidity (cf. also Ref. 19). Right at the critical point we find unusual frequency behavior of the complex diamagnetic susceptibility:

$$\chi_{\mathscr{M}}(\omega)\propto\sqrt{i\omega}\ln\omega.$$

The frustrated nature of couplings in our array and comparison with previous results[7] in the classical version of the same model indicates that the high-J state is a quantum glassy superconductor. The $T=0$ nonergodic properties (irreversibility, aging) remain an open question; note here that a recent study[15] of nonequilibrium glassy behavior in a p-spin spherical quantum model assumed strongly dissipative (overdamped) dynamics, whereas the dynamics relevant to a Josephson array at $T=0$ must be underdamped.

*)ABBYY, p.b.#19, 105568, Moscow, Russia.
†)E-mail: feigel@landau.ac.ru

1)Random version of the same Hamiltonian was considered in Ref. 12, where an approximate dependence of the transition temperature $T_g(E_C, E_J)$ on the ratio of charging to Josephson energy E_C/E_J was calculated. Unfortunately, even in the classical limit $E_C \to 0$ the result of Ref. 12 for T_g differs from the one obtained in Ref. 18 by the method which is free from any approximations or assumptions.

[1] J. P. Bouchaud and M. Mezard, J. Phys. I **4**, 1109 (1994); S. Franz and J. Herz, Phys. Rev. Lett. **74**, 2114 (1995); L. F. Cugliandolo, J. Kurchan, R. Monasson, and G. Parisi, J. Phys. A **29**, 1347 (1996).
[2] E. Marinari, G. Parisi, and F. Ritort, J. Phys. A **28**, 4481 (1995); M. Potters and G. Parisi, J. Phys. A **28**, 5267 (1996) and references therein.
[3] J.-P. Bouchaud, L. Cugliandolo, J. Kurchan, and M. Mezard, Physica A **226**, 243 (1996).
[4] P. Chandra, L. B. Ioffe, and D. Sherrington, Phys. Rev. Lett. **75**, 713 (1995).
[5] P. Chandra, M. V. Feigelman, and L. B. Ioffe, Phys. Rev. Lett. **76**, 4805 (1996).
[6] P. Chandra, M. V. Feigelman, M. E. Gershenson, and L. B. Ioffe, in *Complex Behavior of Glassy Systems*, ed. by M. Rubi and C. Perez-Vicente, Springer-Verlag, Berlin (1997), p. 376.
[7] P. Chandra, M. V. Feigelman, L. B. Ioffe, and D. M. Kagan, Phys. Rev. B **56**, 11553 (1997).
[8] H. R. Shea and M. Tinkham, E-prints archive cond-mat/9706179.
[9] J. Miller and D. A. Huse, Phys. Rev. Lett. **70**, 3147 (1993).
[10] J. Ye, S. Sachdev, and N. Read, Phys. Rev. Lett. **70**, 4011 (1993).
[11] N. Read, S. Sachdev, and J. Ye, Phys. Rev. B **52**, 384 (1995).
[12] J. Jose and T. K. Kopec, Phys. Rev. B **52**, 16140 (1995).
[13] M. Mezard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, J. de Phys. **45**, 843 (1984); M. Mezard, G. Parisi, and M. Virasoro, J. de Phys. Lett. **46**, L217 (1985).
[14] A. Ekert and R. Jozsa, Rev. Mod. Phys. **68**, 733 (1996); L. B. Ioffe *et al.*, Nature **398**, 679 (1999).
[15] L. Cugliandolo and G. Lozano, Phys. Rev. Lett. **80**, 4979 (1998).
[16] H. S. J. van der Zant, F. C. Fritschy, W. J. Elion, L. J. Geerlings, and J. E. Mooji, Phys. Rev. Lett. **69**, 2971 (1992); H. S. J. van der Zant, W. J. Elion, L. J. Geerlings, and J. E. Mooji, Phys. Rev. B **54**, 10081 (1996).
[17] P. Delsing, C. D. Chen, D. B. Haviland, Y. Harada, and T. Claeson, Phys. Rev. B **50**, 3959 (1994).
[18] R. Fazio and G. Schön, Phys. Rev. B **43**, 5307 (1991); R. Fazio *et al.*, Helv. Phys. Acta **65**, 228 (1992).
[19] V. M. Vinokur, L. B. Ioffe, A. I. Larkin, and M. V. Feigelman, Zh. Éksp. Teor. Fiz. **93**, 343 (1987) [Sov. Phys. JETP **66**, 198 (1987)].
[20] D. Sherrington and S. Kirkpatrick, Phys. Rev. B **35**, 1792 (1975).
[21] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philos. Mag. **35**, 593 (1977).

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

Critical behavior of conductivity in two-phase, highly inhomogeneous composites

A. A. Snarskiĭ*)

National Technical University (Kiev Polytechnic Institute), 252056 Kiev, Ukraine; Scientific and Manufacturing Enterprise ''Nauka-Servis,'' 103473 Moscow, Russia

K. V. Slipchenko

National Technical University (Kiev Polytechnic Institute), 252056 Kiev, Ukraine

V. A. Sevryukov

Scientific and Manufacturing Enterprise ''Nauka-Servis,'' 103473 Moscow, Russia

(Submitted 12 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1462–1476 (October 1999)

We study the critical behavior of the effective conductivity of a two-phase, highly inhomogeneous medium with a finite ratio of phase conductivities. We find the critical exponents above and below the percolation threshold and at the percolation threshold itself, and establish the criteria for the applicability of a percolation-like description. We also determine the requirements for linearization of the medium as a whole at the percolation threshold. Finally, we investigate the homogenization of the system near critical fields. © *1999 American Institute of Physics.* [S1063-7761(99)02410-5]

1. INTRODUCTION

The physics of nonlinear composites near the percolation threshold is currently the focus of a great deal of attention (see, e.g., the proceedings of the ETOPIM conferences[1]). The special role that nonlinearity plays in highly inhomogeneous media is related to the fact that in such media the distribution of fields and currents is highly nonuniform, which leads to deviations from Ohm's law.

There are roughly two types of nonlinearity: weak and strong. In the first, for weak fields and currents, the current–field relation for the phase with a weak nonlinearity is close to the linear Ohm's law: $\mathbf{j}=\sigma_0\mathbf{E}+\chi|\mathbf{E}|^2\mathbf{E}$, where $\mathbf{j}$ is the current density, $\mathbf{E}$ is the electric field, σ_0 is the ''ordinary'' ohmic conductivity, and χ is a constant characterizing the deviation from linearity; it is assumed that $\chi|\mathbf{E}|^2 \ll \sigma_0$. The literature devoted to the study of the effective properties of two-component composites near the percolation threshold for weakly nonlinear phases is vast (see, e.g., the literature cited in the review by Bergman and Stroud[2] and in Ref. 3).

In contrast to the current–field relation for a medium with weak nonlinearity, the relation for a medium with strong nonlinearity is far from linear even in very weak fields:

$$\mathbf{j}=\sigma|\mathbf{E}|^{\beta}\mathbf{E}. \tag{1}$$

To simplify matters, in (1) we have denoted the coefficient characterizing the properties of the medium by the same letter σ as for ordinary conductivity, and we even call it conductivity. Note, however, that only when $\beta=1$ dpes this coefficient becomes the ordinary ohmic conductivity.

When studying the effective properties of randomly inhomogeneous medium, we must first do two things: a) find the current–field relation for the medium as a whole, and b) determine the dependence of the parameters of the current–field relations on the concentrations of the phases.

Here we study the standard percolation problem: The medium consists of two phases with very different properties (in our case, conductivity), with the concentration p of the highly conducting phase (which we call the ''black'' phase) being close to the percolation threshold p_c, or $\tau \equiv (p-p_c)/p_c \ll 1$.

Kenkel and Straley[4] were the first to study the effective properties of percolation media with the law (1). They examined the situation above the percolation threshold ($p>p_c$) in which the second phase is an insulator, i.e., only the ''black'' phase conducts, and found that the volume average of the current is related to the volume average of the field $\langle\mathbf{E}\rangle$:

$$\langle\mathbf{j}\rangle=\sigma_1\tau^{t(\beta)}|\langle\mathbf{E}\rangle|^{\beta}\langle\mathbf{E}\rangle, \quad p>p_c. \tag{2}$$

Thus, since all finite resistors in the medium adhere to the same current–field relation (1), the medium as a whole naturally has the same parameter β in (2). According to Ref. 4, the critical exponent $t(\beta)$, which characterizes the changes in the current (under a constant field) as the percolation threshold is approached, is given by $t(\beta)=(d-2-\beta)\nu+(1+\beta)\zeta_R$, where ν is the critical exponent of the correlation length, d is the dimensionality of the problem, and ζ_R is close to unity for all values of d.

Blumenfeld *et al.*[5,6] found that ζ_R is essentially independent of β; according to the nodes–links–blobs (NLB) model,[7] $\zeta_R=1$. In percolation structure models[8,9] it is assumed that $\zeta_R=\alpha_1=t-\nu(d-2)$, where t is the conductivity critical exponent above the percolation threshold in the linear case. On the basis of these models, the analytic expression $t(\beta)=t+[t-\nu(d-1)]\beta$ was obtained in Ref. 10. It was found that this expression is in good agreement with the numerical and model ideas.[4–6]

Ohtsuki and Keyes[11] studied the opposite case, i.e., the first phase (''black'' resistors) is a perfect conductor, so that the potential drop across it can be neglected. The medium is below the percolation threshold and the entire potential drop is across the second phase (''white'' resistors), with a current–field relation of type (1). In this case, instead of (2) we have

$$\langle \mathbf{j} \rangle = \sigma_2 |\tau|^{-q(B)} |\langle \mathbf{E} \rangle|^{\beta} \langle \mathbf{E} \rangle, \quad p < p_c, \tag{3}$$

where, according to Ref. 11 $q(\beta) = \zeta_G - (d-2-\beta)\nu$, with ζ_G a quantity characterizing what is known as singly-disconnected bonds (SDCB) of the poorly conducting phase (see Refs. 11–13). If we put $\zeta_G = \alpha_2 = q + \nu(d-2)$ (Refs. 8, 9, 11, and 13), we obtain $q(\beta) = q + \nu\beta$ (Ref. 10), where q is the critical exponent of conductivity below the percolation threshold in the linear case. The two phases were accounted for simultaneously ($\sigma_1 \neq \infty$ and $\sigma_2 \neq 0$) in Ref. 10, where in particular it was shown that at the percolation threshold (in the smearing region)

$$\langle \mathbf{j} \rangle = (\sigma_1^{q(\beta)} \sigma_2^{t(\beta)})^{1/[t(\beta)+q(\beta)]} |\langle \mathbf{E} \rangle|^{\beta} \langle \mathbf{E} \rangle. \tag{4}$$

Scaling expressions for a two-phase medium were discussed by Hon-Chor Lee *et al.*,[14] who also gave results of computer modeling. There are also many papers devoted to the analysis of highly nonlinear media using variants of the effective medium methods (see, e.g., Refs. 15–19), using nonlinear fractal resistor networks (see, e.g., Refs. 5 and 20), and using series expansion methods (see, e.g., Refs. 6 and 21).

In all of the work cited above, it was assumed that either both phases have the same current–field relation, i.e., $\beta_1 = \beta_2$, or that one phase is a perfect conductor ($\sigma_1 = \infty$) or insulator ($\sigma_2 = 0$). It is obvious, however, that for two-phase, nonlinear media with finite phase ratios, the case $\beta_1 = \beta_2$ is more the exception than the rule. For instance, for a medium consisting of a mixture of nonlinear and linear phases, $\beta_1 \neq 0$ and $\beta_2 = 0$. Below we consider the general case in which $\beta_1 \neq \beta_2$, i.e., both phases are highly nonlinear but have differing current–field relations. In particular, this can signify that one phase is highly nonlinear while the other is linear with finite conductivity.

In Sec. 2 we deal with a special case of phase nonlinearity in the percolation region. This makes it possible, on the one hand, to introduce the main concepts (the critical local field, restrictions on the average fields, and the smearing region) and, on the other, to avoid cumbersome expressions. Section 3 is devoted to the two-dimensional case with ''reciprocal'' nonlinearity of the phases [see Eq. (34)], for which the Dykhne method provides a way to obtain exact results at the percolation threshold. In Sec. 4 we generalize the percolation approach to the d-dimensional case and arbitrary nonlinearity of the phases. Finally, in Sec. 5 we allow for field strengths close to local critical values, where the method of the self-consistent field (of the Bruggeman–Landauer type) can be applied, and fields far stronger than critical, where one can again use the percolation approach.

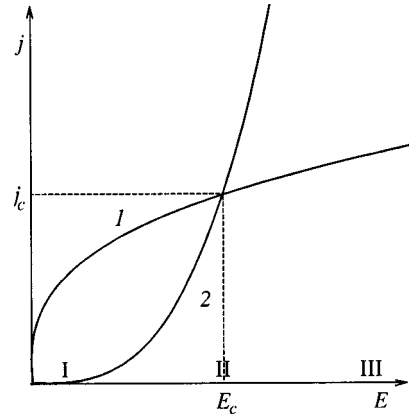


FIG. 1. Current–field relations for the first and second phases (curves *1* and *2*, respectively).

2. SPECIAL CASE OF A THREE-DIMENSIONAL TWO-PHASE NONLINEAR MEDIUM

To avoid cumbersome calculations, we begin with the special case of ''opposite'' nonlinear current–field relations for the phases (Fig. 1):

$$\mathbf{E} = \rho_1 |\mathbf{j}|^2 \mathbf{j}, \quad \mathbf{j} = \sigma_1 |\mathbf{E}|^{-2/3} \mathbf{E}, \tag{5}$$

$$\mathbf{j} = \sigma_2 |\mathbf{E}|^2 \mathbf{E}, \quad \mathbf{E} = \rho_2 |\mathbf{j}|^{-2/3} \mathbf{j}, \tag{6}$$

where $\sigma_1 = \rho_1^{-1/3}$ and $\rho_2 = \sigma_2^{-1/3}$ are constants characterizing the phase conductivities.

To determine the effective properties of the medium, we use the second stage of the hierarchical model of a percolation structure (see, e.g., Refs. 9, 22–25), which allows for percolation in each phase both above and below p_c (Fig. 2).

The resistors of the bridge and the layer (they are the same both above and below the percolation threshold) can be written in the form of voltage–current relations:

$$U_1 = \frac{N_1}{a_0^5 \sigma_1^3} J_1^3, \quad U_2 = \frac{a_0}{\sigma_2 N_2} J_2^{1/3}, \tag{7}$$

where U_1 and U_2 are the potential drops across the bridge and the layer, respectively; J_1 and J_2 are the currents, and a_0 is the minimum size in the system (the size of a ''unit'' resistor). In (7) we have allowed for the fact that the currents

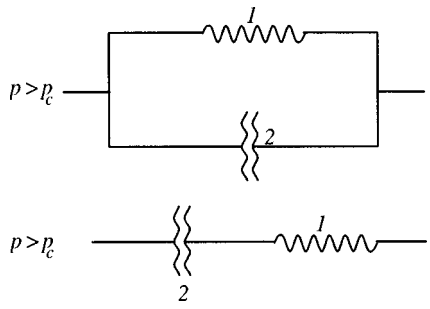


FIG. 2. Hierarchical model of a percolation structure: *1*—the bridge, which is a collection of N_1 singly-connected bonds (SCB) of the highly conducting, or ''black,'' phase, with $N_1 = |\tau|^{-\alpha_1}$, where $\alpha = \zeta_R = t - \nu(d-2)$; and *2*—the layer, which is a collection of N_2 singly-disconnected bonds (SDCB) of the poorly conducting, or ''white,'' phase, with $N_2 = |\tau|^{-\alpha_2}$, where $\alpha_2 = \zeta_G = q + \nu(d-2)$.

flowing through unit resistors in the bridge are the same and are equal to the total current through the bridge, and the potential drops across the resistors of the layer are the same and are equal to the total drop across the layer:

$$N_1=|\tau|^{-\alpha_1}, \quad N_2=|\tau|^{-\alpha_2}. \tag{8}$$

Above the percolation threshold, the main current flows in this case through the principal element of the percolation structure, the bridge. The layer yields a correction, which can be found by allowing for the fact that in this case $U_1=U_2-U_\xi$ and $J_1+J_2=J_\xi$ (Fig. 2), where U_ξ is the potential drop over dimensions of the order of the correlation length ξ, and J_ξ is the current through a cross-sectional area of order ξ^2:

$$J_\xi=\left(\frac{a_0^5\sigma_1}{N_1}\right)^{1/3}U_\xi^{1/3}\left[1+\frac{\sigma_2}{\sigma_1}\frac{N_2N_1^{1/3}}{a_0^{8/3}}U_\xi^{8/3}\right], \quad p>p_c. \tag{9}$$

Analogously, we can find the correction to the conductivity of the system below the percolation threshold provided by the bridge:

$$U_\xi=\left(\frac{\alpha_0}{\sigma_2N_2}\right)^{1/3}J_\xi^{1/3}\left[1+\left(\frac{\sigma_2N_2}{a_0}\right)^{1/3}\frac{N_1}{a_0^5\sigma_1^3}J_\xi^{8/3}\right], \quad p<p_c. \tag{10}$$

Note that both (9) and (10) hold only under certain conditions. Strong inhomogeneity means that the local conductivity of the ''black'' phase is greater than that of the ''white'' phase, and in our case this is possible only if the local fields $E_{\rm loc}$ and currents $j_{\rm loc}$ are much less than the respective critical values E_c and j_c, determined by the intersection of the local current–field relations (Fig. 1):

$$E_{\rm loc}\ll E_c, \quad j_{\rm loc}\ll j_c, \tag{11}$$

where (5) and (6) yield

$$E_c=(\sigma_1/\sigma_2)^{3/8}, \quad j_c=(\sigma_1^9/\sigma_2^{1/8})^{3/8}. \tag{12}$$

The local electric field strength in the percolation structure peaks where the ''black'' phases are separated by the thinnest layer of the ''white'' phase, $E_{\max}=U_2/a_0$, and the largest local current is at the bridge, $j_{\max}=J_1/a_0^2$, where a_0^2 is the cross-sectional area of the bridge. Passing to fields and currents averaged over the volume, $|\langle\mathbf{E}\rangle|=U_\xi/\xi$ and $|\langle\mathbf{j}\rangle|=J_\xi/\xi^2$, and allowing for the fact that essentially the entire potential drop U_ξ across a volume of the medium with a characteristic size ξ amounts to the potential drop across the layer, i.e., $U_\xi\approx U_2$, and that essentially the entire current flowing through the correlation volume, J_ξ, is concentrated at the bridge, $J_\xi\approx J_1$, we obtain expressions for $E_{\max}$ and $J_{\max}$:

$$E_{\max}=|\langle\mathbf{E}\rangle||\tau|^{-\nu}, \quad j_{\max}=|\langle\mathbf{j}\rangle||\tau|^{-2\nu}. \tag{13}$$

Using (12) and (13), we can write the conditions (11) in the form

$$|\langle\mathbf{E}\rangle|\ll\left(\frac{\sigma_1}{\sigma_2}\right)^{3/8}|\tau|^{\nu}, \quad |\langle\mathbf{j}\rangle|\ll\left(\frac{\sigma_1^9}{\sigma_2}\right)^{1/8}|\tau|^{2\nu}. \tag{14}$$

Note that the closer the system is to p_c, the smaller the value of τ and the more stringent the conditions (14).

In addition to the restrictions (14) related to the strong inhomogeneity of the two-phase medium, there are other restrictions directly related to the percolation structure. According to the hierarchical model, above the percolation threshold the principal element of the correlation volume is the bridge, while the layer provides first-order corrections: $Q_1\gg Q_2$, where Q_i is the energy dissipation in bridge and layer. Since $U_1=U_2$ for $p>p_c$ (Fig. 2), this condition becomes

$$J_1\gg J_2, \tag{15}$$

which can be combined with (7) to yield

$$U_2\ll a_0\left(\frac{\sigma_1}{\sigma_2}\right)^{3/8}|\tau|^{(3\alpha_2+\alpha_1)/8}, \quad p>p_c \tag{16}$$

According to the hierarchical model, $\alpha_1=t-\nu(d-2)$ and $\alpha_2=q+\nu(d-2)$, so that in the three-dimensional case ($d=3$) the exponent of τ is $(3q-t+4\nu)/3$. Note that expression (16) is simply the requirement that the second term in square brackets in (9) be small compared to unity.

A similar restriction exists for the case in which $p<p_c$. Here the principal element is the layer, and the condition $Q_2\gg Q_1$ with allowance for $J_1=J_2$ (Fig. 2) yields

$$U_2\gg U_1, \tag{17}$$

which can be combined with (7) to yield

$$J_1\ll a_0^2\left(\frac{\sigma_1^9}{\sigma_2}\right)^{1/8}|\tau|^{(3\alpha_1+\alpha_2)/8}, \quad p<p_c, \tag{18}$$

with $(3\alpha_1+\alpha_2)/8=(3t-q-2\nu)/8$ in the three-dimensional case. This condition means that the second term in square brackets in (10) is much less than unity.

Passing to currents and fields averaged over the volume, $|\langle\mathbf{j}\rangle|=J_\xi/\xi^2$ and $|\langle\mathbf{E}\rangle|=U_\xi/\xi$, allowing for the fact that $\xi=a_0|\tau|^{-\nu}$, and using (9) and (10), we obtain

$$\langle\mathbf{j}\rangle=\sigma_1\tau^{\tilde{t}}|\langle\mathbf{E}\rangle|^{1/3}\frac{\langle\mathbf{E}\rangle}{|\langle\mathbf{E}\rangle|}\left[1+\frac{\sigma_2}{\sigma_1}\tau^{-\tilde{\varphi}}|\langle\mathbf{E}\rangle|^{8/3}\right], \quad p>p_c, \tag{19}$$

$$\langle\mathbf{j}\rangle=\sigma_2\tau^{-\tilde{q}}|\langle\mathbf{E}\rangle|^{3}\frac{\langle\mathbf{E}\rangle}{|\langle\mathbf{E}\rangle|}\times\left[1-3\left(\frac{\sigma_2}{\sigma_1}|\tau|^{-\tilde{\varphi}}|\langle\mathbf{E}\rangle|^{8/3}\right)^3\right], \quad p<p_c, \tag{20}$$

where

$$\tilde{t}=\frac{5\nu+\alpha_1}{3}, \quad \tilde{q}=\nu+\alpha_2, \quad \tilde{\varphi}=\tilde{t}+\tilde{q}=\frac{8\nu+\alpha_1+3\alpha_2}{3}. \tag{21}$$

If we now express α_1 and α_2 in terms of t, q, and ν, in the three-dimensional case ($d=3$) we obtain

$$\tilde{t}=\frac{4\nu+t}{3}, \quad \tilde{q}=2\nu+q, \quad \tilde{\varphi}=\frac{t+3q+10\nu}{3}. \tag{22}$$

Substituting $\nu=0.88$, $t=2.0$, and $q=0.73$ (see Ref. 7), we obtain the numerical values of the critical exponents: $\tilde{t}=1.84$, $\tilde{q}=2.49$, and $\tilde{\varphi}=4.33$.

The conditions (16) and (18) can be written in terms of fields and currents averaged over the volume, and they immediately follow from the requirement that the second terms in (19) and (20) be small:

$$|\langle \mathbf{E}\rangle| \ll (\sigma_1/\sigma_2)^{3/8}\tau^{(t+3q+10\nu)/8}, \quad p>p_c, \tag{23}$$

$$|\langle \mathbf{j}\rangle| \ll (\sigma_1^9/\sigma_2)^{1/8}\tau^{(3t+q+14\nu)/8}, \quad p<p_c. \tag{24}$$

At first glance it would appear that the conditions for $p>p_c$ and $p<p_c$ are different, but if, for example, we take condition (24) and replace $|\langle \mathbf{j}\rangle|$ with $|\langle \mathbf{E}\rangle|$ (from (20) it follows that $|\langle \mathbf{j}\rangle| \approx \sigma_2|\tau|^{-\tilde{q}}|\langle \mathbf{E}\rangle|^3$), we obtain condition (23). Thus, the requirement that the second terms in (19) and (20) be small actually leads to a single condition, which can be written either for the average field [condition (23)] or for the average current [condition (24)]. In addition to the requirement that $|\tau|$ be much less than unity, these conditions mean that the system is in the critical region but outside the smearing region[26] (an analog of the smearing region of a phase transition in the theory of second-order phase transitions[27]).

Upon a further decrease in τ, the conditions (23) and (24) break down. This occurs when $|\tau| \approx \Delta$, where Δ is the size of the smearing region. The value of Δ can be found either by assuming that the second terms in (19) and (20) become approximately equal to unity or by assuming that the terms in (19) and (20) are equal (this is also the approach taken in the linear case):

$$\sigma_1\Delta^{\tilde{t}}|\langle \mathbf{E}\rangle|^{1/3} \approx \sigma_2\Delta^{-\tilde{q}}|\langle \mathbf{E}\rangle|^3. \tag{25}$$

In either case we have

$$\Delta = \left(\frac{\sigma_2}{\sigma_1}\right)^{3/(t+3q+10\nu)} |\langle \mathbf{E}\rangle|^{8/(t+3q+10\nu)}. \tag{26}$$

Since in the critical region $|\tau| \ll 1$, Δ must, of course, be much less than unity. From (26) it immediately follows that

$$|\langle \mathbf{E}\rangle| \ll (\sigma_1/\sigma_2)^{3/8}, \tag{27}$$

and since $|\tau|^{\nu}<1$, it is clear that if the condition (14) is met, the condition (27) is sure to be met.

Thus, the range of applicability of the hierarchical model in the problem with a strong nonlinearity is limited by the conditions (14), (23), and (24).

Substituting the expression (26) for τ into (19) or (20), we obtain the first term of the current–field relation in the smearing region:

$$\langle \mathbf{j}\rangle = (\sigma_1^{3(q+2\nu)}\sigma_2^{t+4\nu})^{1/(t+3q+10\nu)} \times |\langle \mathbf{E}\rangle|^{(3t+q+14\nu)/(t+3q+10\nu)} \frac{\langle \mathbf{E}\rangle}{|\langle \mathbf{E}\rangle|}. \tag{28}$$

Substituting the numerical values of the critical exponents yields $(3t+q+14\nu)/(t+3q+10\nu) \approx 1.47$. Thus, in the smearing region the ''transconductance'' assumes an intermediate value between the corresponding values above and below p_c.

3. EXACT SOLUTION OF THE TWO-PHASE PROBLEM

It is easy to write the expressions for the current–field relations above, below, and inside the smearing region for the two-dimensional case:

$$\langle \mathbf{j}\rangle = \sigma_1\tau^{(\alpha_1+2\nu)/3}|\langle \mathbf{E}\rangle|^{1/3}\frac{\langle \mathbf{E}\rangle}{|\langle \mathbf{E}\rangle|} \equiv \sigma_1\tau^{(t+2\nu)/3}|\langle \mathbf{E}\rangle|^{1/3}\frac{\langle \mathbf{E}\rangle}{|\langle \mathbf{E}\rangle|}, \quad p>p_c, \tag{29}$$

$$\langle \mathbf{j}\rangle = \sigma_2\tau^{-(\alpha_2+2\nu)/3}|\langle \mathbf{E}\rangle|^{3}\frac{\langle \mathbf{E}\rangle}{|\langle \mathbf{E}\rangle|} \equiv \sigma_2\tau^{-(q+2\nu)}|\langle \mathbf{E}\rangle|^{3}\frac{\langle \mathbf{E}\rangle}{|\langle \mathbf{E}\rangle|}, \quad p<p_c. \tag{30}$$

Here, to simplify matters, we have limited ourselves to the first term, and of course the critical exponents refer to the two-dimensional case: $t=q=1.33$ and $\nu=4/3$ (Ref. 7).

In the smearing region, (29) and (30) yield

$$\Delta = \left(\frac{\sigma_2}{\sigma_1}\right)^{8/(\alpha_1+3\alpha_2+8\nu)} |\langle \mathbf{E}\rangle|^{8/(\alpha_1+3\alpha_2+8\nu)} \equiv \left(\frac{\sigma_2}{\sigma_1}\right)^{3/4(t+2\nu)} |\langle \mathbf{E}\rangle|^{2/(t+2\nu)}. \tag{31}$$

Surprisingly, the current–field relations for such a two-dimensional system with nonlinear characteristics of the phases [Eqs. (5) and (6)] are linear in the smearing region. Indeed, substituting (31) into (29) or (30), we immediately obtain

$$\langle \mathbf{j}\rangle = (\sigma_1^3\sigma_2)^{1/4}\langle \mathbf{E}\rangle. \tag{32}$$

The condition for the validity of (32) is the same as in the three-dimensional case, (27).

Using the method proposed by Dykhne,[28] we can rigorously prove that at the percolation threshold ($p=p_c$) the system with the local current–field relations (5) and (6) is linear. To this end we examine the two-phase Dykhne medium,[28] whose phases occupy geometrically equivalent positions, i.e., interchange of the phases does not alter the properties of the medium as a whole, and in particular the current–field relations do not change. Examples of such media include a randomly inhomogeneous medium at the percolation threshold ($p=p_c$), and a medium with a special arrangement of phases, e.g., a chessboard, in which the black and white squares correspond to the ''black'' and ''white'' phases. As shown in Ref. 28, the effective conductivity of such media in the linear case is independent of the actual structure. According to Dykhne,[28] by using the transformations

$$\mathbf{j}(\mathbf{r}) = \Lambda(\mathbf{n}\times\tilde{\mathbf{E}}(\mathbf{r})), \quad \mathbf{E}(\mathbf{r}) = \frac{\mathbf{n}\times\tilde{\mathbf{j}}(\mathbf{r})}{\Lambda}, \tag{33}$$

where $\mathbf{n}$ is a unit vector normal to the two-dimensional medium, we can pass from the basic medium with the local laws

$$\mathbf{E}(\mathbf{r}) = \rho_1|\mathbf{j}(\mathbf{r})|^2\mathbf{j}(\mathbf{r}), \quad \mathbf{r}\in O_1,$$

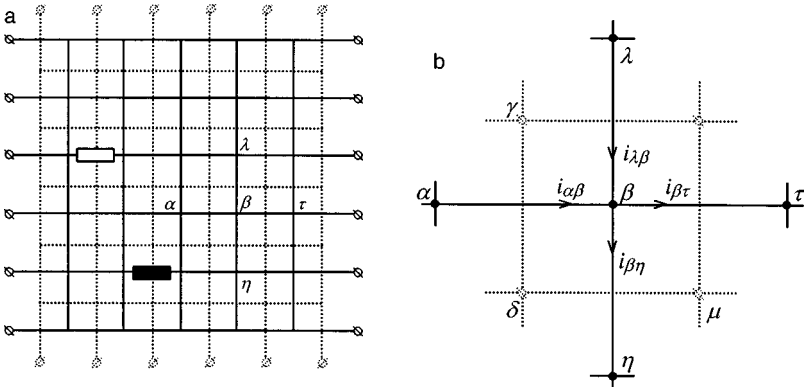


FIG. 3. Network variant of the Dykhne medium (a). The bonds depicted by solid lines are the resistors of the initial medium and by dotted lines, the resistors of the reciprocal medium. Each bond corresponds to a resistor given by (5) or (6). Here resistors of one type in the initial network intersect the resistors of the other type in the reciprocal network, and vice versa. If the reciprocal network coincides with the initial one, the network is self-dual. The black and write rectangles are two resistors corresponding to one another in a self-dual network. Figure 3(b) depicts a section of the network in (a).

$$\mathbf{j}(\mathbf{r})=\rho_2|\mathbf{E}(\mathbf{r})|^2\mathbf{E}(\mathbf{r}),\quad \mathbf{r}\in O_2, \tag{34}$$

to the reciprocal medium with the laws:

$$\tilde{\mathbf{j}}(\mathbf{r})=\Lambda^4\rho_1|\tilde{\mathbf{E}}(\mathbf{r})|^2\tilde{\mathbf{E}}(\mathbf{r}),\quad \mathbf{r}\in O_1,$$

$$\tilde{\mathbf{E}}(\mathbf{r})=\Lambda^{-4}\sigma_2|\tilde{\mathbf{j}}(\mathbf{r})|^2\tilde{\mathbf{j}}(\mathbf{r}),\quad \mathbf{r}\in O_2. \tag{35}$$

Here O_1 and $\tilde{O}_1$ are the regions with the first (''black'') phase, and O_2 and $\tilde{O}_2$ are the regions occupied by the second (''white'') phase in the basic and reciprocal media, respectively.

The requirement of reciprocity (the ''white'' phase in the initial medium becomes the ''black'' phase with the tilde, and the ''black'' phase in the initial medium becomes the ''white'' phase with the tilde) fixes the value of Λ:

$$\Lambda^4=\sigma_2/\rho_1. \tag{36}$$

By using the averaged transformations $\langle\mathbf{j}\rangle=\Lambda(\mathbf{n}\times\langle\tilde{\mathbf{E}}\rangle)$ and $\langle\mathbf{E}\rangle=\lambda^{-1}(\mathbf{n}\times\langle\tilde{\mathbf{j}}\rangle)$ with Λ from (36) we pass from the volume average of the current–field relation for the initial medium,

$$\langle\mathbf{j}\rangle=\sigma_e(|\langle\mathbf{E}\rangle|^2,|\langle\mathbf{j}\rangle|^2)\langle\mathbf{E}\rangle, \tag{37}$$

to that of the reciprocal medium,

$$\langle\tilde{\mathbf{j}}\rangle=\frac{\Lambda^2}{\sigma_e(|\langle\mathbf{j}\rangle|^2/\Lambda^2,\Lambda^2|\langle\mathbf{E}\rangle|^2)}\langle\mathbf{E}\rangle. \tag{38}$$

Since the initial medium and the medium with the tilde exhibit the same effective properties, we ought to have

$$\sigma_e(|\langle\mathbf{E}\rangle|^2,|\langle\mathbf{j}\rangle|^2)=\frac{\Lambda^2}{\sigma_e(|\langle\mathbf{j}\rangle|^2/\Lambda^2,\Lambda^2|\langle\mathbf{E}\rangle|^2)}. \tag{39}$$

We write the functional equation (39) in terms of the more convenient quantities $x=|\langle\mathbf{E}\rangle|^2$ and $y=|\langle\mathbf{j}\rangle|^2$:

$$f(y/\Lambda^2,\Lambda^2x)f(x,y)=\Lambda^2. \tag{40}$$

Now we can find the different solutions of this equation, e.g.,

$$f(x,y)=x/y,\quad f(x,y)=\Lambda\exp\{\beta(\Lambda^2x-y)\},\ldots. \tag{41}$$

Substituting these solutions into (37) yields the linear Ohm's law (32).

Our result, the linearization of a highly nonlinear local medium at the percolation threshold by a two-dimensional two-phase medium, holds in the case of the network variant of the Dykhne medium (see Fig. 3 and the caption to it). The network analog of the transformation (33) can be written (Fig. 3)

$$i_{\alpha\beta}=\Lambda\tilde{u}_{\gamma\delta},\quad u_{\alpha\beta}=\Lambda^{-1}\tilde{i}_{\gamma\delta}, \tag{42}$$

where $i_{\alpha\beta}$ and $\tilde{i}_{\gamma\delta}$ are the currents in the bonds $\alpha\beta$ and $\gamma\delta$ of the initial and reciprocal media, and $u_{\alpha\beta}$ and $\tilde{u}_{\gamma\delta}$ are potential drops across these bonds.

Given the symmetry specified in the caption to Fig. 3, a resistor of the network is determined uniquely and is independent of the network size. Figure 4 shows two examples of networks with such symmetry. Here, irrespective of the network size, there is at least one resistor that maps onto itself, the resistor *3* in Fig. 4. Recall that there are always places in

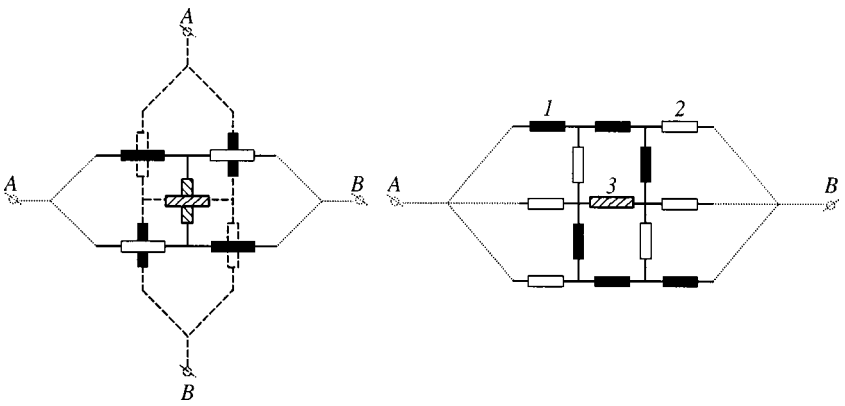


FIG. 4. Self-dual 2×2 and 3×3 networks: *1* designates resistors with the voltage–current relation $u=r_1i^3$ and are depicted by black rectangles; *2* designates resistors with the current–voltage relation $i=g_2u^3$ and are depicted by white rectangles; and *3* the resistor $r_3=\Lambda=(r_1/g_2)^{1/4}$. In the linear case, both networks yield the Dykhne result;[28] a numerical calculation yields the linear current–voltage relation $I=\Lambda U$ between the terminals A and B.

Dykhne media (points if one is dealing with a continuous problem and resistors if a network problem is involved) at which the two phases merge. In the ''chessboard'' realization of the medium such places are the corners of the squares. More details about the behavior of the current density and field strength near these corners can be found in Refs. 29 and 30.

Three networks were chosen for a numerical verification of (32), the minimum network and the two that follow it (Fig. 4). The numerical run shows, to high accuracy, that the entire medium is linear, i.e., it yields a linear current–field relation between the terminals A and B (Fig. 4). Note that since circuits consisting of nonlinear resistors with strictly monotonically increasing current–field relations can be in no more than one regime,[31] the solutions of the Kirchhoff equations for the networks depicted in Fig. 4 are unique.

The foregoing is also true of the more general (but of course still special) current–field relations for the phases:

$$\mathbf{j}=\sigma_2|\mathbf{E}|^{\gamma-1}\mathbf{E}, \quad \mathbf{E}=\rho_1|\mathbf{j}|^{\gamma-1}\mathbf{j}. \tag{43}$$

Furthermore, it can be shown that the relationship between the average field and average current in a two-dimensional medium is linear, $\langle\mathbf{j}\rangle=\Lambda\langle\mathbf{E}\rangle$, if the current–field relations for the phases exhibit a certain symmetry property: the current–field relation for one phase can be obtained via a specular reflection of that for the other phase through an oblique straight line $|\mathbf{j}|=\Lambda|\mathbf{E}|$ with appropriately chosen scales on the axes. Here two conditions must be met: both current–field relations must be single-valued monotonically increasing functions, and the disposition of the phases nust be geometrically equivalent.

4. GENERALIZATION TO ARBITRARILY STRONG NONLINEARITIES

We now examine the d-dimensional case, in which the local current–field relations for the first and second phases are, respectively,

$$\mathbf{E}=\rho_1|\mathbf{j}|^{\gamma-1}\mathbf{j}, \quad \mathbf{j}=\sigma_1|\mathbf{E}|^{(1-\gamma)/\gamma}\mathbf{E}, \tag{44}$$

$$\mathbf{j}=\sigma_2|\mathbf{E}|^{\beta-1}\mathbf{E}, \quad \mathbf{E}=\rho_2|\mathbf{j}|^{(1-\beta)/\beta}\mathbf{j}, \tag{45}$$

where $\rho_1=1/\sigma_1^{\gamma}$, and $\rho_2=1/\sigma_2^{1/\beta}$. Here γ and β are the parameters that specify the strong nonlinearity.

In particular, with $\gamma=1/\beta$ and $\sigma_2/\sigma_1\neq 0$ we obtain the case considered in Ref. 10, where both cases have the same current–field relations. The necessary calculations are carried out as in Sec. 2, so here we simply give the result.

Above and below the percolation threshold,

$$\langle\mathbf{j}\rangle=\sigma_1\tau^{\tilde{t}}|\langle\mathbf{E}\rangle|^{1/\gamma}\frac{\langle\mathbf{E}\rangle}{|\langle\mathbf{E}\rangle|}\left[1+\frac{\sigma_2}{\sigma_1}\tau^{-\tilde{\varphi}}|\langle\mathbf{E}\rangle|^{\beta-1/\gamma}\right], \quad p>p_c, \tag{46}$$

$$\langle\mathbf{j}\rangle=\sigma_2\tau^{-\tilde{q}}|\langle\mathbf{E}\rangle|^{\beta}\frac{\langle\mathbf{E}\rangle}{|\langle\mathbf{E}\rangle|}\left[1-\left(\frac{\sigma_2}{\sigma_1}|\pi|^{-\tilde{\varphi}}|\langle\mathbf{E}\rangle|^{\beta-1/\gamma}\right)^{\gamma}\right], \quad p<p_c, \tag{47}$$

where the conductivity critical exponents $\tilde{t}$ and $\tilde{q}$ have the form

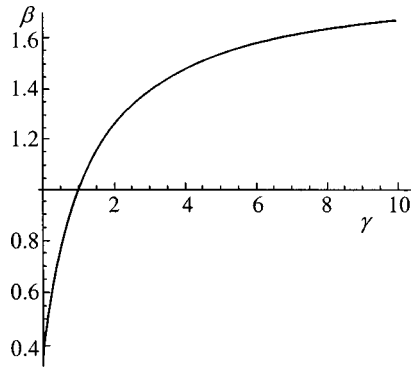


FIG. 5. The range of values of γ and β within which the current–field relation for the entire medium in the smearing region in the three-dimensional case becomes linear. In the two-dimensional case, this range is the straight line $\gamma=\beta$.

$$\tilde{t}=\frac{\alpha_1-\nu+\nu\gamma(d-1)}{\gamma}=\frac{t+\nu(d-1)(\gamma-1)}{\gamma}, \tag{48}$$

$$\tilde{q}=\alpha_2+\nu\beta-\nu(d-1)=q+\nu(\beta-1), \tag{49}$$

and the so-called scaling critical exponent is

$$\tilde{\varphi}=\tilde{t}+\tilde{q}=\frac{\alpha_1+\gamma\alpha_2+\nu(\beta\gamma-1)}{\gamma}=\frac{t+\gamma q+\nu[(\gamma-1)(d-1)+\gamma(\beta-1)]}{\gamma}. \tag{50}$$

Note that the second term in square brackets in (46) coincides with the expression in parentheses in (47), which means that there is a unique scaling exponent. This, in particular, implies that the problem has a unique smearing region Δ, whose value is determined by the condition that the second term in (46) or (47) is of order 1:

$$\Delta=\left(\frac{\sigma_2}{\sigma_1}|\langle\mathbf{E}\rangle|^{\beta-1/\gamma}\right)^{1/\tilde{\varphi}}. \tag{51}$$

Substituting this into $\langle\mathbf{j}\rangle=\sigma_1\tau^{\tilde{t}}|\langle\mathbf{E}\rangle|^{1/\gamma}\langle\mathbf{E}\rangle/|\langle\mathbf{E}\rangle|$ or $\langle\mathbf{j}\rangle=\sigma_2\tau^{-\tilde{q}}|\langle\mathbf{E}\rangle|^{\beta}\langle\mathbf{E}\rangle/|\langle\mathbf{E}\rangle|$ yields the current–field relation in the smearing region:

$$\langle\mathbf{j}\rangle=(\sigma_1^{\tilde{q}}\sigma_2^{\tilde{t}})^{1/\tilde{\varphi}}|\langle\mathbf{E}\rangle|^{(1/\tilde{\varphi})(\beta\tilde{t}+\tilde{q}/\gamma)}\frac{\langle\mathbf{E}\rangle}{|\langle\mathbf{E}\rangle|}. \tag{52}$$

So as not to complicate the discussion, the criteria of applicability of the percolation approach for the general case (they are similar to the criteria for the special case (11), (14), (16), (18), etc.) are given in the Appendix.

Note that the linearization of the system in the two-dimensional case at the percolation threshold, Eq. (32), is more the exception than the rule, and occurs only for specially selected local laws of the type (5), (6), or (43). In the parameter space (γ,β), the linearization regions of the system are the same in the two- and three-dimensional cases. The shape of this region can easily be obtained by setting the exponent of $\langle\mathbf{E}\rangle$ in (52) to unity (Fig. 5).

In contrast to the linear case, the smearing region (51) now depends on the field applied to the sample, so that a change in $\langle\mathbf{E}\rangle$ can initiate a transition of the system from one

region to another, for example, from (46) or (47) to (52). Such a transition changes the shape of the current–field relation. Note that even if one phase is linear, e.g., $\gamma=1$, the current–field relation in the smearing region fails to match a single current–field relation for the phases constituting the medium.

5. BEYOND REGIONS WITH $E \ll E_c$ AND $j \ll j_c$

The aforesaid holds under certain conditions (see the Appendix), i.e., for fields and currents that are small compared to E_c and j_c. In this section we investigate the behavior of the system near $|\langle \mathbf{E} \rangle| = E_c$.

It can easily be shown that at $|\langle E \rangle| = E_c$ the system becomes homogeneous. To prove this, we consider the network variant of the problem. For the sake of definiteness we assume that the terminals on a $\xi \times \xi \times \xi$ sample are arranged from left to right and a fixed potential drop U_ξ is maintained across the terminals. To each resistor positioned along the terminals we apply a voltage U_c ($U_c = E_c a_0$). Then all the resistors through which the current flows become identical. At each of these elements of the medium the field strength is E_c, and the average field strength $|\langle \mathbf{E} \rangle|$ is also equal to E_c. Such a distribution of fields satisfies the boundary conditions for any concentration p of the ''black'' phase. Since the solution of the problem (the regime in which the medium operates) is unique for monotonic current–field relations,[31] the proposed distribution of fields ($\mathbf{E} = \langle \mathbf{E} \rangle$ and $|\langle \mathbf{E} \rangle| = E_c$) is the true distribution.

In the continuum variant of the medium, the local laws (44) and (45) in the phases are the same at $|\langle \mathbf{E} \rangle| = E_c$:

$$\mathbf{E}(\mathbf{r}) = E_c \mathbf{k}_0, \quad \mathbf{j}(\mathbf{r}) = j_c \mathbf{k}_0, \tag{53}$$

where $\mathbf{k}_0$ is the direction of $\langle \mathbf{E} \rangle$ independent of the coordinates. Thus, these fields and currents (53) satisfy the equations (div $\mathbf{j}=0$ and curl $\mathbf{E}=0$) and the boundary conditions at the phase boundary ($\mathbf{E}_\tau|_1 = \mathbf{E}_\tau|_2$ and $\mathbf{j}_n|_1 = \mathbf{j}_n|_2$). Since the solution is unique, (53) is also the solution of the problem at the point $|\langle \mathbf{E} \rangle| = E_c$.

Near $|\langle \mathbf{E} \rangle| = E_c$, the medium is weakly inhomogeneous, so that to determine the effective characteristics we can use the effective-average approximation. Hui *et al.*[16] used this approximation to describe conductivity in systems with

$$\mathbf{j}(\mathbf{r}) = \chi_a |\mathbf{E}|^\alpha \mathbf{E}, \quad \mathbf{j}(\mathbf{r}) = \sigma_b \mathbf{E}. \tag{54}$$

In our notation, $\alpha+1 \to \gamma$, $\beta \to 1$, $\chi_\alpha \to \sigma_1$, and $\sigma_b \to \sigma_2$. According to Ref. 16, $\langle \mathbf{j} \rangle = \sigma_e(\langle \mathbf{E} \rangle)\langle \mathbf{E} \rangle$, where

$$\sigma_e = \frac{(1-2p)(\sigma_2 - \sigma_1 \langle |\mathbf{E}|^{\gamma-1} \rangle_1)}{2} - \frac{\sqrt{(1-2p)^2(\sigma_2 - \sigma_1 \langle |\mathbf{E}|^{\gamma-1} \rangle_1)^2 + 4\sigma_2\sigma_1 \langle |\mathbf{E}|^{\gamma-1} \rangle_1}}{2}. \tag{55}$$

Here $\langle \cdots \rangle_1$ designates averaging over the first phase, and p is the concentration of the nonlinear (first) phase; it is assumed that $\langle |\mathbf{E}|^{\gamma-1} \rangle_1 \approx (\langle |\mathbf{E}|^2 \rangle_1)^{(\gamma-1)/2}$ for the average fields, and that the average field $\langle |\mathbf{E}|^2 \rangle_1$ can be found from the self-consistency conditions:

$$\langle |\mathbf{E}|^2 \rangle_1 = \frac{|\langle \mathbf{E} \rangle|}{2p} \left\{ -(1-2p) + \frac{2\sigma_2 - (1-2p)^2(\sigma_2 - \sigma_1 \langle |\mathbf{E}|^2 \rangle_1^{(\gamma-1)/2})}{\sqrt{(1-2p)^2(\sigma_2 - \sigma_1 \langle |\mathbf{E}|^2 \rangle_1)^{(\gamma-1)/2} + 4\sigma_2\sigma_1 (\langle |\mathbf{E}|^2 \rangle_1)^{(\gamma-1)/2}}} \right\}. \tag{56}$$

We introduce a parameter ε that characterizes the small deviation of the field strength from E_c:

$$\langle |\mathbf{E}|^2 \rangle_1 = E_c^2 (1+\varepsilon)^2. \tag{57}$$

Then to first order in ε we obtain

$$|\langle \mathbf{E} \rangle| = E_c \left[1 + \left(\frac{\gamma-1}{4}(1-p) - 1 \right) \varepsilon \right]. \tag{58}$$

Substituting (57) and (58) into (55) yields the concentration dependence of the effective conductivity near E_c:

$$\sigma_e = \sigma_2 \left(1 - 2 \frac{\gamma-1}{(1-p)(\gamma-1) - 2} p \frac{|\langle \mathbf{E} \rangle| - E_c}{E_c} \right). \tag{59}$$

Of course, this expression is valid when the second term in the parentheses is small compared to unity. This condition imposes additional restrictions on the values of γ and p, in addition to the requirement that $|\langle \mathbf{E} \rangle|$ be close to E_c, i.e., that $(|\langle \mathbf{E} \rangle| - E_c)/E_c$ be much less than unity. For instance, at $\gamma=2$ the concentration dependence of σ_e is valid for any value of p, while at $\gamma=4$ it is valid only for values of p far from 0.5.

How does the system behave when $|\mathbf{E}| \gg E_c$? We assume that local fields can only increase as the average field $|\langle \mathbf{E} \rangle|$ increases. This means that when $|\langle \mathbf{E} \rangle| \gg E_c$, the elements of the ''black'' and ''white'' phases across which the potential drop is the greatest and through which the main current flows exchange places. Thus, we have the same percolation structure as when $|\langle \mathbf{E} \rangle| \ll E_c$, only now the bridges consist of the ''white'' phase and the layers of the ''black'' phase, and in (46), (47), (51), and (52) we must interchange σ_1 and σ_2 and replace p with $1-p$.

6. CONCLUSION

We have come to the conclusion that in the problem of the effective conductivity in a two-phase, highly nonlinear medium with a finite phase ratio there are three regions of fields. For the first and third regions (I, $|\mathbf{E}| \ll E_c$; III,

$|\mathbf{E}| \gg E_c$) it is possible to use a percolation-like approach, and in a certain sense the medium is highly inhomogeneous. In the second, intermediate, region (II, $|\mathbf{E}| \approx E_c$) the medium is weakly inhomogeneous and self-consistent field methods apply. At the point $|\mathbf{E}| = E_c$ the system becomes homogeneous and the effective conductivity can be found exactly. We believe that such homogenization must lead to a sharp decrease in the level of flicker noise, which, being the second moment of Joule heating, increases with the degree of inhomogeneity.

We are grateful to P. M. Hui for sending copies of important publications and to A. M. Satanin for helpful discussions. This work was partially supported by the International Science and Education Program (Projects QSU082187 and PSU082057).

APPENDIX

The restrictions [similar to (23) and (24)] imposed on the expressions (46) and (47) by the requirement that the principal element of a percolation structure above the percolation threshold be a bridge and below the threshold, a layer, have the form

$$|\langle \mathbf{j} \rangle| \ll \left(\frac{\sigma_1^\beta}{\sigma_2^{1/\gamma}} \tau^{\beta \tilde{t} + \tilde{q}/\gamma} \right)^{1/(\beta - 1/\gamma)}, \tag{A1}$$

$$|\langle \mathbf{E} \rangle| \ll \left(\frac{\sigma_1}{\sigma_2} \tau^{\tilde{\varphi}} \right)^{1/(\beta - 1/\gamma)}. \tag{A2}$$

As in (46) and (47), the expressions transform into one another when $|\langle \mathbf{j} \rangle|$ is replaced by $|\langle \mathbf{E} \rangle|$, or $|\langle \mathbf{E} \rangle|$ by $|\langle \mathbf{j} \rangle|$.

The restrictions imposed on (46) and (47) by the requirement that the local fields and currents be weak in comparison to the critical values j_c and E_c can be written in the form of conditions for the average fields and currents:

$$|\langle \mathbf{j} \rangle| \ll \left(\frac{\sigma_1^\beta}{\sigma_2^{1/\gamma}} \right)^{1/(\beta - 1/\gamma)} \tau^{\nu(d-1)}, \tag{A3}$$

$$|\langle \mathbf{E} \rangle| \ll \left(\frac{\sigma_1}{\sigma_2} \right)^{1/(\beta - 1/\gamma)} |\tau|^{\nu}. \tag{A4}$$

These conditions are similar to (14), where (A3) follows from the restriction on the maximum local current density in the bridge, and (A4) from the restriction on the maximum local field in the layer. In this form, (A3) and (A4) are applicable both above and below the percolation threshold. However, (A3) can be written as a restriction on the average field, and (A4) as a restriction on the average current density:

$$|\langle \mathbf{E} \rangle| \ll \left(\frac{\sigma_1}{\sigma_2} \right)^{1/(\beta - 1/\gamma)} \tau^{-\gamma[\tilde{t} - \nu(d-1)]}, \quad p > p_c, \tag{A5}$$

$$|\langle \mathbf{E} \rangle| \ll \left(\frac{\sigma_1}{\sigma_2} \right)^{1/(\beta - 1/\gamma)} |\tau|^{(1/\beta)[\tilde{q} + \nu(d-1)]}, \quad p < p_c, \tag{A6}$$

$$|\langle \mathbf{j} \rangle| \ll \left(\frac{\sigma_1^\beta}{\sigma_2^{1/\gamma}} \right)^{1/(\beta - 1/\gamma)} \tau^{\tilde{t} + \nu/\gamma}, \quad p > p_c, \tag{A7}$$

$$|\langle \mathbf{j} \rangle| \ll \left(\frac{\sigma_1^\beta}{\sigma_2^{1/\gamma}} \right)^{1/(\beta - 1/\gamma)} |\tau|^{-\tilde{q} + \beta\nu}, \quad p < p_c, \tag{A8}$$

In the smearing region, the requirement that Δ be small yields

$$|\langle \mathbf{E} \rangle| \ll \left(\frac{\sigma_1}{\sigma_2} \right)^{1/(\beta + 1/\gamma)}, \tag{A9}$$

which is similar to the inequality (27). If we use the average current density, (A9) can be combined with (52) to obtain

$$|\langle \mathbf{j} \rangle| \ll \left(\frac{\sigma_1^\beta}{\sigma_2^{1/\gamma}} \right)^{1/(\beta + 1/\gamma)}. \tag{A10}$$

The conditions that are imposed on (52) in the smearing region by the requirement that the local fields and current be small compared to the critical values [which are similar to (A3)–(A8)] coincide with (A9) and (A10) if expressed in terms of the average field and currents.

*)E-mail: asnar@phys.carrier.kiev.ua

[1] *Proceedings of the Fourth International Conference on Electrical Transport and Optical Properties of Inhomogeneous Media*, Physica A **241**, 1 (1997).
[2] D. J. Bergman and D. Stroud, Solid State Phys. **46**, 147 (1992).
[3] A. A. Snarskiĭ and S. I. Buda, Élektrichestvo No. **2**, 67 (1998); Zh. Tekh. Fiz. **68**(6), 5 (1998) [Tech. Phys. **43**, 619 (1998)].
[4] S. W. Kenkel and J. P. Straley, Phys. Rev. Lett. **49**, 767 (1982); J. P. Straley and S. W. Kenkel, Phys. Rev. B **29**, 6299 (1984).
[5] R. Blumenfeld and A. Aharony, J. Phys. A **18**, L443 (1985).
[6] Y. Meir, R. Blumenfeld, A. Aharony, and A. B. Harris, Phys. Rev. B **34**, 3424 (1986).
[7] D. Staufter and A. Aharony, *Introduction to Percolation Theory*, Taylor & Francis, London (1992).
[8] A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **91**, 1405 (1986) [Sov. Phys. JETP **64**, 828 (1986)].
[9] A. A. Snarskii, A. E. Morozovsky, A. Kolek, and A. Kusy, Phys. Rev. E **53**, 5596 (1996).
[10] A. E. Morozovsky and A. A. Snarskii, Int. J. Electron. **73**, 925 (1992).
[11] T. Ohtsuki and T. Keyes, J. Phys. A **17**, L559 (1984).
[12] A. Coniglio and H. E. Stanley, Phys. Rev. Lett. **52**, 1068 (1984).
[13] D. C. Wright, D. J. Bergman, and Y. Kantor, Phys. Rev. B **33**, 396 (1986).
[14] Hon-Chor Lee, Wing-Hon Siu, and K. W. Yu, Phys. Rev. B **52**, 4217 (1995).
[15] P. M. Hui, P. Cheung, and Y. R. Kwong, Physica A **241**, 301 (1997).
[16] P. M. Hui, Y. F. Woo, and W. M. V. Wan, J. Phys.: Condens. Matter **7**, L593 (1995).
[17] Tat-Sang Choy and K. W. Yu, Phys. Rev. B **52**, 3341 (1995).
[18] Hon-Chor Lee, Kin-Pong Yuen, and K. W. Yu, Phys. Rev. B **51**, 9317 (1995).
[19] Hon-Chor Lee, K. W. Yu, and G. Q. Gu, J. Phys.: Condens. Matter **7**, 8785 (1995).
[20] R. Blumenfeld, Y. Meir, A. B. Harris, and A. Aharony, J. Phys. A **19**, L791 (1986).
[21] Y. Meir, R. Blumenfeld, A. B. Harris, and A. Aharony, Phys. Rev. B **36**, 3950 (1987).
[22] A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **95**, 1844 (1989) [Sov. Phys. JETP **68**, 1066 (1989)].
[23] A. E. Morozovskiĭ and A. A. Snarskiĭ, JETP Lett. **52**, 244 (1990).
[24] A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **102**, 683 (1992) [Sov. Phys. JETP **75**, 366 (1992)].
[25] A. E. Morozovsky and A. A. Snarskii, Int. J. Phys. **78**, 135 (1995).
[26] A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
[27] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3rd ed., Pergamon Press, Oxford (1980).
[28] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].

[29] A. M. Satanin, V. V. Skuzovatkin, and S. V. Khor'kov, JETP Lett. **64**, 538 (1996).
[30] A. M. Satanin, V. V. Skuzovatkin, and S. V. Khor'kov, Zh. Éksp. Teor. Fiz. **112**, 643 (1997) [JETP **85**, 351 (1997)].
[31] L. V. Danilov, P. N. Matkhanov, and E. S. Filippov, *Theory of Nonlinear Electric Circuits* [in Russian], Énergoatomizdat, Leningrad (1980), p. 177.

Translated by Eugene Yankovsky

Hypersensitivity to weak ac signals induced by colored multiplicative noise in a system with *on-off* intermittency

O. V. Gerashchenko*)

B. P. Konstantinov St. Petersburg Institute of Nuclear Physics, Russian Academy of Sciences, 188350 Gatchina, Leningrad Region, Russia
(Submitted 17 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1477–1483 (October 1999)

An experimental study is made of the effect of the finite correlation time of controlling multiplicative noise on the hypersensitivity to weak ac signals in systems with *on-off* intermittency. It is shown that the gain has a maximum for an optimal choice of correlation time and intensity of the controlling noise. © *1999 American Institute of Physics.* [S1063-7761(99)02510-X]

1. INTRODUCTION

A novel phenomenon, namely hypersensitivity to ac signals induced by white multiplicative noise, has been shown[1–3] theoretically and experimentally to occur in the case of a simple nonlinear system with *on-off* intermittency, an overdamped Kramers oscillator. Real stochastic processes, however, always have a finite correlation time, i.e., white noise is an abstraction. Thus, a study of the effect of colored noise on the hypersensitivity is a natural step in the study of this interesting phenomenon.

In this paper the effect of the correlation time of controlling multiplicative noise on the hypersensitivity phenomenon in a nonlinear system which displays *on-off* intermittency is investigated experimentally in an analog electronic circuit.

Previous papers[1–3] examined an overdamped Kramers oscillator with controlling multiplicative δ-correlated noise. We recall the basic results in brief. Consider the stochastic differential equation for this system:

$$\frac{dx}{dt}=[\lambda+\beta\eta(t)]x-Ux^3+E(t), \tag{1}$$

where $\eta(t)$ is a δ-correlated Gaussian random process with intensity β, and $E(t)$ is a periodic zero-mean signal. This is the equation for a massless particle in a biquadratic potential with a potential barrier whose height is modulated by white Gaussian noise. In the adiabatic approximation, when the time to establish a stationary probability density $F(x)$ is much shorter than the period of the signal $E(t)$, we can write $F(x)$ in the form

$$F(x)=C|x|^{\alpha-1}\theta(\mathrm{sgn}(E(t)x))\phi(x,x_1,\beta), \tag{2}$$

where θ is the Heaviside step function, C is a normalization constant, and the form of the function $\phi(x,x_1,\beta)$ is determined by the form of the nonlinearity in the potential, leading to a cutoff of $x(t)$ at the characteristic scale length x_1 of the system.

Clearly, the main parameter governing the dynamics of this kind of system is α:

$$\alpha=2\lambda/\beta^2. \tag{3}$$

The condition $|\alpha|\ll 1$ is the condition for *on-off* intermittency for the system (1), where the system is suddenly excited from the laminar phase to a cutoff x_1 determined by the size of the system, and then returns to the laminar phase. Here the stationary probability density has the power-law form

$$F(x)\sim|x|^{\alpha-1}. \tag{4}$$

Note the characteristic features of the solution (2): first, it has a power-law dependence of the probability density over a wide range $A_0\sim\exp(-1/|\alpha|)<x<x_1$ for $|\alpha|\ll 1$, which is typical of the *on-off*-effect,[4–6] and second, it is sensitive to the sign of the signal $E(t)$ for signal amplitudes $A>A_0$. We emphasize that A_0 is exponentially small in the parameter $1/|\alpha|$.

The expression for the gain when $|\alpha|\ll 1$ has the form

$$I=\frac{\langle|x(t)|\rangle}{A}\sim\begin{cases}(1/A)\ln(1/A), & z\ll 1,\\ |\alpha|/A^{1-|\alpha|}, & z\gg 1, \quad \alpha<0,\end{cases} \tag{5}$$

where

$$z=|\alpha|\ln(1/A). \tag{6}$$

Thus, in the *on-off* intermittency region, a stochastic system controlled by external multiplicative noise can amplify an ultraweak ac signal up to a cutoff determined by the typical dimensions of the system with a gain $I\sim 1/A$, where A is the signal amplitude.[1–3]

It can be shown that in the case of ''colored'' dichotomous noise, the parameter α is given by:

$$\alpha=\frac{\lambda\gamma}{\Delta^2-\lambda^2}\approx\frac{\lambda\gamma}{\Delta^2}, \quad \lambda\ll\Delta, \tag{7}$$

where $1/\gamma$ is the correlation time for the noise, Δ is the noise amplitude, and λ is a term linear in the dynamic variable in Eq. (11). $F(x)$ has the power-law form (4) in the *on-off* intermittency region. Since *on-off* intermittency is observed when $|\alpha|\ll 1$, it can be seen from Eq. (7) that the transition to the ''white'' noise limit requires a simultaneous reduction

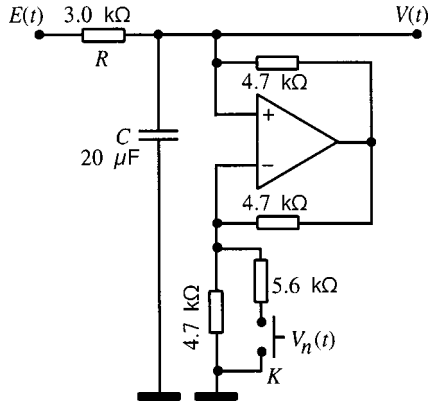


FIG. 1. Analog electronic circuit used in the experiment. The operational amplifier is a KR544UD1 and the relay *K* is an RÉS49.

in λ, so that for a finite value of λ and sufficiently large γ, the system will not be hypersensitive and the gain will decrease. If λ remains constant, the noise intensity Δ must increase to maintain *on-off* intermittency (and hypersensitivity).

2. EXPERIMENT

Figure 1 shows the analog circuit employed in the experiment. It consists of a capacitor and nonlinear element with conductivity $G(t)$ and control voltage $V_n(t)$ connected in parallel, the latter being implemented with an operational amplifier and switch. The controlling voltage $V_n(t)$ consisted of dichotomous (telegraphic) nose. The input rectangular signal $E(t)$, with zero mean, amplitude A, and period T, was fed through the resistor R.

The current–voltage characteristic of the circuit is shown in Fig. 2. It is clear that the controlling telegraphic noise changes the sign of the conductivity of the circuit.

Let us write the Kirchhoff equation for our circuit:

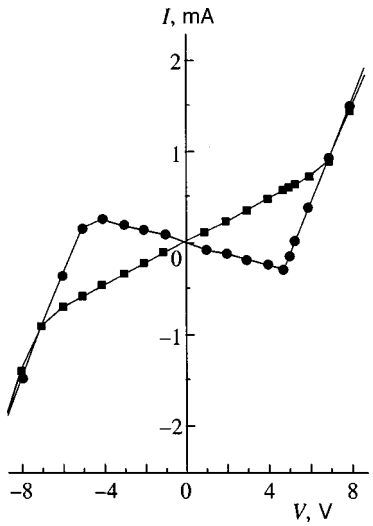


FIG. 2. Static current–voltage characteristic of the electronic circuit shown in Fig. 1. For $|V|<V_0\approx 5$ V, multiplicative noise modulates the conductivity of the circuit. At $|V|>7$ V, the branches of the current–voltage characteristic coincide.

$$\frac{E(t)-V(t)}{R}=C\frac{dV}{dt}+I(V), \tag{8}$$

where $V(t)$ is the unknown (output) voltage, CdV/dt is the current through the capacitor, $I(V)$ is the current through the nonlinear element with fluctuating conductivity, and $E(t)$ is the input signal.

Since Eq. (2) implies that the main contribution to the stationary probability density $F(V)$ comes from low $V(t)$, while the role of the nonlinearity reduces to a cutoff at higher voltages $V(t)\approx V_1$, where V_1 is determined from the condition $I(V_1)=0$, the behavior of the system in the linear region is of interest. We rewrite Eq. (8) for $|V(t)|<V_0$. Then, for the current $I(V)$ we can write

$$I(V)=(G_0+g(t))V=[G_0+g_0\xi(t)]V, \tag{9}$$

where $g(t)$ is the fluctuating conductivity, which takes values $\pm g_0$ at random. Its correlation function is

$$\langle g(0)g(t)\rangle=g_0^2\langle\xi(0)\xi(t)\rangle=g_0^2\exp(-\gamma_0|t|). \tag{10}$$

Substituting Eq. (9) in Eq. (8) and transforming to dimensionless time $\tau=t/RC$, we obtain

$$\frac{dV}{d\tau}=[\lambda+\Delta\xi(\tau)]V+E(\tau), \tag{11}$$

with

$$\langle\xi(0)\xi(\tau)\rangle=\exp(-\gamma|\tau|),\quad \lambda=-R|G_0|,$$
$$\Delta=Rg_0,\quad \gamma=RC\gamma_0. \tag{12}$$

Clearly, noise enters Eq. (11) multiplicatively, i.e., it is multiplied by the variable V.

In our case, the current–voltage characteristic of Fig. 2 yields $\lambda\approx-0.09$ and $\Delta\approx 0.27$.

Figure 3 shows the output voltage $V(t)$ when a rectangular signal of amplitude $A=50$ mV and frequency $f_s=0.3$ Hz is applied to the input of the circuit for different values of γ. It is clear that the system reacts to a small input signal by ''surging'' to the cutoff $|V_1|\approx 5$ V, and thereby amplifies it. This behavior is typical of *on-off* intermittency, where a system that is in a laminar phase is excited to cutoff and again undergoes a transition to the laminar phase. It is also clear that for $\gamma<f_s$, switching does not occur for every change of sign in the signal (adiabaticity breaks down), and the fundamental harmonic of the signal is essentially absent from the spectrum. Increasing γ leads to switching at each half period of the signal, and therefore to further amplification.

Beyond the scaling dependence (4) for the stationary probability density, *on-off* intermittency is characterized by a power-law probability density FOR the length of the laminar phase,[4–6]

$$P_L(L)\propto L^{-3/2}, \tag{13}$$

where L is the length of the laminar segment.

The laminar phase was determined experimentally in the following way: the generated $V(t)$ contained a maximum spike $V_{\max}\approx V_1\approx 5$ V; the laminarity threshold p was specified in fractions of $V_{\max}$ and the laminar phase $V(t)$ was

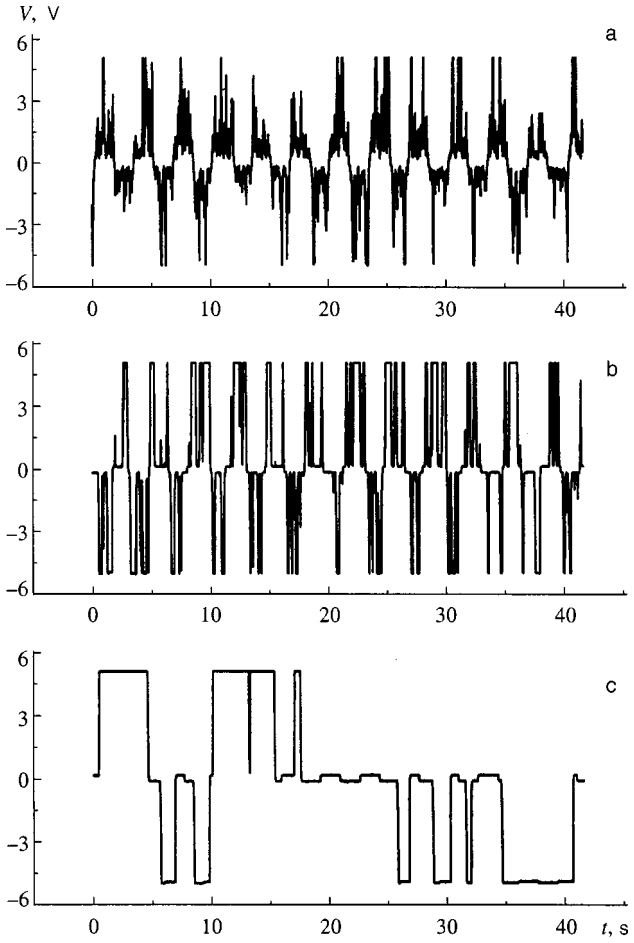


FIG. 3. Output voltage $V(t)$ for a rectangular signal of amplitude $A=50$ mV and frequency 0.3 Hz (0.0018 in dimensionless units) applied to the circuit, for $\gamma=1.14$ (a), 0.065 (b), and 0.0068 (c).

defined by the condition $V(t)<pV_{\max}$. Figure 4 shows a histogram of $P_L(L)$ at constant input voltage $E(t)=E=15$ mV for a laminarity threshold $p=0.1$, which is in good agreement with the theoretical dependence (13). Deviations from Eq. (13) occur over characteristic times of $1/\gamma$ (a reduction in P_L for $L\sim 1/\gamma$) and $1/\lambda$ (an exponential decrease at large $L\sim 1/|\lambda|$). Measuring the variation $P_L(L)$ provides another way to estimate the parameters γ and λ. From Fig. 4 we obtain $\gamma\sim 1$ and $|\lambda|\sim 0.1$, which agrees with a direct measurement of γ from the spectrum of the controlling noise, and with the value $|\lambda|=0.09$ derived from the current–voltage characteristic in Fig. 2.

Figure 5 shows a plot of the gain as a function of γ. The gain is defined by

$$K(A)=\sqrt{S_V(f_s)\Delta f}/A, \tag{14}$$

where $S_V(f_s)$ is the spectral intensity of the first harmonic of the output signal and $\Delta f=0.03$ Hz is the spectral bandwidth for the measurement of $S_V(f)$.

For $\gamma<f_s$, the adiabaticity condition for the signal relative to the noise is clearly violated, so the gain is reduced. For large γ, according to Eq. (7) the parameter α approaches

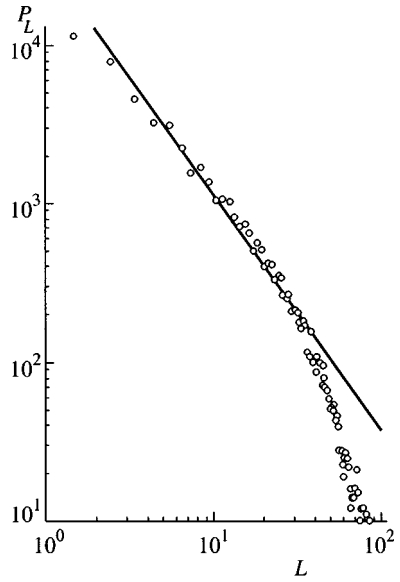


FIG. 4. Distribution of the length of the laminar phase for constant input signal $E(t)=E=15$ mV and laminar threshold of $p=0.1$, for $\gamma=1.14$. The straight line has slope $-3/2$.

unity and the gain also falls off. Thus, it is possible to choose an optimal correlation time for the controlling noise so as to maximize the gain .

3. CONCLUSIONS

In this paper we have shown experimentally that one can observe hypersensitivity to weak ac signals in a system with *on- off* intermittency when a multiplicative noise driver with finite correlation time $1/\gamma$ and sufficiently high intensity Δ is applied to it. Here the signal gain is greatest within an optimal range of noise correlation times.

The author thanks S. L. Ginzburg and M. A. Pustovoit for sustained interest and useful discussions.

This work was supported by the state program ''Physics of Quantum and Wave Processes'' (subprogram ''Statistical Physics,'' project VIII- 3), the state program ''Neutron Studies of Materials,'' and the Russian Fund for Fundamental Research (Project No. 99-02-17545).

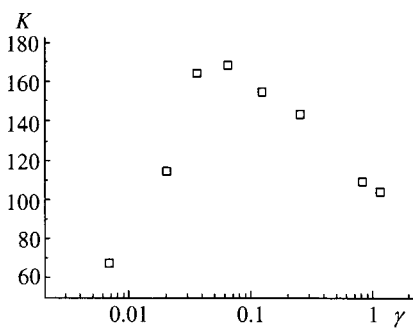


FIG. 5. Gain as a function of γ for a rectangular input signal of amplitude $A=10$ mV and period 3.3 s (dimensionless signal frequency 0.0018).

*)E-mail: gerashch@hep486.pnpi.spb.ru

[1] S. L. Ginzburg and M. A. Pustovoit, Phys. Rev. Lett. **80**, 4840 (1998).
[2] S. L. Ginzburg and M. A. Pustovoit, JETP Lett. **67**, 627 (1998).
[3] O. V. Gerashchenko, S. L. Ginzburg, and M. A. Pustovoit, JETP Lett. **67**, 997 (1998).
[4] N. Platt, E. A. Spiegel, and C. Tresser, Phys. Rev. Lett. **70**, 279 (1993).
[5] J. F. Heagy, N. Platt, and S. M. Hammel, Phys. Rev. E **49**, 1140 (1994).
[6] N. Platt, S. M. Hammel, and J. F. Heagy, Phys. Rev. Lett. **72**, 3498 (1994).

Translated by D. H. McNeill

Hypersensitivity of a nonlinear system with multiplicative colored noise to an external periodic signal

S. L. Ginzburg and M. A. Pustovoit*)

B. P. Konstantinov St. Petersburg Institute of Nuclear Physics, Russian Academy of Sciences, 188350 Gatchina, Leningrad Region, Russia
(Submitted 17 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 1484–1498 (October 1999)

A simple nonlinear stochastic system, an overdamped Kramers oscillator with multiplicative colored noise, is studied analytically and by numerical simulation. It is shown that in the region where *on-off* intermittency occurs, the system becomes hypersensitive to weak external periodic signals. © *1999 American Institute of Physics.* [S1063-7761(99)02610-4]

1. INTRODUCTION

It is now generally recognized that in a number of stochastic systems noise can lead to effects that at first glance defy common sense. One widely known example of such a phenomenon is stochastic resonance, where noise makes a system amplify an external (as a rule, periodic) signal.[1] Another class of such phenomena is known as noise-induced transport, where a stochastic system with an asymmetric potential and colored noise operates as a ratchet, and the flow of matter is driven by noise.[2] These systems call to mind Maxwell's demon. Yet another example is noise-induced phase transitions in systems with multiplicative (parametric) noise.[3] Here, adding noise to a small multiplicative bifurcation parameter leads to the development of a phase state that does not exist in the absence of the noise.

Recently we observed yet another unusual phenomenon that illustrates the constructive role of noise in nonlinear systems. It turned out that a nonlinear system (a Kramers oscillator) with multiplicative white noise acted upon by an ultraweak periodic signal in the *on-off* intermittency regime is capable of amplifying that signal by many orders of magnitude.[4] We referred to this phenomenon as noise-induced hypersensitivity.

For further clarification of the physics of this phenomenon, however, it must be borne in mind that white noise is an abstraction, since any real physical noise process is colored, i.e., has a finite correlation time τ. In a theoretical treatment of the problem in its general form, it is necessary to provide for a limiting transition to white noise, where as τ decreases, the noise amplitude increases as $\tau^{-1/2}$. In fact this condition is essentially never satisfied, and therefore the white noise approximation is inapplicable. In addition, the finite correlation time must be taken into account when that time is comparable to the period of the external signal. Thus, in this paper we carry out a theoretical and simulation study of a Kramers oscillator with multiplicative colored noise from the very outset.

Since it is known that a stochastic differential equation with colored Gaussian noise does not have an exact solution,[3] we carry out a theoretical analysis for dichotomous colored noise, and numerical simulation for Gaussian colored noise as well. The results in these two cases are essentially identical. We determine the dependence of the gain for a periodic signal on the amplitude and correlation time of the colored noise and show that the hypersensitivity phenomenon shows up over a wide range of these parameters. The signal gain increases by many orders of magnitude when both the amplitude and correlation time of the noise increase from zero to their optimum values. Here the dependence of the gain on these parameters is the same for different types of noise, both Gaussian and dichotomous (telegraphic).

This paper is organized as follows: in Sec. 2 we describe our model and derive the basic equations for a one-particle probability density function. In Sec. 3 we briefly review the main results of Ref. 4 for white noise, and determine the one-particle probability density function and its moments for colored dichotomous noise in the adiabatic approximation. In Sec. 4 we present results from a numerical simulation for white, colored dichotomous, and colored Gaussian noise. Here we find the range of system parameters within which hypersensitivity occurs. The conclusions are presented in Sec. 5.

2. BASIC EQUATIONS

We study the stochastic differential equation for an overdamped Kramers oscillator with multiplicative noise in the periodic field of a rectangular signal,

$$\frac{dx}{dt}=\lambda x+z(t)x-Ux^3+AR(t),$$
$$R(t+T)=R(t)=\begin{cases}1, & 0<t\leqslant T/2,\\ -1, & T/2<t\leqslant T.\end{cases} \tag{1}$$

Here λ, U, and A are constant parameters, and $z(t)$ is a random variable with autocorrelation function

$$\langle z(t)z(t')\rangle=\Delta^2 e^{-\gamma|t-t'|}. \tag{2}$$

If $z(t)$ is a Gaussian random function, then it satisfies the Ornstein–Uhlenbeck condition,

$$\frac{dz}{dt} = -\gamma z + \Delta\sqrt{2\gamma}\,\xi(t),$$
$$\langle \xi(t)\xi(t')\rangle = \delta(t-t'), \tag{3}$$

where $\xi(t)$ is white noise.

For dichotomous noise, $z(t)$ can be written in the form

$$z(t) = S(t)\Delta, \quad S(t) = \pm 1, \tag{4}$$

where $S(t)$ is a random quantity that changes sign with probability $\gamma/2$ per unit time. Then $z(t)$ is an discontinuous variable, which must be described in the following way. We introduce the discrete time $t_k = k\Delta t$ and write the mapping

$$S_{k+1} = \frac{1}{2}(\xi_{+,k} - \xi_{-,k}) + \frac{1}{2}(\xi_{+,k} + \xi_{-,k})S_k$$
$$= \begin{cases} \xi_{+,k}S_k, & S_k = 1, \\ \xi_{-,k}S_k, & S_k = -1, \end{cases} \tag{5}$$
$$f(\xi_{\pm,k}) = \frac{\gamma}{2}\delta(\xi_{\pm,k}+1)\Delta t + \left(1 - \frac{\gamma}{2}\Delta t\right)\delta(\xi_{\pm,k}-1).$$

Here $\xi_{+,k}$ and $\xi_{-,k}$ are two random independent quantities whose values are determined by the distribution $f(\xi)$. If the corresponding value if negative, then $S(t)$ changes sign. In the limit $\Delta t \to 0$, we obtain random telegraphic (symmetric dichotomous) noise. Taking

$$\Delta^2 = \gamma\beta^2/2, \quad \gamma \to \infty \tag{6}$$

in Eq. (2), we obtain

$$\langle z(t)z(t')\rangle = \beta^2\delta(t-t'),$$

which means that both Gaussian and dichotomous colored noise transform into Gaussian white noise. Equation (1) is to be interpreted here in the sense of Stratonovich.

In the case of white noise, Eq. (1) determines a one-dimensional Markov process, for which one can write the Fokker–Planck equation

$$\frac{\partial F(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left\{\left[\left(\lambda + \frac{\beta^2}{2}\right)x - Ux^3 + AR(t)\right]F(x,t)\right\}$$
$$+ \frac{\beta^2}{2}\frac{\partial^2}{\partial x^2}[x^2F(x,t)]. \tag{7}$$

For colored noise, $x(t)$ and $z(t)$ determine a two-dimensional Markov process for which there is a two-dimensional Fokker–Planck equation. For Gaussian noise, it takes the form

$$\frac{\partial F(x,z,t)}{\partial t} = -\frac{\partial}{\partial x}\{[(\lambda + z)x - Ux^3 + AR(t)]F(x,z,t)\}$$
$$+ \gamma\frac{\partial}{\partial z}[zF(x,z,t)] + \gamma\Delta^2\frac{\partial^2}{\partial z^2}F(x,z,t). \tag{8}$$

Since the latter equation does not admit of a potential,[5] it has no exact solution.

For dichotomous noise the variable $z(t)$ can take on only two values, so instead of the partial differential equation (8), we obtain two ordinary differential equations,[3]

$$\frac{\partial F(x,\Delta,t)}{\partial t} = -\frac{\partial}{\partial x}\{[(\lambda + \Delta)x - Ux^3 + AR(t)]F(x,\Delta,t)\}$$
$$+ \frac{\gamma}{2}[F(x,-\Delta,t) - F(x,\Delta,t)],$$
$$\frac{\partial F(x,-\Delta,t)}{\partial t} = -\frac{\partial}{\partial x}\{[(\lambda - \Delta)x - Ux^3 + AR(t)]F(x,$$
$$-\Delta,t)\} + \frac{\gamma}{2}[F(x,\Delta,t) - F(x,-\Delta,t)]. \tag{9}$$

3. SOLUTION OF THE EQUATIONS IN THE ADIABATIC APPROXIMATION

We now solve Eqs. (7) and (9) in the adiabatic approximation (the characteristic relaxation time of the system is much less than the period of the signal), and then estimate the switching times for the output signal. This estimate makes it possible to determine whether the adiabatic approximation is appropriate.

We begin with Eq. (7). For small signal amplitudes (henceforth we assume that $A \approx 10^{-10}$) its solution in the adiabatic approximation has the form[4]

$$F(x,t) = C|x|^{\alpha-1}\theta(R(t)x)\exp\left\{-\frac{2AR(t)}{\beta^2 x} - \frac{Ux^2}{\beta^2}\right\},$$
$$\alpha = 2\lambda/\beta^2, \tag{10}$$

where θ is the Heaviside step function. The asymptotic form of the normalizing factor C for small α is

$$C = \begin{cases} \alpha, & \alpha > 0, \quad \zeta \gg 1, \\ [\ln(1/A)]^{-1}, & \zeta \ll 1, \\ |\alpha|A^{|\alpha|}, & \alpha < 0, \quad \zeta \gg 1, \end{cases}$$
$$\zeta = |\alpha|\ln(1/A), \quad |\alpha| \ll 1, \quad U/\beta^2 \sim 1. \tag{11}$$

The solution (10) has two interesting properties. First, $F(x) \propto \theta(x)$ for $R(t) > 0$ and $F(x) \propto \theta(-x)$ for $R(t) < 0$. This means that when the sign of the input signal changes, the output signal also changes sign after some time T_0, which we estimate below. The solution (10) was obtained subject to the condition $T \gg T_0$. Its detailed derivation is given elsewhere.[4] Second, in the case of an ultraweak signal, the distribution $F(x)$ has a scaling form over a wide range of x:

$$F(x) = C|x|^{\alpha-1}\theta(R(t)x), \quad A \ll x \ll x_0 \approx \beta/\sqrt{U}. \tag{12}$$

From Eq. (11) it is then clear that for $\zeta \sim 1$, i.e., for a signal amplitude

$$A_0 = \exp(-1/|\alpha|),$$

crossover occurs, i.e., for $|\alpha| \ll 1$ an ultraweak signal $A_0 \ll A \ll x_0$ changes the density function in a fundamental way. Horsthemke and Lefever[3] only consider $A \ll A_0$; when $\lambda < 0$ the function $F(x)$ approaches $\delta(x)$, i.e., $F(x) \to \delta(x)$. In contrast, here we study the region $A_0 \ll A \ll x_0$ for small absolute values of α. That is precisely the region in which the effects of interest can be detected. To see them immediately,

we calculate the moments of $F(x,t)$. Given the explicit form of $R(t)$ in Eq. (1), in the adiabatic approximation we obtain for small ζ

$$\begin{aligned}
&\langle x(t)\rangle = \frac{\beta}{2}\sqrt{\frac{\pi}{U}}\frac{1}{\ln(1/A)}R(t),\\
&\langle x^2(t)\rangle = \frac{\beta^2}{2U}\frac{1}{\ln(1/A)},\\
&\frac{\langle x(t)\rangle^2}{\langle x^2(t)\rangle} = \frac{\pi}{2\ln(1/A)} \ll 1,\\
&I = \frac{\langle x(t)\rangle}{AR(t)} = \sqrt{\frac{\pi}{4U}}\frac{\beta}{A\ln(1/A)}.
\end{aligned} \tag{13}$$

Here I is the signal gain.

This expression for I holds only in the adiabatic approximation. In general the gain is defined by

$$I^2 = \frac{1}{T}\int_0^T \frac{\langle x(t)\rangle^2}{A^2}dt. \tag{14}$$

Since $\ln(1/A)$ is a very slowly varying function, the first moment is of order x_0, while the gain is enormous. (For $\beta=0.7$, $U=1$, and $A=10^{-11}$, $I=2.5\times10^9$.) Note that $x(t)$ fluctuates strongly, and the standard deviation is large compared to the mean. However, since $\langle x(t)\rangle$ is an ac quantity and $\langle x^2(t)\rangle$ is not, the signal can easily be detected using standard spectral techniques.

We termed the foregoing phenomenon, which is induced by strong multiplicative noise, ithypersensitivity to variable weak signals.

Based on simple physical considerations, we now estimate the signal switching time T_0 when the sign of $R(t)$ change. When the sign of $R(t)$ switches, the trajectory $x(t)$ changes sign only when $|x|$ reaches A. We must therefore determine from Eq. (10) the probability that x lies in the region $|x|\leqslant A$, i.e., that it will be ready to pass through zero. T_0 is obviously inversely proportional to that probability, i.e.,

$$\begin{aligned}
&T_0^{-1} \propto \int_0^A F(x)dx \propto CA^{\alpha},\\
&T_0 \propto \begin{cases} A^{-\alpha}/\alpha, & \alpha>0, \quad \zeta\gg 1,\\ \ln(1/A), & \zeta\ll 1,\\ 1/|\alpha|, & \alpha<0, \quad \zeta\gg 1. \end{cases}
\end{aligned} \tag{15}$$

Clearly, adiabaticity fails here (T_0 becomes large) for positive α. For $\alpha<0$ and $\zeta\gg1$ it is always satisfied, while for $\zeta\ll1$ there is crossover to nonadiabatic behavior. The numerical simulations shown below demonstrate this clearly.

We now proceed to the case of dichotomous noise, i.e., to the solution of Eq. (9) in the adiabatic approximation. It has been shown[3] that for a constant input signal ($R(t)=1$), it is possible to find a stationary solution of Eq. (9). Going to an adiabatically varying signal, we write down the solutions for positive and negative input signals separately. For $R(t)=1$, Eq. (9) yields the following closed set of expressions for the probability density function $F(x)=F(x,\Delta)+F(x,-\Delta)$:[3]

$$\begin{aligned}
F(x) &= C\frac{g(x)}{\Delta^2 g^2(x)-f^2(x)}\\
&\quad\times\exp\left[-\frac{\gamma}{2}\int dz\left(\frac{1}{f(z)-g(z)\Delta}+\frac{1}{f(z)+g(z)\Delta}\right)\right],\\
F(x,\Delta) &= \frac{g(x)\Delta-f(x)}{2g(x)\Delta}F(x),\\
F(x,-\Delta) &= \frac{g(x)\Delta+f(x)}{2g(x)\Delta}F(x),\\
f(x) &= \lambda x - Ux^3 + A, \quad g(x)=x.
\end{aligned} \tag{16}$$

Above all, we note that for the limiting transition to white noise (6) this yields the standard solution of Eq. (7):[3,5]

$$F(x) = \frac{C}{g(x)}\exp\left[\frac{2}{\beta^2}\int\frac{f(x)}{g^2(x)}dx\right]. \tag{17}$$

In addition, there are three constants with the same dimensions in the problem, Δ, λ, and γ. We are interested in small λ, so in the following we take $|\lambda|<\Delta$. We assume no other relationships among these constants *a priori*. Then, after the integrals are evaluated, the solution (16) takes the form

$$\begin{aligned}
F(x) &= Cx\left(x+\frac{A}{\Delta+\lambda}\right)^{-1-\gamma/2(\Delta+\lambda)}\\
&\quad\times\left(x-\frac{A}{\Delta-\lambda}\right)^{-1+\gamma/2(\Delta-\lambda)}\\
&\quad\times\left(\frac{\Delta+\lambda}{U}-x^2\right)^{-1+\gamma/4(\Delta+\lambda)}\\
&\quad\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-1-\gamma/4(\Delta-\lambda)},\\
F(x,\Delta) &= \frac{CU}{2\Delta}\left(x+\frac{A}{\Delta+\lambda}\right)^{-1-\gamma/2(\Delta+\lambda)}\\
&\quad\times\left(x-\frac{A}{\Delta-\lambda}\right)^{\gamma/2(\Delta-\lambda)}\\
&\quad\times\left(\frac{\Delta+\lambda}{U}-x^2\right)^{-1+\gamma/4(\Delta+\lambda)}\\
&\quad\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-\gamma/4(\Delta-\lambda)},\\
F(x,-\Delta) &= \frac{CU}{2\Delta}\left(x+\frac{A}{\Delta+\lambda}\right)^{-\gamma/2(\Delta+\lambda)}\\
&\quad\times\left(x-\frac{A}{\Delta-\lambda}\right)^{-1+\gamma/2(\Delta-\lambda)}\\
&\quad\times\left(\frac{\Delta+\lambda}{U}-x^2\right)^{\gamma/4(\Delta+\lambda)}\\
&\quad\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-1-\gamma/4(\Delta-\lambda)},
\end{aligned} \tag{18}$$

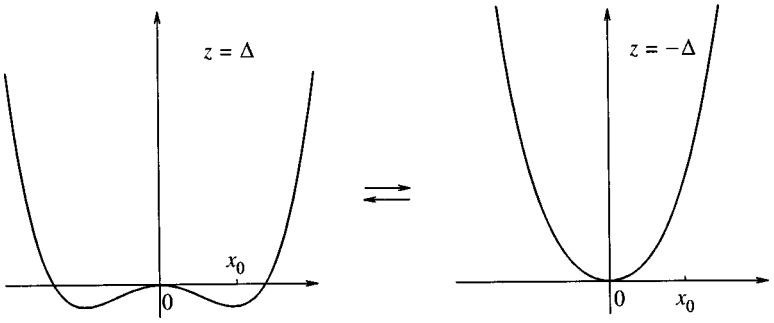


FIG. 1. Potential curves for Eq. (1) with two values of the dichotomous noise in Eq. (4) and $A=0$.

$$R(t)=+1.$$

Following Ref. 3, it can be shown that $F(x)$ and $F(x,\pm\Delta)$ are specified on the interval $(A/(\Delta-\lambda),\sqrt{(\Delta+\lambda)/U})$, which is referred to as the support of $F(x)$.

It is clear from Eq. (18) that for small γ and $z=\Delta$ there is a singularity only at $x=\sqrt{(\Delta+\lambda)/U}$, and when $z=-\Delta$ only at $x=A/(\Delta-\lambda)$, i.e., as $\gamma\to 0$ there is a very strong correlation between x and z, which is perfectly natural.

Then, for a negative input signal we have

$$\begin{aligned}F(x)=C\cdot(-x)&\left(-x+\frac{A}{\Delta+\lambda}\right)^{-1-\gamma/2(\Delta+\lambda)}\\&\times\left(-x-\frac{A}{\Delta-\lambda}\right)^{-1+\gamma/2(\Delta-\lambda)}\\&\times\left(\frac{\Delta+\lambda}{U}-x^2\right)^{-1+\gamma/4(\Delta+\lambda)}\\&\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-1-\gamma/4(\Delta-\lambda)},\\R(t)=-1,\quad x\in&\left(-\sqrt{\frac{\Delta+\lambda}{U}},-\frac{A}{\Delta-\lambda}\right)\end{aligned}\tag{19}$$

and similarly for $F(x,\pm\Delta)$.

Equations (18) and (19), therefore, transform into one another when x goes into $-x$ under reflection, which implies reflection symmetry of the probability density function about zero. Thus, in the following we mainly consider just positive input signals.

We initially take $A=0$ in Eq. (18). Then

$$\begin{aligned}F(x)=Cx^{-1+\alpha}&\left(\frac{\Delta+\lambda}{U}-x^2\right)^{-1+\gamma/4(\Delta+\lambda)}\\&\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-1-\gamma/4(\Delta-\lambda)},\\F(x,\Delta)=\frac{UC}{2\Delta}x^{-1+\alpha}&\left(\frac{\Delta+\lambda}{U}-x^2\right)^{-1+\gamma/4(\Delta+\lambda)}\\&\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-\gamma/4(\Delta-\lambda)},\end{aligned}\tag{20}$$

$$\begin{aligned}F(x,-\Delta)=\frac{UC}{2\Delta}x^{-1+\alpha}&\left(\frac{\Delta+\lambda}{U}-x^2\right)^{\gamma/4(\Delta+\lambda)}\\&\times\left(\frac{\Delta-\lambda}{U}+x^2\right)^{-1-\gamma/4(\Delta-\lambda)},\\\alpha=\frac{\lambda\gamma}{\Delta^2-\lambda^2},\quad x\in&\left(0,\sqrt{\frac{\Delta+\lambda}{U}}\right).\end{aligned}$$

It is clear from Eqs. (18) and (20) that for $\gamma=\gamma_c=4(\Delta+\lambda)$, the character of the probability density function changes. For $\gamma<\gamma_c$ there is a singularity in the functions $F(x)$ and $F(x,\Delta)$ at $x=x_0=\sqrt{(\Delta+\lambda)/U}$ (i.e., the system spends a lot of its time near x_0), and for $\gamma>\gamma_c$ it vanishes. Furthermore, it is clear from Eq. (20) that over a wide range $A\ll x\ll x_0$ the probability density function has the form $F(x)\propto x^{\alpha-1}$, as in the case of white noise, only with an otherwise determined parameter α. (In the white noise limit, these two definitions coincide.) It also follows from Eq. (20) that for all the functions $F(x)$ and $F(x,\pm\Delta)$, there is the same singularity at $x=0$, while for $x=x_0$ the functions $F(x)$ and $F(x,\Delta)$ are singular and $F(x,-\Delta)$ goes to zero. In addition, for $\alpha<0$ the normalizing integral diverges, i.e., the finiteness of the signal amplitude A must be taken into account.

The factor C can be expressed in terms of hypergeometric functions. However, it is clear that since the singularity at $x=0$ is cut off by the boundary of the support of $F(x)$, a reasonable estimate of C can be obtained by simply taking

$$C^{-1}=B^{-1}\int_A^{x_0}x^{\alpha-1}dx=B^{-1}\frac{x_0^\alpha-A^\alpha}{\alpha},$$

where B is a constant independent of A. For small α, we have

$$C=\begin{cases}B\alpha, & \alpha>0,\quad \zeta\gg 1,\\ B/\ln(1/A), & \zeta\ll 1,\\ B|\alpha|A^{|\alpha|}, & \alpha<0,\quad \zeta\gg 1.\end{cases}\tag{21}$$

This expression coincides with the normalization (11). Since the hypersensitivity effect is related to the dependence of C on α, it is clear that introducing colored dichotomous noise does not alter the physics, and only affects the determination of α. Nevertheless, it is interesting that in the case of strongly correlated noise, i.e., when γ decreases, α also decreases, that is, hypersensitivity is induced as well.

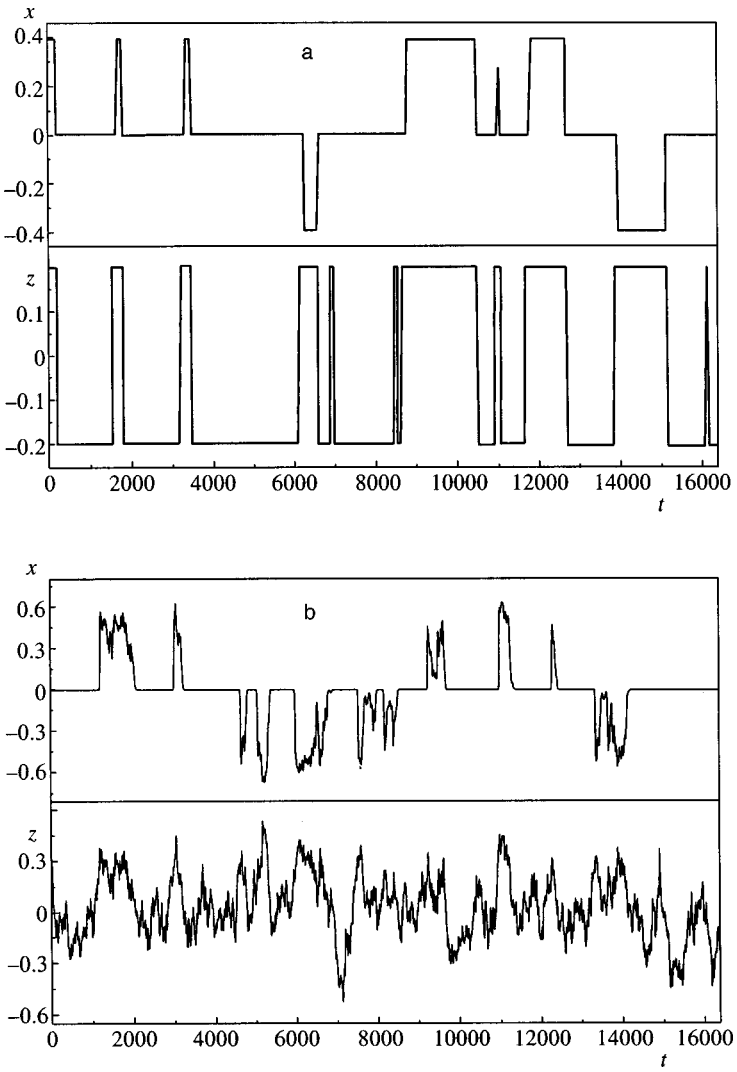


FIG. 2. Time behavior of multiplicative telegraphic (a) and Gaussian (b) noise and the corresponding output signal of a system with parameters $\gamma=0.0005$, $\Delta=0.2$, $\lambda=-0.05$, $A=10^{-11}$, $T=8192$, and $U=1$.

Let us examine the case of small γ separately. As can be seen from Eq. (17), when $\gamma\to 0$ all the exponents go to unity. In order to understand the structure of the response, let us consider the case $A=0$, i.e., Eq. (20). The main factors in these formulas are $x^{\alpha-1}$ and $(x_0-x)^{-1+\gamma/4(\Delta+\lambda)}$. We therefore examine just these, i.e., we write the model probability density function (for positive λ):

$$F(y)=Cy^{\alpha-1}(1-y)^{\varepsilon-1},$$

$$\varepsilon=\frac{\gamma}{4(\Delta+\lambda)}, \quad \alpha=\frac{\lambda\gamma}{\Delta^2-\lambda^2}, \quad y=\frac{x}{x_0}, \tag{22}$$

$$C=\frac{\Gamma(\alpha+\varepsilon)}{\Gamma(\alpha)\Gamma(\varepsilon)} \xrightarrow[\alpha,\varepsilon\to 0]{} \frac{\alpha\varepsilon}{\alpha+\varepsilon}.$$

Averaging an arbitrary function $f(y)$ with respect to the distribution (22), we obtain

$$\langle f(y)\rangle=\frac{\alpha\varepsilon}{\alpha+\varepsilon}\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}\frac{\Gamma(\alpha+n)\Gamma(\varepsilon)}{\Gamma(\alpha+n+\varepsilon)} \xrightarrow[\alpha,\varepsilon\to 0]{} f(0)$$
$$+\frac{\alpha}{\alpha+\varepsilon}\sum_{n=1}^{\infty}\frac{f^{(n)}(0)}{n!}=\frac{\varepsilon}{\alpha+\varepsilon}f(0)$$
$$+\frac{\alpha}{\alpha+\varepsilon}f(1). \tag{23}$$

It is clear from Eq. (23) that as the parameters α and ε approach zero,

$$F(y)\to\frac{\varepsilon}{\alpha+\varepsilon}\delta(y)+\frac{\alpha}{\alpha+\varepsilon}\delta(y-1), \tag{24}$$

i.e., the probability density vanishes in Eq. (22), and therefore in Eq. (19) for $\gamma\to 0$, it consists of two smeared-out delta functions. Physically, this is obvious, since for small γ the system is mainly in the $x=0$ and $x=x_0$ states, to which it relaxes exponentially. However, if we simply examine the adiabatic approximation with respect to the noise correlation time, then, instead of Eq. (24), we obtain $F(x)=[\delta(x)$

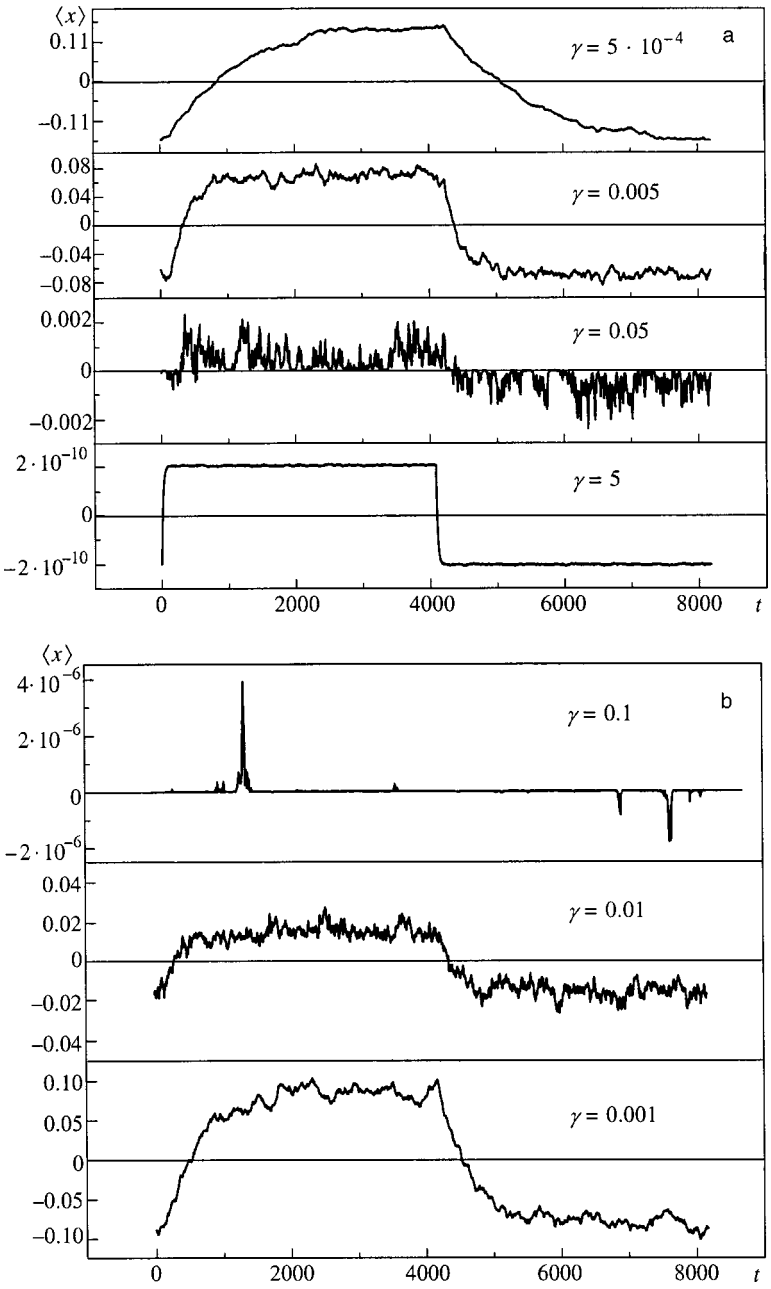


FIG. 3. Averages $\langle x(t)\rangle$ over 500 periods of the signal for different values of γ: (a) telegraph input noise, (b) Gaussian input noise. The remaining parameters are as in Fig. 2.

$+\delta(x-x_0)]/2$. This means that the adiabatic approximation with respect to the correlation time is not correct, but Eq. (24) is.

The reason for this behavior is clear from the following considerations. When the noise switches from $z=\Delta$ to $z=-\Delta$, the system relaxes exponentially to zero from $x=x_0$, and when it reverses direction, relaxation begins at very small x, so that the system is greatly delayed for $|x|\ll 1$. This is clear in Fig. 1, which shows the potential for different levels of noise. For $z=\Delta$ a potential barrier appears at $x=0$, where the system is delayed, and for $z=-\Delta$ the barrier vanishes, so the system rapidly relaxes to the unique potential minimum. The inequivalence of switching also follows from Eqs. (18) and (20) for $F(x,\Delta)$ and $F(x,-\Delta)$. In addition, the smearing out of the delta functions in Eqs. (20) and (22), as well as in Eq. (18), is far from standard.

We now calculate the moments of the distribution in the adiabatic approximation with respect to the period of the signal ($T\gg T_0$). Given that their magnitudes are determined by the region $|x|\gg A$, we can use Eq. (20) (and the analogous formula for $R(t)=-1$). Then, given Eq. (21), by analogy with Eq. (13) for $\zeta\ll 1$ we obtain

$$\langle x(t)\rangle\sim\frac{x_0}{\ln(1/A)}R(t),\quad \langle x^2(t)\rangle\sim\frac{x_0^2}{\ln(1/A)},$$

$$\frac{\langle x(t)\rangle^2}{\langle x^2(t)\rangle}\sim\frac{1}{\ln(1/A)},\quad I=\frac{\langle x(t)\rangle}{AR(t)}\sim\frac{x_0}{A\ln(1/A)} \tag{25}$$

and analogous expressions for the other asymptotes. Comparing Eqs. (25) and (13) we see that colored dichotomous noise also initiates hypersensitivity to a weak ac signal. The significant difference from white noise here is that the condition for hypersensitivity ($|\alpha|\ll 1$) is not only satisfied at

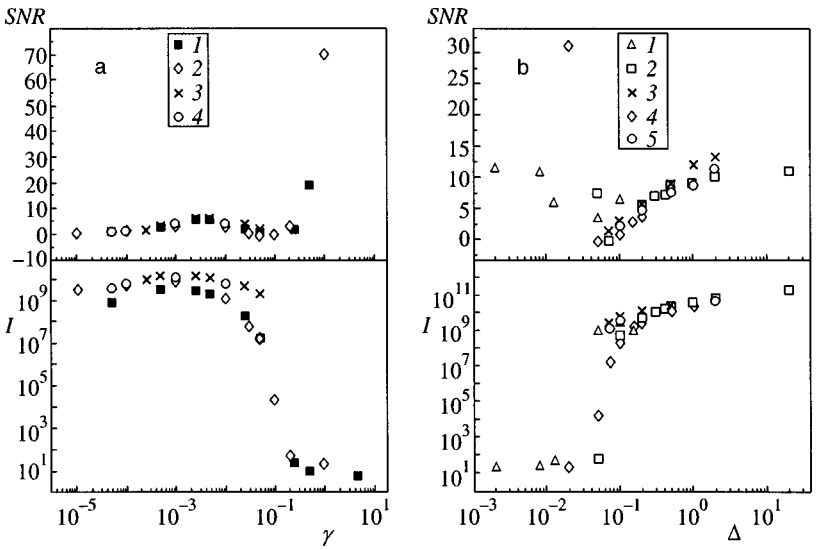


FIG. 4. a) Signal gain and signal-to-noise ratio (*SNR*) as functions of the reciprocal of the correlation time for colored noise. Notation: *1*) telegraphic noise, $\lambda=-0.05$; *2*) Gaussian noise, $\lambda=-0.05$; *3*) telegraphic noise, $\lambda=0.02$; *4*) Gaussian noise, $\lambda=0.02$. The values of Δ, A, and T are the same as in the preceding figures. b) I and *SNR* as functions of the noise amplitude Δ for Gaussian white noise (6) (*1*, $\lambda=-0.05$), for colored telegraphic noise (*2*, $\lambda-0.05$; *3*, $\lambda=0.02$), and for colored Gaussian noise (*4*, $\lambda=-0.05$; *5*, $\lambda=0.02$). For colored noise $\gamma=0.0005$. For white noise, Δ was calculated from its intensity β using Eq. (6) with $\gamma=0.0005$.

small λ, but also at small γ, i.e., for long noise relaxation times τ. This, of course, is true only for $\tau \ll T$. Otherwise, the signal cannot switch and a new class of phenomena (in particular, nonergodicity) occurs, which we shall not discuss.

We conclude this section by considering the coupling between the hypersensitivity phenomenon and *on-off* intermittency. We showed previously[4] that the system of Eqs. (1) with white noise manifests *on-off* intermittency. The phenomenon essentially entails giant fluctuations of the physical quantities, which, with similar probabilities, can take finite values (the burst phase) or become vanishingly small over long quiet periods (the laminar phase). This behavior is associated with the scaling character of the probability density function $F(x)$ for $A \ll x \ll x_0$. Since all the probability density functions for telegraphic noise have the same scaling behavior, it is natural to expect *on-off* intermittency in this case as well. The generally accepted criterion for *on-off* intermittency in a given system is a 3/2 power-law dependence for the distribution of the duration of laminar segments, i.e.,[6]

$$\rho(l) \propto l^{-3/2}. \tag{26}$$

In the case of correlated noise, Eq. (26) can be observed only for $l > \tau = 1/\gamma$.

4. NUMERICAL SIMULATION

We have carried out numerical simulations of process (1) for colored telegraphic [Eqs. (4) and (5)] and colored Gaussian [Eq. (3)] noise. The integration was carried out using an Eulerian scheme with a time step of $\Delta t = 0.01$. Further reductions in the step size have no effect on the system characteristics of interest to us, the input signal transition switching time and the system gain.

The behavior of the system with telegraphic noise is clearly illustrated in Fig. 2a, which shows fragments of the simulated output signal and the corresponding noise input for a long correlation time $\tau = 1/\gamma = 2000$. The output signal pulses rapidly follow the noise pulses and their sign is determined by the sign of the ultraweak periodic actuating signal. The inequivalence in the noise switching described above, when the signal cannot rise to x_0, is also clear. For comparison, we show analogous data for Gaussian colored noise with the same parameters (Fig. 2b).

Subsequently, the hypersensitivity effect shows up distinctly in the ensemble average $\langle x(t) \rangle$ (Fig. 3) and in the magnitude of the gain (Fig. 4). The gain I in Eq. (14) can be defined in terms of the output spectral density $S(\omega)$ as follows. It is known[7] that

$$S(\omega) = 2\pi \sum_k |x_k|^2 \delta(\omega - k\Omega) + S_{\text{noise}}(\omega), \tag{27}$$

where x_k is the Fourier coefficient of the periodic function $\langle x(t) \rangle$ and S_{noise} is the stochastic component. It is clear from Eqs. (14) and (27) that

$$I^2 = \frac{1}{A^2} \sum_k |x_k|^2 \frac{\delta\omega}{2\pi A^2} \sum_i (S_i - S_{\text{noise}}), \tag{28}$$

where S_i is the ith harmonic of the signal in the spectrum and $\delta\omega$ is the spectral bandwidth.

It is clear from Fig. 4 that there are ranges of the amplitude and correlation time of colored noise within which the gain coefficient has a maximum. This happens because the

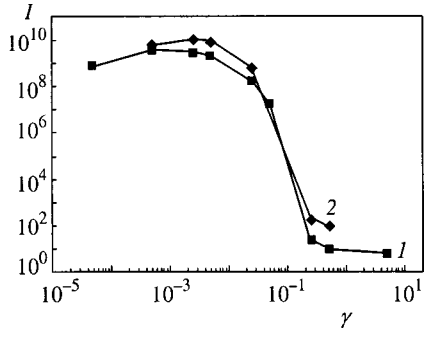


FIG. 5. Resonance plots of $I(\gamma)$ for input signal periods $T = 8192$ (*1*) and $T = 819$ (*2*); $\Delta = 0.2$, $\lambda = -0.05$, $A = 10^{-11}$.

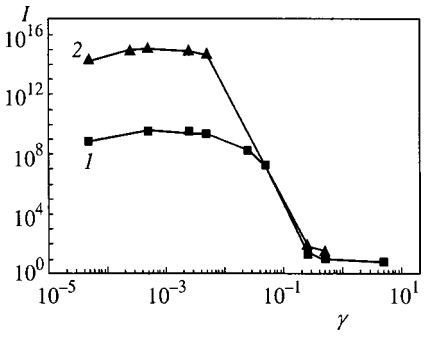


FIG. 6. Resonance plots of $I(\gamma)$ for $U=1$ (*1*) and $U=10^{-10}$ (*2*); $T=8192$ and the remaining parameters are as in Fig. 5.

relaxation time T_0 (15) depends on both Δ and γ. The system has maximum sensitivity to a periodic signal when

$$T_0(\gamma,\Delta)\sim T/2. \tag{29}$$

The condition (29) is similar to the condition for an ordinary stochastic resonance in the classical problem with a two-well potential,[8] which describes a phenomenon of an entirely different character.

The signal-to-noise ratio, defined as

$$SNR^2=\frac{S_i-S_{\text{noise}}}{S_{\text{noise}}}, \tag{30}$$

also has a resonant dependence on the noise, which emphasizes the similarity of the effect we are studying to a classical stochastic resonance. Note that in the region where the sharp rise in gain takes place, this ratio passes through a minimum, which means that there are strong fluctuations in the system (this is also clear in the curves for $\gamma=0.05$ in Fig. 3a and with $\gamma=0.1$ in Fig. 3b). The reasons for this behavior of the system are as yet unclear.

Figure 4 illustrates well the universality of the hypersensitivity phenomenon for different kinds of noise. This is to be expected, since when the condition $|\alpha|\ll 1$ for hypersensitivity is satisfied, the gain depends only on α, as can be seen from the equations of the previous section.

Figure 5 shows resonance plots of I for two signal periods and Fig. 6, for two cutoffs x_0. In our model, $x_0 \propto U^{-1/2}$. It is quite clear that in fact $I\propto x_0/A$; that is, at the peak of the resonance curve the output signal $x(t)$ varies from $x\sim A=10^{-11}$ to $x\sim x_0=10^5$, i.e., by 16 orders of magnitude.

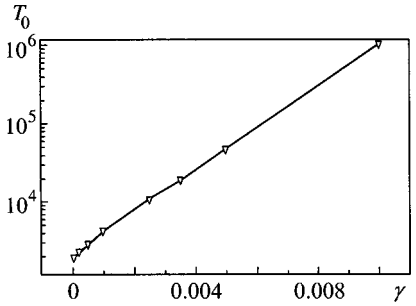


FIG. 7. The time for the system to reach a threshold of $x=10^{50}$ for $U=10^{-100}$ as a function of the reciprocal of the correlation time for telegraph noise. The amplitude of the constant input signal is $AR(t)=10^{-11}$, $\Delta=0.2$, and $\lambda=-0.05$.

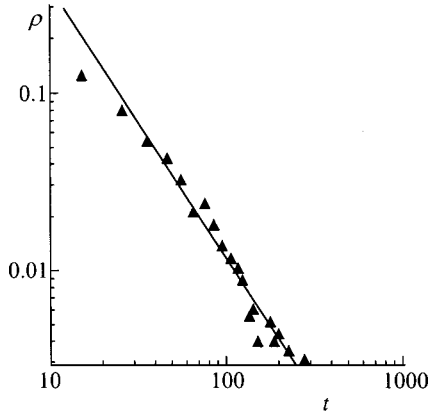


FIG. 8. Probability density function for the length of the laminar segments in the system with telegraph noise, $\gamma=0.025$, $\Delta=0.2$, a constant input signal $AR(t)=10^{-11}$, $\lambda=-0.05$, and $U=1$. The laminar threshold is $x_{\text{thr}}=0.1x_0$.

Figure 7 is a plot of the relaxation time T_0 as a function of γ. As Eq. (15) implies, this dependence is exponential. Although Eq. (15) was derived for white noise, only the scaling form of the probability density function for $|x|<x_0$ was of importance in deriving it, so that Eq. (15) should be universal, as confirmed by the numerical calculations.

In conclusion, we present results indicative of *on-off* intermittency in thise system. The distribution of the lengths of the laminar sections was calculated in a separate computer experiment with a constant signal $(R(t)=1)$ and dichotomous noise. The laminar segments were identified in the calculated output signal in accordance with the criterion $x<x_{\text{thr}}=0.1x_0$. The probability density of the durations calculated in this manner is shown in Fig. 8. Equation (26) does indeed apply for $l>\tau$, i.e., *on-off* intermittency occurs in a system with colored noise as well.

5. CONCLUSIONS

Noise-induced hypersensitivity to weak signals, which is typical of systems with *on-off* intermittency, is a very widespread phenomenon. It turns out that it can develop under the influence of both telegraph and Gaussian correlated (colored) noise, as well as white noise. We have found that the gain for an ultraweak periodic signal in a system with colored multiplicative noise depends resonantly on both the correlation time and amplitude of the noise.

This work was supported by the Russian Fund for Funcamental Research (Project No. 99-02-17545), the state program on the ''Physics of Quantum and Wave Processes'' (subprogram on ''Statistical Physics,'' Project VIII- 3), and the state program on ''Neutron Studies of Matter.''

*)E-mail: markp@hep486.pnpi.spb.ru

[1] K. Wiesenfeld and F. Jaramillo, Chaos **8**, 539 (1998).
[2] R. D. Astumian and F. Moss, Chaos **8**, 533 (1998).
[3] W. Horsthemke and R. Lefever, *Noise-induced Transitions*, Springer-Verlag, Berlin (1984).
[4] S. L. Ginzburg and M. A. Pustovoit, Phys. Rev. Lett. **80**, 4840 (1998).
[5] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer-Verlag, Berlin (1985).
[6] J. F. Heagy, N. Platt, and S. M. Hammel, Phys. Rev. E **49**, 1140 (1994).
[7] P. Jung, Phys. Rep. **234**, 175 (1993).
[8] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).

Translated by D. H. McNeill

ERRATA

ERRATUM: Squeezed states and quantum chaos [JETP 86, 61–70 (January 1998)]

K. N. Alekseev and D. S. Primak

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian, Academy of Sciences, 660036 Krasnoyarsk, Russia
Zh. Éksp. Teor. Fiz. **116**, 1499 (October 1999)

[S1063-7761(99)02710-9]

In Ref. 29 and in Conclusions, ''Rui-Hua Xie'' should appear instead of ''Rui-Pue Xie.'' We thank Xie for calling our attention to this misprint.

Translated by M. E. Alferieff

ERRATUM: Spectral investigations of surface ordering in ultrathin molecular films [JETP 88, 1005–1009 (May 1999)]

V. K. Dolganov, V. M. Zhilin, and K. P. Meletov

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia
Zh. Éksp. Teor. Fiz. **116**, 1499 (October 1999)

[S1063-7761(99)02810-3]

In Fig. 6 the ordinate should be designated as $E_d \cdot 10^3$.

In Fig. 7 the scale divisions along the ordinate should range from 2400 to 3200.

Translated by M. E. Alferieff

