Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 457–467.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 515–527.

ATOMS, MOLECULES, OPTICS

Recovery of Parameters of Delayed-Feedback Systems from Chaotic Time Series

V. I. Ponomarenko, M. D. Prokhorov*, A. S. Karavaev, and B. P. Bezruchko

Institute of Radio Engineering and Electronics, Saratov Branch, Russian Academy of Sciences, Saratov, 410019 Russia

**e-mail: sbire@sgu.ru*

Received August 5, 2004

Abstract—Methods for reconstructing a delay differential equation from the time series of an observable quantity are proposed for various classes of time-delay systems. The methods rely on knowledge of the distributions of extrema of the time series of observed oscillations and projection of the infinite-dimensional phase spaces of time-delay systems onto special low-dimensional subspaces. The effectiveness of the proposed methods is demonstrated by reconstructing delay differential equations from their chaotic solutions, including those corrupted by noise, and by constructing models of real time-delay systems from chaotic time series. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Systems whose behavior is determined not only by their current state, but also by their past states, are widespread in nature [1]. Their models are generally formulated as delay differential equations. Models of this kind have been successfully applied in various areas of physics, biology, physiology, and chemistry. For example, a delay in population dynamics arises because individuals must go through an adolescent period before they contribute to reproduction [2]. In a spatially distributed system, the delay is determined by the time required for a signal to travel a distance with a finite velocity [3]. In particular, the class of delay systems includes the standard Ikeda model of a passive optical cavity [4], the Lang–Kobayashi equations describing the dynamics of semiconductor lasers with optical feedback [5], the Mackey–Glass equation modeling the production of red blood cells [6], and many other models of processes in living organisms ranging from glucose metabolism to infectious disease propagation [7].

A general model of a time-delay system is described by the following equation:

$$\varepsilon_n x^{(n)}(t) + \varepsilon_{n-1} x^{(n-1)}(t) + \ldots + \varepsilon_1 \dot{x}(t) = F(x(t), x(t-\tau_1), \ldots, x(t-\tau_k)), \quad (1)$$

where $x(t)$ is the state of the system at a time t; $x^{(n)}(t)$ is the nth time derivative; $\tau_1, \ldots, \tau_k$ are delays; and the parameters $\varepsilon_1, \ldots, \varepsilon_n$ characterize the system's inertia. Even first-order delay equations may exhibit extremely high-dimensional chaotic oscillations [3, 8]. This complicates reconstruction of a model from time series and necessitates the development of special methods. In most reconstruction methods, the infinite-dimensional phase space of a time-delay system is projected onto low-dimensional subspaces. Criteria for evaluating the quality of reconstruction of a time-delay system include the minimal prediction error [9–11], the minimal value of information entropy [12], and various measures of complexity of the projected time series [13–16]. Some methods for analyzing time-delay systems rely on regression analysis [17–19].

In this paper, we describe an original technique for determining the time delay based on statistical analysis of the time intervals between the extrema of time series. Relying on the technique, we propose new reconstruction methods for high-dimensional time-delay systems characterized by several delays. We also propose methods for reconstructing delay-feedback ring systems from the time series of various observable quantities.

2. CHARACTERISTICS OF TIME SERIES OF TIME-DELAY SYSTEMS

Statistical analysis of the time intervals between the extrema of the time series of various real time-delay systems and their models reveals the following general properties. For a delay-feedback system, the probability N of observing a pair of extrema separated by a time interval τ in a time series, as a function of τ, has a distinct minimum at τ corresponding to the time delay (see Fig. 1a). This can be explained by analyzing one of the most widely used time-delay differential equations,

$$\varepsilon_1 \dot{x}(t) = -x(t) + f(x(t-\tau_1)). \quad (2)$$

In radio engineering, Eq. (2) is used as the general model of a delayed-feedback oscillator represented by a circuit containing a nonlinear device, an inertial element, and a delay line [20] (see Fig. 2). In the realistic case of an inertial system ($\varepsilon_1 > 0$), a time series $x(t)$ can

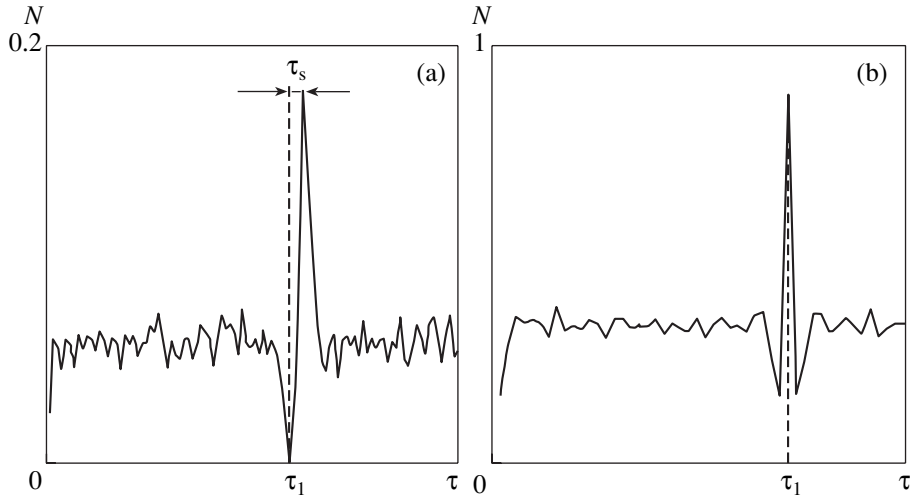

Fig. 1. Qualitative behavior of the number N of pairs of extrema separated by an interval τ, normalized to the total number of extrema in time series: (a) inertial system; (b) zero-inertia system.

be approximated by quadratic functions in the neighborhoods of extremum points; i.e.,

$$\dot{x}(t) = 0, \quad \ddot{x}(t) \neq 0$$

at these points. Indeed, the condition

$$\dot{x}(t) = \ddot{x}(t) = 0$$

holds at inflection and nonquadratic extremum points and on time intervals where the state variable is constant. In this case, the time series virtually has no extrema separated by the delay τ_1. Differentiating (2) with respect to t, we obtain

$$\varepsilon_1 \ddot{x}(t) = -\dot{x}(t) + \frac{df(x(t-\tau_1))}{dx(t-\tau_1)}\dot{x}(t-\tau_1). \qquad (3)$$

If $\dot{x}(t) = 0$ and $\ddot{x}(t) \neq 0$ for $\varepsilon_1 \neq 0$, then it follows from Eq. (3) that $\dot{x}(t-\tau_1) \neq 0$. In other words, there is no extremum separated by τ_1 from a quadratic extremum; i.e., $N(\tau_1) \longrightarrow 0$. When $\tau \neq \tau_1$, the derivatives $\dot{x}(t)$ and $\dot{x}(t-\tau)$ can vanish simultaneously; i.e., extrema separated by τ can exist.

An analogous property is characteristic of time-delay systems having a more general form,

$$\dot{x}(t) = F(x(t), x(t-\tau_1)). \qquad (4)$$

Differentiating Eq. (4) with respect to time, we obtain

$$\ddot{x}(t) = \frac{\partial F(x(t), x(t-\tau_1))}{\partial x(t)}\dot{x}(t) + \frac{\partial F(x(t), x(t-\tau_1))}{\partial x(t-\tau_1)}\dot{x}(t-\tau_1); \qquad (5)$$

i.e., $\dot{x}(t)$ and $\dot{x}(t-\tau_1)$ cannot vanish simultaneously in the typical case of a quadratic extremum.

When $\varepsilon_1 = 0$, differentiation of Eq. (2) with respect to t leads to

$$\dot{x}(t) = \frac{df(x(t-\tau_1))}{dx(t-\tau_1)}\dot{x}(t-\tau_1); \qquad (6)$$

i.e., $\dot{x}(t) = 0$ if $\dot{x}(t-\tau_1) = 0$. In other words, each extremum of the time series of $x(t)$ is separated from another one by τ_1 if $\varepsilon_1 = 0$; i.e., $N(\tau)$ has a maximum at $\tau = \tau_1$ (see Fig. 1b). In the case of a limited-bandwidth system ($\varepsilon_1 > 0$), the most probable time interval between extrema in the time series of $x(t)$ is wider than τ_1 (see Fig. 1a). This can be explained by the effect of the additional delay due to the inertial element in Fig. 2 on the system's dynamics. For such a system, extrema are most frequently separated by $\tau_1 + \tau_s$. In particular, a numerical analysis of Eq. (2) with a nonlinear function

$$f(x) = \lambda - x^2$$

leads to the following estimate when λ is large:

$$\tau_s \approx \varepsilon_1/2.$$

Noisy time series exhibit additional extrema, which

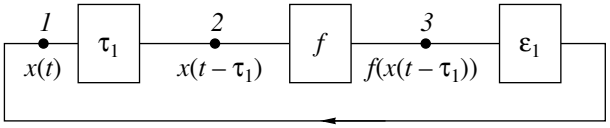

Fig. 2. Block diagram of a delay-feedback ring system. The elements denoted by τ_1, f, and ε_1 represent a delay line, a nonlinear device, and an inertial element, respectively. Points *1–3* are locations where the state variable can be measured.

are not related to the intrinsic dynamics of time-delay systems. The probability of finding a pair of extrema separated by τ increases with the intensity of high-frequency noise, and so does the probability of observing extrema separated by τ_1. At moderate noise levels, the latter probability remains lower than the former; i.e., the qualitative behavior of $N(\tau)$ is preserved.

3. RECONSTRUCTION OF FIRST-ORDER TIME-DELAY SYSTEMS

We describe a method for recovering first-order time-delay systems, using Eq. (2) as an example. To find the delay τ_1, one should locate the extrema of a time series and find the number N of pairs of extrema separated by an interval τ for a sufficiently large set of τ values. The delay τ_1 corresponds to the location of an absolute minimum of $N(\tau)$ adjacent to an absolute maximum.

To recover the parameter ε_1 and the function f from a chaotic time series, we rewrite Eq. (2) as

$$\varepsilon_1 \dot{x}(t) + x(t) = f(x(t-\tau_1)). \quad (7)$$

Then, the graph of f can be recovered as a set of points with coordinates $(x(t-\tau_1), \varepsilon_1 \dot{x}(t) + x(t))$. Since ε_1 is not known a priori, we calculate the sets of points $(x(t-\tau_1), \varepsilon \dot{x}(t) + x(t))$ corresponding to a number of values of ε. Using the fact that $\varepsilon = \varepsilon_1$ only for a set of points that represents a single-valued function, we find the desired ε_1 and nonlinear function by minimizing the length $L(\varepsilon)$ of the polygonal line connecting the points $(x(t-\tau_1), \varepsilon \dot{x}(t) + x(t))$ ordered with respect to $x(t-\tau_1)$. Note that other methods for recovering this function make use of extremum points or points satisfying certain conditions only [13, 14], whereas every data point in a time series can be utilized in the procedure proposed here. Thus, a nonlinear function can be recovered even by using relatively short, weakly chaotic time series.

To demonstrate the effectiveness of this procedure, we have reconstructed various time-delay differential equations having the form of (2), using the time series obtained by solving them numerically. In particular, we have applied our method to time series generated by the Ikeda model of a passive optical cavity [4],

$$\dot{x}(t) = -x(t) + \mu \sin(x(t-\tau_1) - x_0), \quad (8)$$

which corresponds to Eq. (2) with $\varepsilon_1 = 1$. Figure 3a shows a data segment illustrating the dynamics on the high-dimensional chaotic attractor described by Eq. (8) with $\mu = 20$, $\tau_1 = 2$, and $x_0 = \pi/3$ [21]. The full data set of 20000 points, including 1100 extremum points, was obtained at a sampling rate of 200 points per delay interval $\tau_1 = 2$.

Figure 3b shows the graph of $N(\tau)$ obtained by counting the number of points where both $\dot{x}(t)$ and $\dot{x}(t-\tau)$ vanished for values of τ taken with a step of 0.01. The derivative $\dot{x}(t)$ was evaluated by using a local quadratic approximation. The absolute minimum of $N(\tau)$ is located at $\tau = \tau_1 = 2.00$, which corresponds exactly to the time delay. Figure 3c shows the graph of $L(\varepsilon)$ obtained by varying ε with a step of 0.01. The minimum of $L(\varepsilon)$ is located exactly at $\varepsilon = \varepsilon_1 = 1.00$. Figure 3d shows the multimodal function f (virtually identical to the nonlinear function in (8)) recovered by using only 2000 data points. Its amplitude corresponds to the parameter μ, and the its value at $x(t-\tau_1) = 0$ can be used to find x_0.

To evaluate the scope of our method as applied to noisy time series, we added zero-mean Gaussian white noise to the time series generated by Eq. (8). Even at a relative noise amplitude of 20%, we recovered the exact time delay $\tau_1' = 2.00$ and the almost exact value $\varepsilon_1' = 0.98$ of the inertia parameter by using the locations of the absolute minimum of $N(\tau)$ and the minimum of $L(\varepsilon)$, respectively. The corresponding recovered nonlinear function is shown in Fig. 3e. Note that the quality of its reconstruction is much higher than that attained in [19], where the same parameter values were used in the Ikeda model, despite a relatively high noise level and numerical error of ε_1.

In another example, we applied our method to the time series generated by a real delayed-feedback ring system. The elements of its schematic representation by Fig. 2 are a delay line characterized by a delay τ_1, an amplifier characterized by a dynamic transfer function f, and a first-order RC filter with resistance R and capacitance C. The system is described by the equation

$$RC\dot{V}(t) = -V(t) + f(V(t-\tau_1)), \quad (9)$$

where $V(t)$ and $V(t-\tau_1)$ are the delay-line input and output, respectively. Equation (9) is equivalent to Eq. (2) with $\varepsilon_1 = RC$.

Figure 4a shows the input waveform sampled with an analog-to-digital (A/D) converter at a sampling frequency $f_s = 4$ kHz for $\tau_1 = 31.7$ ms and $\varepsilon_1 = 1.007$ ms. Since the delay τ_1 is not a multiple of the sampling interval $T_s = 0.25$ ms, we could not find the exact value of τ_1. The absolute minimum of the $N(\tau)$ curve plotted in Fig. 4b with a step T_s in τ is located at $\tau_1' = 31.75$ ms. The $L(\varepsilon)$ curve plotted in Fig. 4c with a step of 0.025 ms in ε for this τ_1' exhibits a minimum at $\varepsilon_1' = 1.000$ ms. The recovered nonlinear function shown in Fig. 4d is virtually identical to the dynamic transfer function of the amplifier.

4. RECONSTRUCTION OF DELAYED-FEEDBACK RING SYSTEMS

The state variable of the delay-feedback ring system described by Eq. (2) can be measured at its points represented as *1–3* in Fig. 2. However, some of these mea-

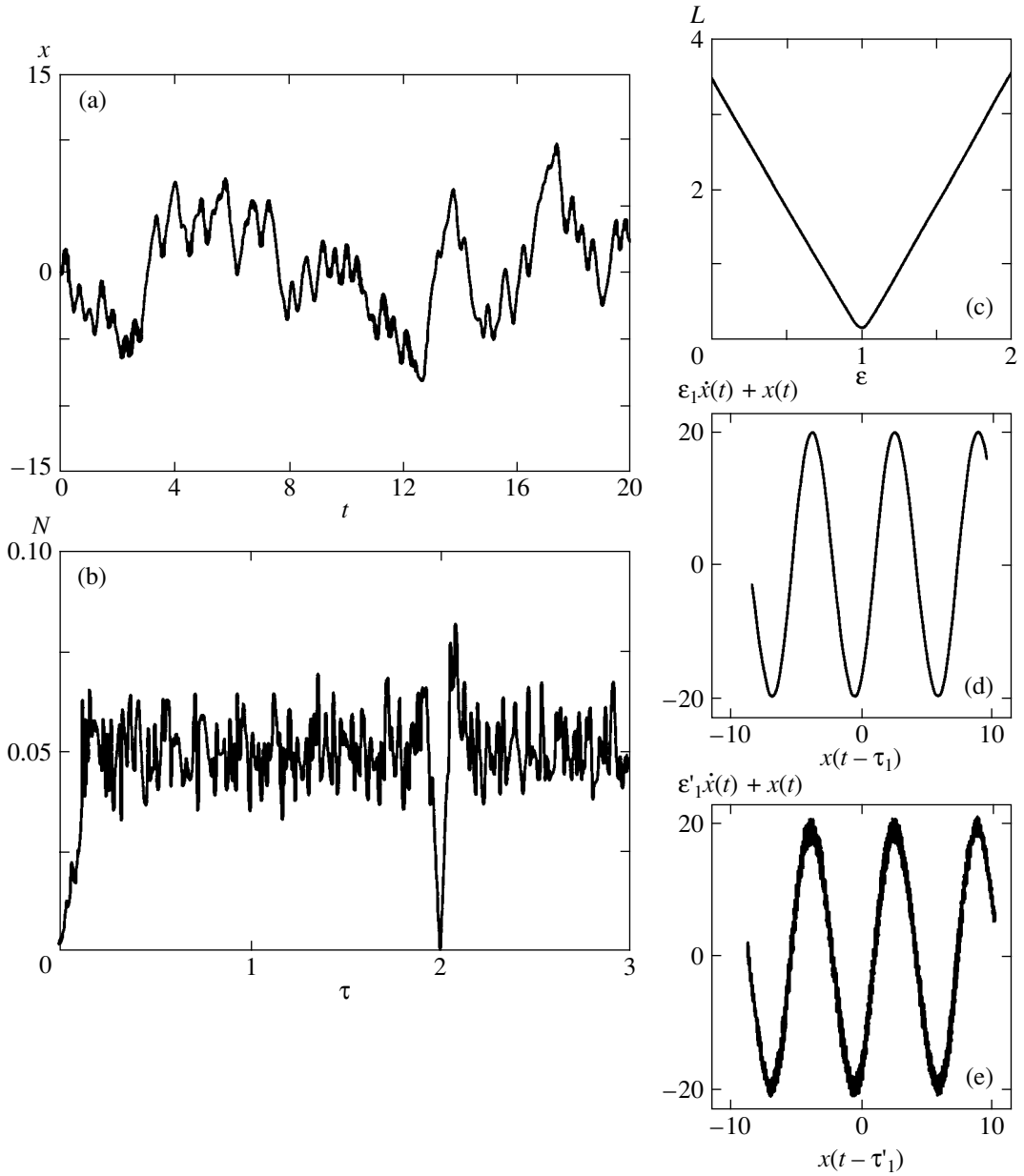

Fig. 3. Reconstruction of the Ikeda model: (a) time series; (b) number N of pairs of extrema separated by an interval τ, normalized to the total number of extrema in the time series; (c) length $L(\varepsilon)$ of the polygonal line connecting the points $(x(t-\tau_1), \varepsilon\dot{x}(t) + x(t))$ ordered with respect to $x(t-\tau_1)$, normalized to the total number of points, $L_{\min}(\varepsilon) = L(1.00)$; (d) recovered nonlinear function; (e) nonlinear function recovered from a time series with a Gaussian white noise level amplitude of 20%.

surements may be impossible to perform, for example, when the system is a monolithic circuit device or when the physical locations of these points cannot be identified. The analysis presented in the preceding section is developed for $x(t)$ measured at point *1*.

When the $x(t-\tau_1)$ is measured at point *2* in Fig. 2, the parameters of system (2) can be evaluated by using the procedure described above, because $x(t-\tau_1)$ is identical to $x(t)$ shifted by τ_1. For example, the results obtained by analyzing the time series sampled as a delay-line output $V(t-\tau_1)$ for the ring oscillator described by Eq. (9) are qualitatively similar to those presented in Fig. 4 to illustrate the reconstruction of the system from $V(t)$.

Now, we consider a procedure for recovering system (2) in the case when $f(x(t-\tau_1))$ is measured at point *3* in Fig. 2. The time series $f(x(t-\tau_1))$ generated by time-delay system (2) has almost no extrema separated by τ_1, because

$$\frac{df(x(t-\tau_1))}{dt} = \left(\frac{df(x(t-\tau_1))}{dx}\right)\dot{x}(t-\tau_1).$$

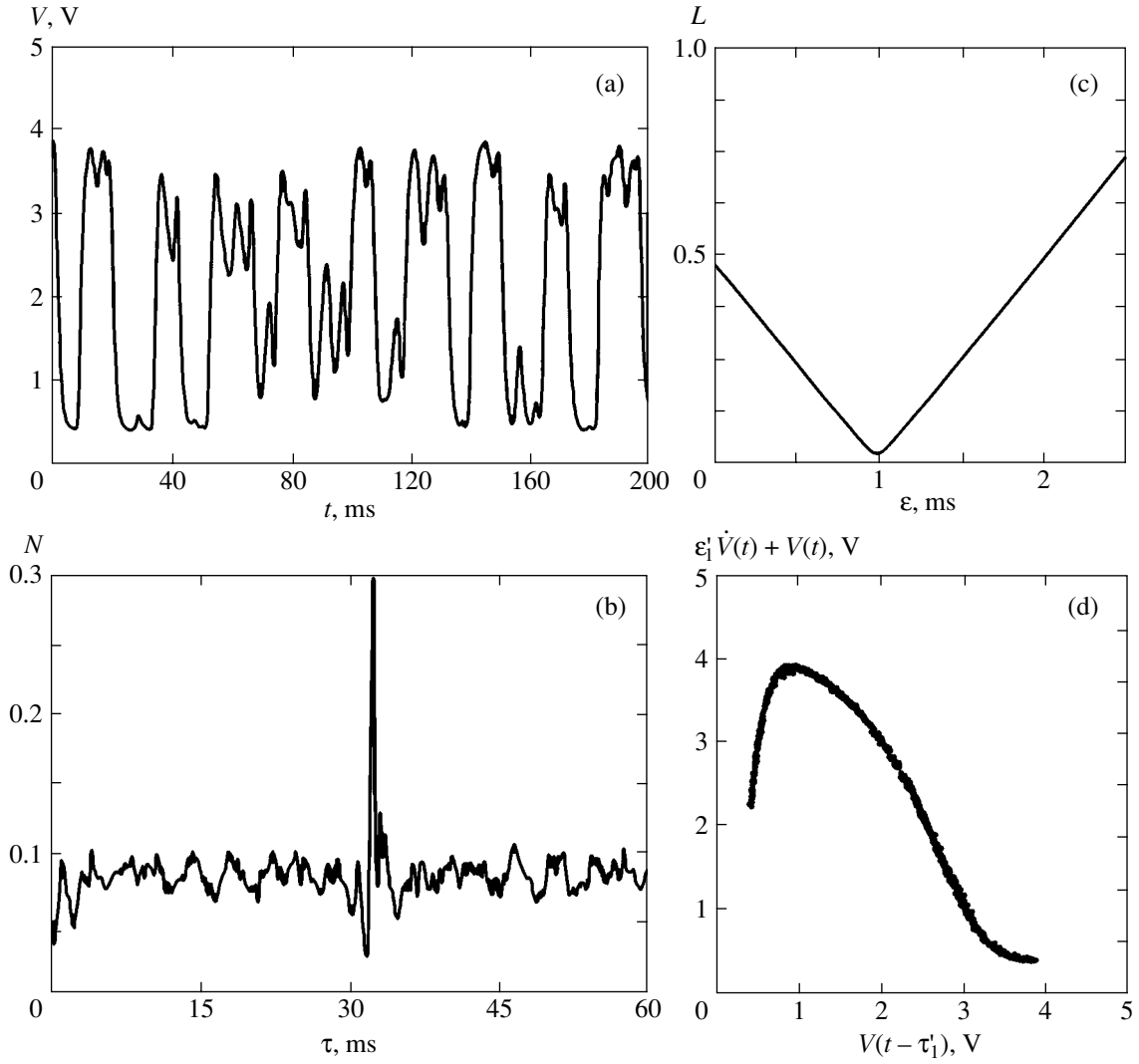

Fig. 4. Reconstruction of a delayed-feedback oscillator: (a) sampled experimental time series; (b) number N of pairs of extrema separated by an interval τ, normalized to the total number of extrema in the time series, $N_{\min}(\tau) = N(31.75\ \text{ms})$; (c) $L(\varepsilon)$ normalized to the total number of points, $L_{\min}(\varepsilon) = L(1.000\ \text{ms})$; (d) recovered nonlinear function.

Therefore, τ_1 can be evaluated by finding the number N of pairs of extrema separated by τ in time series $f(x(t-\tau_1))$ for a sufficiently large set of τ values and locating the absolute minimum of $N(\tau)$.

The function f can be recovered by plotting $f(x(t-\tau_1))$ versus $x(t-\tau_1)$. To find $x(t-\tau_1)$, we use a low-pass filter with cutoff frequency $\nu_1 = 1/\varepsilon_1$ to smooth the available chaotic waveform $f(x(t-\tau_1))$ and shift the filter output by the delay τ_1 determined above. Since neither ε_1 nor ν_1 is known a priori, we calculate the sets of points with coordinates $(u(t-\tau_1), f(x(t-\tau_1)))$ for a number of values of ν, where $u(t-\tau_1)$ is the filter output. Using the fact that $\varepsilon = \varepsilon_1$ only for a set of points that represents a single-valued function, we again find the desired ε_1 and nonlinear function by minimizing the length $L(\varepsilon)$ of the polygonal line connecting the points $(u(t-\tau_1), f(x(t-\tau_1)))$ ordered with respect to the abscissa.

As an illustration, we use a time series $f(x(t-\tau_1))$ to recover the Mackey–Glass equation (a standard model of a first-order delay system) [6],

$$\dot{x}(t) = -bx(t) + \frac{ax(t-\tau_1)}{1 + x^c(t-\tau_1)}, \qquad (10)$$

which takes the form of (2) when divided by b. Our analysis of (10) was performed for

$$a = 0.2, \quad b = 0.1, \quad c = 10, \quad \tau_1 = 300,$$

which correspond to a high-dimensional chaotic attractor [8].

Figure 5a shows a segment of the time series $f(x(t-\tau_1))$ illustrating the dynamics of the Mackey–Glass sys-

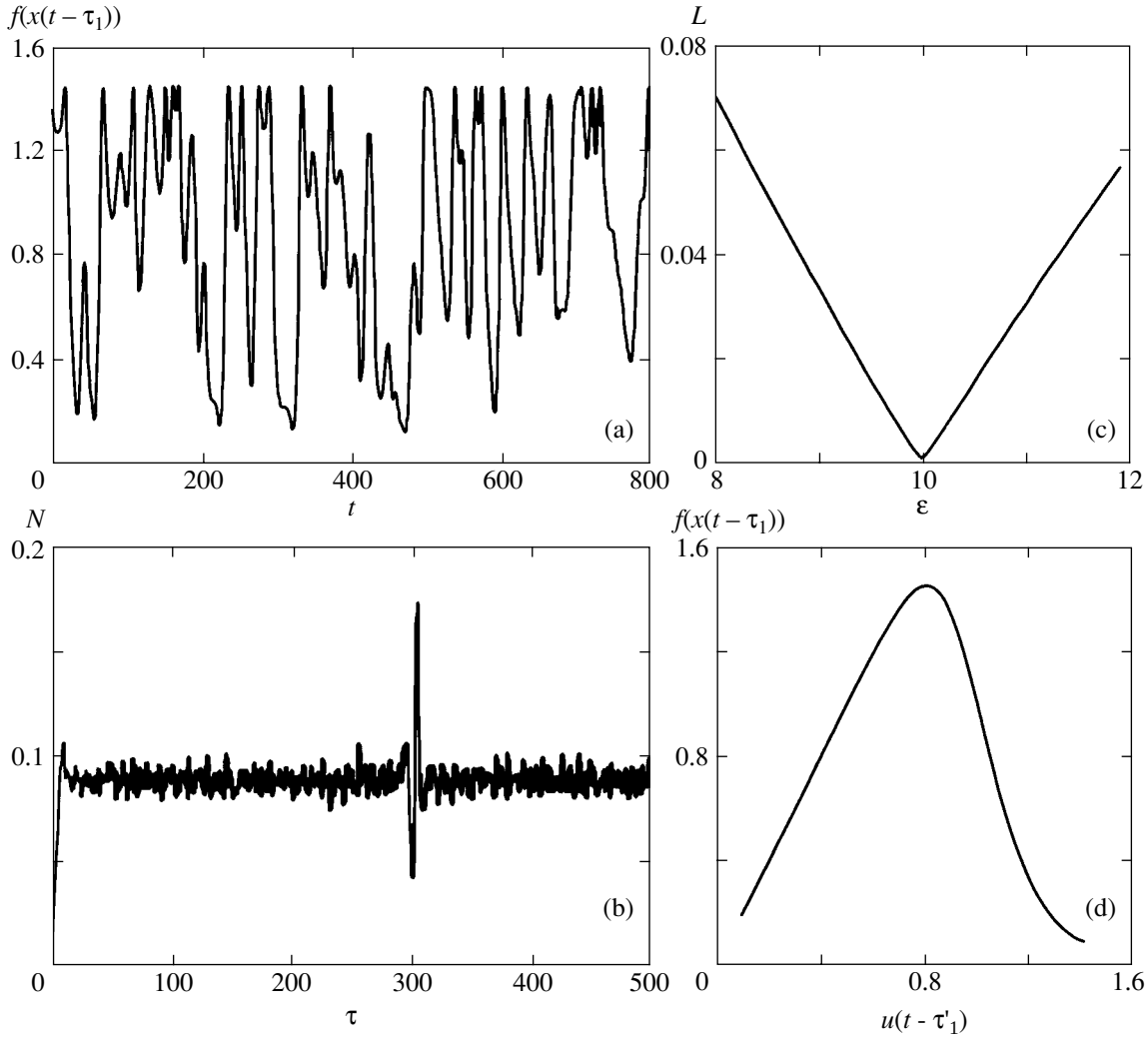

Fig. 5. Reconstruction of the Mackey–Glass system: (a) time series of $f(x(t-\tau_1))$; (b) number N of pairs of extrema in the time series separated by an interval τ, normalized to the total number of extrema, $N_{\min}(\tau) = N(300)$; (c) length $L(\varepsilon)$ of the polygonal line connecting the points with coordinates $(u(t-\tau'_1), f(x(t-\tau_1)))$ ordered with respect to $u(t-\tau'_1)$, normalized to the number of points, $L_{\min}(\varepsilon) = L(10.0)$; (d) recovered nonlinear function.

tem. The absolute minimum of $N(\tau)$ plotted in Fig. 5b with unit step in τ is located at $\tau'_1 = 300$. The minimum of $L(\varepsilon)$ plotted in Fig. 5c with a step of 0.1 is located at $\varepsilon'_1 = 10.0$ ($\varepsilon_1 = 1/b = 10$). Figure 5d shows the function f (virtually identical to the nonlinear function in (10)) recovered by using these values of τ'_1 and ε'_1. The method described above accurately recovers the parameters of the Mackey–Glass system at relative noise amplitudes of up to 10%.

5. RECONSTRUCTION OF HIGH-ORDER DELAYED-FEEDBACK SYSTEMS

The method of statistical analysis of the intervals separating the extrema in time series developed above for systems (2) and (4) can be extended to a higher order time-delay system,

$$\begin{aligned}\varepsilon_n x^{(n)}(t) + \varepsilon_{n-1} x^{(n-1)}(t) + \ldots + \varepsilon_1 \dot{x}(t) \\ = F(x(t), x(t-\tau_1)).\end{aligned} \tag{11}$$

Differentiating (11) with respect to time, we obtain

$$\begin{aligned}\varepsilon_n x^{(n+1)}(t) + \varepsilon_{n-1} x^{n}(t) + \ldots + \varepsilon_1 \ddot{x}(t) \\ = \frac{\partial F(x(t), x(t-\tau_1))}{\partial x(t)} \dot{x}(t) \\ + \frac{\partial F(x(t), x(t-\tau_1))}{\partial x(t-\tau_1)} \dot{x}(t-\tau_1).\end{aligned} \tag{12}$$

If $\dot{x}(t) = 0$, then $\dot{x}(t-\tau_1) \neq 0$ when the left-hand side of Eq. (12) does not vanish. Since the probability that it vanishes is very low in the general case, the graph of $N(\tau)$ must be qualitatively similar to that corresponding to a first-order delay differential equation having the form of (2) or (4).

The proposed method for evaluating the inertia parameter and nonlinear function by projection of the inifinite-dimensional phase space of a time-delay system onto special two-dimensional subspaces can be applied to higher order systems. For example, if the dynamics of a system is governed by the second-order time-delay differential equation

$$\varepsilon_2\ddot{x}(t) + \varepsilon_1\dot{x}(t) = -x(t) + f(x(t-\tau_1)), \qquad (13)$$

then the nonlinear function can be recovered by calculating a set of points with coordinates

$$(x(t-\tau_1), \varepsilon_2\ddot{x}(t) + \varepsilon_1\dot{x}(t) + x(t)).$$

Since neither ε_1 nor ε_2 is known a priori, the graphs of

$$\hat{\varepsilon}_2\ddot{x}(t) + \hat{\varepsilon}_1\dot{x}(t) + x(t)$$

corresponding to a number of values of $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ should be plotted versus $x(t-\tau_1)$. Using the fact that a single-valued function can be obtained only if $\hat{\varepsilon}_1 = \varepsilon_1$ and $\hat{\varepsilon}_2 = \varepsilon_2$, we calculated the minimum length $L(\hat{\varepsilon}_1, \hat{\varepsilon}_2)$ of the polygonal line connecting the points

$$(x(t-\tau_1), \hat{\varepsilon}_2\ddot{x}(t) + \hat{\varepsilon}_1\dot{x}(t) + x(t))$$

ordered with respect to $x(t-\tau_1)$. The quality of reconstruction is poorer than that attained for first-order time-delay systems, because the procedure involves numerical calculation of second derivatives.

This approach is illustrated here by an analysis of a time series generated by a delayed-feedback oscillator containing two series connected low-pass RC filters. The dynamics of the oscillator is governed by Eq. (13), where $x(t)$ and $x(t-\tau_1)$ are the delay-line input and output, respectively;

$$\varepsilon_1 = R_1C_1 + R_2C_2, \quad \varepsilon_2 = R_1C_1R_2C_2,$$

and R_1, C_1 and R_2, C_2 are the respective resistances and capacitances of the filters.

We used an A/D converter to sample the delay-line input at a sampling frequency of $f_s = 4$ kHz for $\tau_1 = 31.7$ ms, $R_1C_1 = 1.007$ ms, and $R_2C_2 = 0.479$ ms ($\varepsilon_1 = 1.486$ ms and $\varepsilon_2 = 0.482$ ms^2). The absolute minimum of the $N(\tau)$ curve plotted in Fig. 6a is located at $\tau_1' = 31.75$ ms. The $L(\hat{\varepsilon}_1, \hat{\varepsilon}_2)$ graph plotted in Fig. 6b with a step of 0.01 ms in $\hat{\varepsilon}_1$ and a step of 0.01 ms^2 in $\hat{\varepsilon}_2$ exhibits a minimum at $\varepsilon_1' = 1.48$ ms and $\varepsilon_2' = 0.48$ ms^2. The filter parameters evaluated by using these results are $(R_1C_1)' = 1.00$ ms and $(R_2C_2)' = 0.48$ ms. The recovered nonlinear function shown in Fig. 6c is virtually identical to the dynamic transfer function of the nonlinear device.

6. RECONSTRUCTION OF HIGH-ORDER DELAYED-FEEDBACK SYSTEMS WITH TWO INDEPENDENT DELAYS

Consider the time-delay system characterized by two independent delays:

$$\dot{x}(t) = F(x(t), x(t-\tau_1), x(t-\tau_2)). \qquad (14)$$

Differentiating (14) with respect to t, we obtain

$$\ddot{x}(t) = \frac{\partial F}{\partial x(t)}\dot{x}(t) + \frac{\partial F}{\partial x(t-\tau_1)}\dot{x}(t-\tau_1) + \frac{\partial F}{\partial x(t-\tau_2)}\dot{x}(t-\tau_2). \qquad (15)$$

Since the time series of Eq. (14) typically has quadratic extrema, $\dot{x}(t) = 0$ and $\ddot{x}(t) \neq 0$ at the extremum points. Therefore, if $\dot{x}(t) = 0$, then

$$a\dot{x}(t-\tau_1) + b\dot{x}(t-\tau_2) \neq 0, \qquad (16)$$

where

$$a = \frac{\partial F(x(t), x(t-\tau_1), x(t-\tau_2))}{\partial x(t-\tau_1)},$$

$$b = \frac{\partial F(x(t), x(t-\tau_1), x(t-\tau_2))}{\partial x(t-\tau_2)}.$$

Condition (16) holds only if $\dot{x}(t-\tau_1) \neq 0$ or $\dot{x}(t-\tau_2) \neq 0$. Therefore, neither $\dot{x}(t)$ and $\dot{x}(t-\tau_1)$ nor $\dot{x}(t)$ and $\dot{x}(t-\tau_2)$ vanish simultaneously; i.e., there is no extremum separated by τ_1 or τ_2 from a quadratic extremum. In other words, the number of extrema separated by intervals τ_1 and τ_2 is much smaller than the number of extrema separated by τ intervals of different length; i.e., the graph of $N(\tau)$ has distinct minima at $\tau = \tau_1$ and $\tau = \tau_2$.

To illustrate the method for recovering the remaining characteristics of a system with two independent delays, we consider the equation

$$\varepsilon_1\dot{x}(t) = -x(t) + f_1(x(t-\tau_1)) + f_2(x(t-\tau_2)). \qquad (17)$$

Differentiating Eq. (17), we obtain

$$\varepsilon_1\ddot{x}(t) = -\dot{x}(t) + \frac{\partial f_1(x(t-\tau_1))}{\partial x(t-\tau_1)}\dot{x}(t-\tau_1) + \frac{\partial f_2(x(t-\tau_2))}{\partial x(t-\tau_2)}\dot{x}(t-\tau_2). \qquad (18)$$

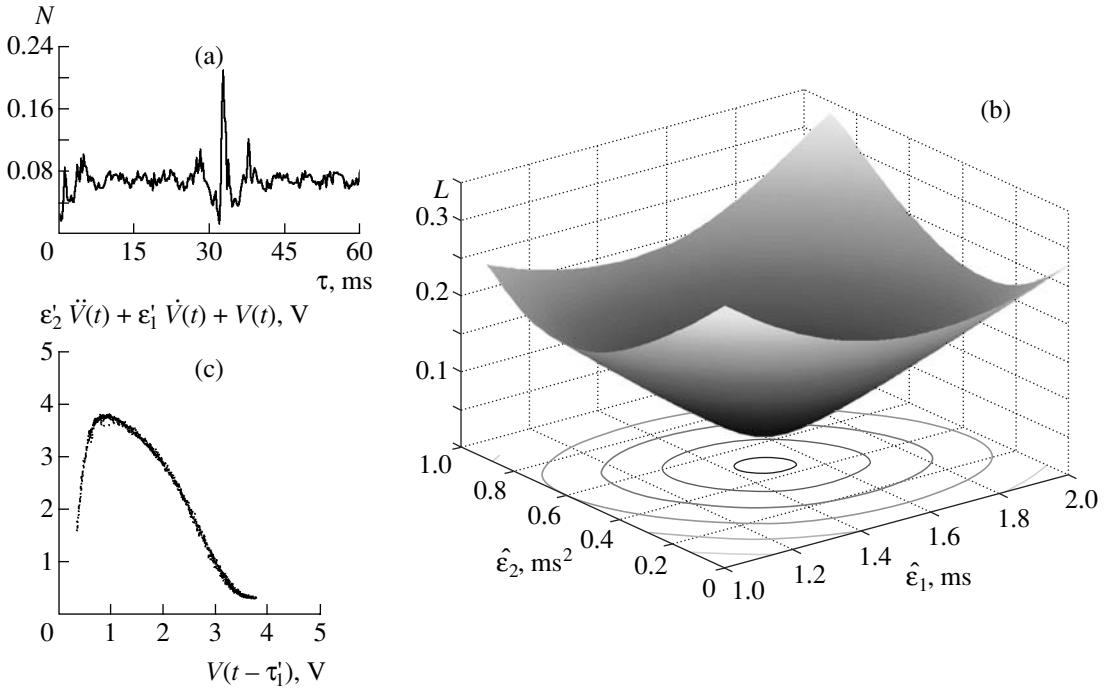

Fig. 6. Reconstruction of the model of a delayed-feedback oscillator with a two-section filter: (a) number N of pairs of extrema in the time series separated by an interval τ, normalized to the total number of extrema, $N_{\min}(\tau) = N(31.75\ \text{ms})$; (b) $L(\hat{\varepsilon}_1, \hat{\varepsilon}_2)$ normalized to the number of points, $L_{\min}(\hat{\varepsilon}_1, \hat{\varepsilon}_2) = L(1.48\ \text{ms}, 0.48\ \text{ms}^2)$; (c) recovered nonlinear function.

By virtue of Eq. (18), if

$$\dot{x}(t-\tau_1) = \dot{x}(t-\tau_2) = 0, \tag{19}$$

then

$$\varepsilon_1 \ddot{x}(t) = -\dot{x}(t),$$

i.e.,

$$\varepsilon_1 = -\frac{\dot{x}(t)}{\ddot{x}(t)}. \tag{20}$$

Thus, the parameter ε_1 can be evaluated by finding the points in the time series of $x(t)$ where condition (19) is satisfied, calculating the first and second derivatives at these points, using (20) to determine the corresponding values of ε_1, and averaging the results.

To recover the nonlinear functions f_1 and f_2, we map the trajectory described by Eq. (17) onto the three-dimensional space

$$(x(t-\tau_1), x(t-\tau_2), \varepsilon_1\dot{x}(t) + x(t)).$$

The trajectory of the system in this space is restricted to a two-dimensional surface, because Eq. (17) can be rewritten as

$$\varepsilon_1\dot{x}(t) + x(t) = f_1(x(t-\tau_1)) + f_2(x(t-\tau_2)). \tag{21}$$

The intersection of the surface with a plane $x(t-\tau_2) = \text{const}$ determines the function f_1 up to an additive constant, because each intersection point satisfies the equation

$$\varepsilon_1\dot{x}(t) + x(t) = f_1(x(t-\tau_1)) + c_1,$$

where

$$c_1 = f_2(x(t-\tau_2))$$

corresponds to the chosen particular constant value of $x(t-\tau_2)$. Similarly, the function f_2 can be recovered by finding the intersection with $x(t-\tau_1) = \text{const}$ as

$$\varepsilon_1\dot{x}(t) + x(t) = f_2(x(t-\tau_2)) + c_2,$$

where

$$c_2 = f_1(x(t-\tau_1))$$

corresponds to a particular value of $x(t-\tau_1)$.

To demonstrate the effectiveness of this method, we recover the generalized Mackey–Glass equation containing an additional delay term,

$$\dot{x}(t) = -bx(t) + \frac{a_1 x(t-\tau_1)}{2(1+x^c(t-\tau_1))} + \frac{a_2 x(t-\tau_2)}{2(1+x^c(t-\tau_2))} \tag{22}$$

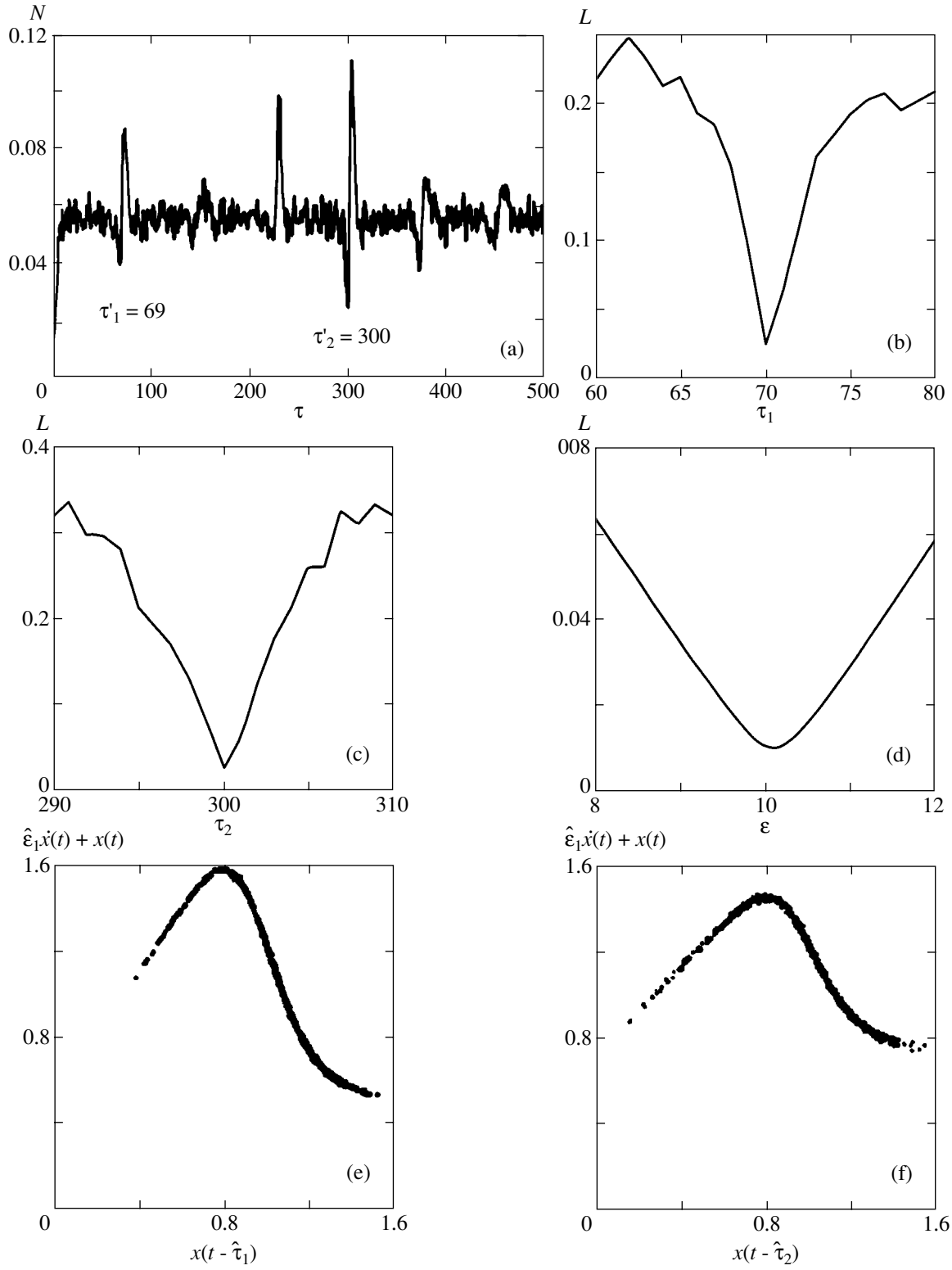

Fig. 7. Reconstruction of Eq. (22): (a) number N of pairs of extrema separated by an interval τ normalized to the total number of extrema in the time series; (b) length L of the polygonal line connecting the intersection points in $x(t - \tau_2') = 1$, ordered with respect to the abscissa, $L_{\min}(\tau_1) = L(70)$; (c) length L of the polygonal line connecting the intersection points in $x(t - \hat{\tau}_1) = 1$, ordered with respect to the abscissa, $L_{\min}(\tau_2) = L(300)$; (d) length L of the polygonal line connecting the intersection points in $x(t - \hat{\tau}_2) = 1$, ordered with respect to the abscissa, $L_{\min}(\varepsilon) = L(10.1)$; (e) nonlinear function f_1 recovered up to $\hat{c}_1 = f_2(x(t - \hat{\tau}_2))$, where $x(t - \hat{\tau}_2) = 1$; (f) nonlinear function f_2 recovered up to $\hat{c}_2 = f_1(x(t - \hat{\tau}_1))$, where $x(t - \hat{\tau}_1) = 1$.

from a corresponding time series. Dividing Eq. (22) by b, we rewrite it as (17) with $\varepsilon_1 = 1/b$. Figure 7a shows the graph of $N(\tau)$ plotted for $a_1 = 0.3$, $a_2 = 0.2$, $b = 0.1$, $c = 10$, $\tau_1 = 70$, and $\tau_2 = 300$. Locating the first pronounced minima of $N(\tau)$, we obtain the delays $\tau_1' = 69$ and $\tau_2' = 300$. Another distinct minimum of $N(\tau)$ is located at a point close to $\tau = \tau_1 + \tau_2$. Using these values of τ_1' and τ_2', we obtain the average $\varepsilon_1' = 9.4$ ($\varepsilon_1 = 1/b = 10$).

To minimize the error of ε_1 given by (20), we discarded the points where $\ddot{x}(t)$ is very small.

Mapping the time series of Eq. (22) onto the three-dimensional space

$$(x(t-\tau_1'), x(t-\tau_2'), \varepsilon_1'\dot{x}(t) + x(t))$$

and calculating its intersections with planes $x(t-\tau_2') = \text{const}$ and $x(t-\tau_1') = \text{const}$, we recovered the nonlinear functions f_1 and f_2 up to additive constants. However, the quality of reconstruction is poor because of the errors of τ_1 and ε_1.

To improve the quality of reconstruction, we refined the desired parameter values as follows. Varying τ_1 in a small neighborhood of $\tau_1' = 69$, we mapped the time series onto a number of embedding spaces,

$$(x(t-\tau_1), x(t-\tau_2'), \varepsilon_1'\dot{x}(t) + x(t)),$$

and calculated their intersections with a plane $x(t-\tau_2') = \text{const}$ in order to find a set of intersection points localized in a narrow strip. The corresponding nearly single-valued function was determined by minimizing the length $L(\tau_1)$ of the polygonal line connecting the intersection points ordered with respect to the abscissa. The $L(\tau_1)$ curve plotted in Fig. 7b has a minimum at $\hat{\tau}_1 = 70$. The value of τ_2 was refined in a similar manner. We mapped the time series onto the spaces

$$(x(t-\hat{\tau}_1), x(t-\tau_2), \varepsilon_1'\dot{x}(t) + x(t))$$

corresponding to a number of τ_2 values in a small neighborhood of $\tau_2' = 300$ and calculated their intersections with $x(t-\hat{\tau}_1) = \text{const}$. Note that we used the refined delay $\hat{\tau}_1 = 70$ in these calculations. The $L(\tau_2)$ curve plotted in Fig. 7c has a minimum at $\hat{\tau}_2 = 300$. If $\hat{\tau}_2 \neq \tau_2'$, then τ_1 should be refined again by calculating the intersections of the embedding spaces with a plane $x(t-\hat{\tau}_2) = \text{const}$ for the refined value of $\hat{\tau}_2$. The parameters τ_1 and τ_2 should be successively refined until constant values are obtained. When the deviations of the starting approximate delays τ_1' and τ_2' from the respective true values are small, both delays can be determined to high accuracy.

After the delays have been refined, a new estimate for $\hat{\varepsilon}_1$ can be found by using (20) as described above. However, a more reliable estimate is obtained by varying ε in a neighborhood of ε_1', mapping the time series onto the spaces

$$(x(t-\hat{\tau}_1), x(t-\hat{\tau}_2), \varepsilon\dot{x}(t) + x(t)),$$

and using all points of intersection with either $x(t-\hat{\tau}_1) = \text{const}$ or $x(t-\hat{\tau}_2) = \text{const}$ to find a nearly single-valued function. In Fig. 7d, the $L(\varepsilon)$ normalized, by analogy with $L(\tau_1)$ and $L(\tau_2)$, to the number of points in the intersection has a minimum at $\hat{\varepsilon}_1 = 10.1$. Note that the proposed procedure of parameter refinement is faster by several orders of magnitude as compared to the simultaneous direct search through the parameters ε_1, τ_1, and τ_2 defining the embedding space

$$(x(t-\tau_1), x(t-\tau_2), \varepsilon_1\dot{x}(t) + x(t)).$$

Figures 7e and 7f show the results obtained by using the refined parameter values $\hat{\varepsilon}_1 = 10.1$, $\hat{\tau}_1 = 70$, and $\hat{\tau}_2 = 300$ to recover the nonlinear functions in (22). To evaluate the scope of the method, we applied it to noisy data sets and found that the system can be recovered to good accuracy at relative noise amplitudes of up to 10%.

7. CONCLUSIONS

We have proposed methods for recovering time-delay systems of certain classes from chaotic time series, based on statistical analysis of the time intervals separating the extrema of the series and mapping the infinite-dimensional phase spaces of the systems onto special low-dimensional subspaces. If dynamical systems describing processes of different nature are similar in structure to certain model equations, then the corresponding time delays, inertia parameters, and nonlinear functions can be recovered even at relatively high noise levels. Since the time delay is determined by performing only comparison and addition operations (without data sorting, estimating approximation error, or calculating any motion-complexity measure), the computational cost of this procedure is relatively low. Since the inertia parameter and nonlinear function for systems with a single time delay are recovered by using all data points in the time series, the procedure can be successfully applied even to relatively short, weakly chaotic time series.

It is shown that the proposed methods can be used to recover the model equations describing delay-feedback ring systems from the time series of various observable state variables. The effectiveness of the methods has been demonstrated by applying them to time series

obtained in numerical experiments, including those corrupted by noise, and to the time series generated by a real device used in radio engineering.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-17593; CRDF, Award REC-006; and INTAS, project no. 03-55-920.

REFERENCES

1. J. K. Hale and S. M. V. Lunel, *Introduction to Functional Differential Equations* (Springer, New York, 1993).
2. Y. Kuang, *Delay Differential Equations with Applications in Population Dynamics* (Academic, Boston, 1993).
3. S. P. Kuznetsov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **25**, 1410 (1982).
4. K. Ikeda, Opt. Commun. **30**, 257 (1979).
5. R. Lang and K. Kobayashi, IEEE J. Quantum Electron. **16**, 347 (1980).
6. M. C. Mackey and L. Glass, Science **197**, 287 (1977).
7. G. A. Bocharov and F. A. Rihan, J. Comput. Appl. Math. **125**, 183 (2000).
8. J. D. Farmer, Physica D (Amsterdam) **4**, 366 (1982).
9. R. Hegger, M. J. Bünner, H. Kantz, and A. Giaquinta, Phys. Rev. Lett. **81**, 558 (1998).
10. C. Zhou and C.-H. Lai, Phys. Rev. E **60**, 320 (1999).
11. V. S. Udaltsov, J.-P. Goedgebuer, L. Larger, *et al.*, Phys. Lett. A **308**, 54 (2003).
12. Y.-C. Tian and F. Gao, Physica D (Amsterdam) **108**, 113 (1997).
13. M. J. Bünner, M. Popp, Th. Meyer, *et al.*, Phys. Lett. A **211**, 345 (1996).
14. M. J. Bünner, M. Popp, Th. Meyer, *et al.*, Phys. Rev. E **54**, 3082 (1996).
15. M. J. Bünner, Th. Meyer, A. Kittel, and J. Parisi, Phys. Rev. E **56**, 5083 (1997).
16. M. J. Bünner, M. Ciofini, A. Giaquinta, *et al.*, Eur. Phys. J. D **10**, 165 (2000).
17. H. Voss and J. Kurths, Phys. Lett. A **234**, 336 (1997).
18. S. P. Ellner, B. E. Kendall, S. N. Wood, *et al.*, Physica D (Amsterdam) **110**, 182 (1997).
19. H. Voss and J. Kurths, Chaos, Solitons and Fractals **10**, 805 (1999).
20. A. S. Dmitriev and V. Ya. Kislov, *Stochastic Oscillations in Radiophysics and Electronics* (Nauka, Moscow, 1989) [in Russian].
21. K. Ikeda and K. Matsumoto, Physica D (Amsterdam) **29**, 223 (1987).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 468–474.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 528–535.

ATOMS, MOLECULES, OPTICS

Diffraction Radiation from a Fast Particle at a Resonance Frequency

M. I. Ryazanov

Moscow Institute of Engineering Physics (State University), Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: ryazanov@theor.mephi.msk.ru

Received August 19, 2004

Abstract—Diffraction radiation generated by a fast particle moving past a dielectric wedge or past a homogeneous medium with the surface of arbitrary shape is considered at a frequency close to the natural frequency of the medium. The angular distribution of diffraction radiation is investigated as a function of the wedge angle and the shape of the surface. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

From the microscopic point of view, diffraction radiation results from the scattering of the field of a uniformly moving charged particle by the atoms of a medium. The scattering cross section of an electromagnetic wave by an atom is maximal near the resonance, when the wave frequency is close to the natural frequency of an atom; therefore, the intensity of diffraction radiation must increase at resonance frequencies. It is well known that the radiation field in the near-field zone is mainly longitudinal and the magnitude of the transverse field is much less than that of the longitudinal field. Therefore, in dense media, the energy transfer from an excited atom to a nonexited one mainly occurs via the longitudinal field through a resonance dipole–dipole interaction rather than through the radiation and absorption of transverse resonance waves. The energy of electron excitation propagates through the medium in the form of excitons. As a result, the interaction between a resonance photon and an atom leads to the annihilation of the photon and the creation of an exciton with the maximum probability. Therefore, the resonance photon, i.e., a photon with energy close to the energy of an exciton, does not penetrate deep into the medium. For the same reason, transverse resonance waves cannot be emitted by atoms situated in the bulk of a dense medium.

Thus, the probability of diffraction radiation at a resonance frequency due to scattering by an atom is much less than the probability of creation of an exciton, and the diffraction radiation results from the scattering of the field of a particle by atoms in a layer close to the surface of the medium. The thickness of this layer is determined by the absorption coefficient of transverse resonance waves. Since this thickness is small, we can restrict the analysis to the single-scattering approximation for the resonance component of the field of a fast particle by the atoms of the medium. Such an approximation in the problem of reflection of resonance electromagnetic waves from the surface of a medium was proposed by Fermi [1]. This approximation made it possible to solve the reflection problem without resorting to the usual macroscopic boundary conditions and showed good agreement with experiment.

It should be noted that the field of a charged particle moving at a constant velocity $\mathbf{v}$ and energy $E \equiv \gamma mc^2$ falls off with distance a as $\exp(-a\omega/\gamma v)$ in the direction perpendicular to the velocity. Therefore, polarization currents that are responsible for the diffraction radiation exponentially decay away from the surface. This fact increases the contribution of the atoms that are situated close to the surface of the medium to the resonance radiation, thus increasing the accuracy of single-scattering approximation for the resonance field component of a fast particle by the atoms of the medium.

Usually, to consider the diffraction radiation, one applies the methods of macroscopic electrodynamics, which take into account the effect of boundaries by ordinary boundary conditions [2–8]. These methods are convenient when applied to the problems of diffraction radiation from surfaces of relatively simple shape. However, for media with a surface of complex shape, rigorous consideration of boundary conditions involves considerable difficulties; therefore, one has to develop special approximation methods for each specific shape of the surface.

A generalization of the method, developed in [1], for analyzing resonance scattering of waves to the problem of diffraction radiation at resonance frequencies provides a relatively simple technique for studying diffraction radiation in the case of surfaces of complex shape. It is worthwhile to apply the Fermi method to the diffraction radiation at a resonance frequency of the medium and to evaluate the intensity of such radiation depending on the shape of the surface of the medium.

2. DIFFRACTION RADIATION FROM A WEDGE AT A RESONANCE FREQUENCY

Consider a charged particle that moves uniformly in vacuum by the law

$$\mathbf{r} = \mathbf{a} + \mathbf{v}t$$

in the plane $x = a$. Let the axis z be directed along the vector $\mathbf{v}$, and the axis x, along $\mathbf{a}$. Then, we can represent the field generated by a moving particle as

$$\begin{aligned}\mathbf{E}_0(\mathbf{r}, t) &= \int d^3p \int d\omega \mathbf{E}_0(\mathbf{p}, \omega)\exp(i\mathbf{p}\cdot\mathbf{r} - i\omega t),\\ \mathbf{E}_0(\mathbf{p}, \omega) &= \mathbf{E}_0(\mathbf{p})\delta(\omega - p_z v),\end{aligned} \tag{1}$$

$$\mathbf{E}_0(\mathbf{p}) = \frac{-ie\{\omega\mathbf{v} - \mathbf{p}c^2\}}{2\pi^2\{p^2c^2 - \omega^2\}}\exp(-ip_x a). \tag{2}$$

Below, we consider the Fourier component of the particle field at a resonance frequency of the medium. Assuming that the wavelength of the field is much greater than the size of a molecule, we can consider the interaction of a molecule with the field in the dipole approximation. The molecules excited by the field of a fast particle give rise to a radiation field; however, because transverse resonance waves are absorbed by the medium, only the resonance component of the field emitted by the molecules that are close to the surface may be irradiated from the medium. Therefore, one can restrict the analysis to the single-scattering approximation of the particle field in the medium. Consider the diffraction radiation of a charged particle when it travels past a homogeneous dielectric wedge given by the formula

$$x = \begin{cases}\xi z, & z < 0,\\ -\eta z, & z > 0\end{cases} \tag{3}$$

($\xi, \eta > 0$), when the concentration of molecules in the medium is n_0. The amplitude of the transverse resonance wave emitted by a molecule situated at the point $\mathbf{R}_a$ exponentially decays as it propagates through the medium. The direction to the observation point is defined by the unit vector $\mathbf{n}$; this direction intersects the surface of the wedge at the point

$$\mathbf{R}_a + \mathbf{R}'(X', Y', Z'),$$

where

$$X'/Z' = n_x/n_z, \quad Y'/Z' = n_y/n_z.$$

The field is irradiated from either face of the wedge depending on the direction of irradiation, i.e., depending on the sign of n_z. Consider, for definiteness, the case of $n_z > 0$, i.e., the forward radiation. Then, the field is irradiated through the plane $x = -\eta z$ and

$$Z' = -\frac{X_a + \eta Z_a}{\eta + n_x/n_z}, \quad \frac{n_x}{n_z} < \eta. \tag{4}$$

When $n_z < 0$, the field is irradiated through the plane $x = \xi z$ of the wedge and

$$Z' = \frac{X_a - \xi Z_a}{\xi - n_x/n_z}, \quad \frac{n_x}{|n_z|} < \xi. \tag{5}$$

The distance between a radiating molecule and the intersection point with the surface is $|Z'/n_z|$. A decrease in the amplitude of the resonance wave in the medium can be taken into account by multiplying the density of intramolecular current by

$$\exp\{-g|Z'(\mathbf{R}_a)/2n_z|\},$$

where g is the absorption coefficient. Hence, to take into account the absorption of the transverse resonance wave in the material of the wedge, we should replace the usual expression for the Fourier image of the density of polarization current in the single-scattering approximation,

$$\begin{aligned}\mathbf{j}(\mathbf{q}, \omega) &= -i\omega(1/2\pi)^3\alpha(\omega)\\ &\times\sum_\alpha \int d^3p \mathbf{E}_0(\mathbf{p}, \omega)\exp\{-i(\mathbf{p} - \mathbf{q})\cdot\mathbf{R}_a\}\end{aligned} \tag{6}$$

by the expression

$$\begin{aligned}\mathbf{j}(\mathbf{q}, \omega) &= -i\omega(1/2\pi)^3\alpha(\omega)\sum_a \int d^3p \mathbf{E}_0(\mathbf{p}, \omega)\\ &\times\exp\{-i(\mathbf{p} - \mathbf{q})\cdot\mathbf{R}_a - g|Z'(\mathbf{R}_a)/2n_z|\}.\end{aligned} \tag{7}$$

Here, $\alpha(\omega)$ is the polarizability of a molecule, $\mathbf{R}_a$ is the radius vector of the center of inertia of a molecule, and the summation is performed over all molecules of the substance. Assuming that the material of the wedge is homogeneous, we can neglect the effect of fluctuations of the polarization current by replacing (7) by its value averaged over the coordinates of the molecules:

$$\begin{aligned}\mathbf{j}(\mathbf{q}, \omega) &= -i\omega(1/2\pi)^3\alpha(\omega)\int d^3p \mathbf{E}_0(\mathbf{p}, \omega)\\ &\times\left\langle\sum_a \exp\{-i(\mathbf{p} - \mathbf{q})\cdot\mathbf{R}_a - g|Z'(\mathbf{R}_a)/2n_z|\}\right\rangle.\end{aligned} \tag{8}$$

The energy of the field irradiated from the medium over the entire flight time of a particle in the frequency interval $d\omega$ near the resonance frequency, into an element

$d\Omega$ of solid angle in the direction of the vector $\mathbf{k} = (\omega/c)\mathbf{r}/r$ can be expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{(2\pi)^6}{c}|[\mathbf{k}\cdot\mathbf{j}(\mathbf{k},\omega)]|^2$$

$$= \frac{\omega^2}{c}\left|\alpha(\omega)\int d^3p[\mathbf{k}\cdot\mathbf{E}_0(\mathbf{p},\omega)]\right. \quad (9)$$

$$\left.\times\left\langle\sum_a \exp\{-i(\mathbf{p}-\mathbf{k})\cdot\mathbf{R}_a - g|Z'(\mathbf{R}_a)/2n_z|\}\right\rangle\right|^2.$$

Equations (4) and (5) show that $Z'(\mathbf{R}_a)$ is independent of Y_a; therefore, averaging with respect to Y_a results in the delta function $\delta(p_y - k_y)$. Since $\mathbf{E}_0(\mathbf{p},\omega)$ contains $\delta(\omega - p_z)$, the integration with respect to p_y and p_z is performed with the use of these delta functions. The remaining integral with respect to p_x is given by ($\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors along the axes x and y)

$$\int_{-\infty}^{\infty} dp_x\mathbf{E}_0(p_x,k_y,\omega/v)\exp\{-ip_x(X_a-a)\} \quad (10)$$
$$= (ie/2\pi Q)\mathbf{L}\exp\{-aQ+QX_a\},$$

$$\mathbf{L} = \mathbf{e}_y k_y - iQ\mathbf{e}_x - \frac{\mathbf{v}\omega(1-v^2/c^2)}{v^2}, \quad (11)$$
$$Q = \left[k_y^2 + \left(\frac{\omega}{v}\right)^2\left(1-\frac{v^2}{c^2}\right)\right]^{1/2}.$$

When considering the forward radiation, one should use expression (4) for Z', so that

$$\frac{g|Z'(\mathbf{R}_a)|}{2n_z} = -\frac{g}{2}\frac{X_a+\eta Z_a}{\eta n_z + n_x} \equiv -h(X_a+\eta Z_a). \quad (12)$$

Hence, after averaging with respect to Y_a and integrating with respect to p_x, it remains to determine the mean value with respect to the coordinates X_a and Z_a of the form

$$\left\langle\sum_a \exp\{ik_xX_a + Q(X_a-a) - i(k_z-\omega/v)Z_a + h(X_a+\eta Z_a)\}\right\rangle$$
$$= n_0\exp(-aQ)\int_{-\infty}^{0} dX\exp\{ik_xX + QX + hX\}\int_{X/\xi}^{-X/\eta} dZ\exp\{i(k_z-\omega/v)Z + h\eta Z\} \quad (13)$$
$$= \frac{n_0(\eta+\xi)\exp(-aQ)}{[k_x\xi + (k_z-\omega/v) - ih\eta - i(Q+h)\xi][k_x\eta - (k_z-\omega/v) + iQ\eta]}.$$

The distribution of the radiated energy can be expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2\omega^2}{4c\pi^2Q^2}|n_0\alpha(\omega)[\mathbf{k}\times\mathbf{L}]|^2\frac{(\xi+\eta)^2\exp(-2aQ)}{[(k_x\xi + k_z - \omega/v)^2 + (Q\xi + h\eta + h\xi)^2][(k_x\eta - k_z + \omega/v)^2 + (Q\eta)^2]}. \quad (14)$$

The radiated energy as a function of the distance a between the particle trajectory and the edge of the wedge is given by the exponential function

$$\exp(-2aQ) = \exp\{-2a[k_y^a + (\omega/\gamma v)^2]^{1/2}\},$$

which decreases with increasing distance much faster in the nonrelativistic case than in the ultrarelativistic case, similar to the behavior of the field of a fast particle. This exponential function is also essential for the dependence of the radiation intensity on the azimuthal angle φ. In the nonrelativistic case,

$$Q \approx \frac{\omega}{v} \gg k_y$$

and the function $\exp(-2aQ)$ takes values of the same order of magnitude for different φ. In the ultrarelativistic case, for very small k_y, $Q \sim \omega/\gamma v$, whereas, for large k_y,

$$Q \sim k_y \gg \frac{\omega}{\gamma v}.$$

Therefore, in the ultrarelativistic case, the function $\exp(-2aQ)$ strongly suppresses the radiation for finite k_y. As a result, when the particle moves perpendicular to the edge of the wedge, the whole radiation is virtually concentrated near the plane xz. Taking into account that

$$\mathbf{L} \approx -\frac{i\mathbf{e}_x\omega}{\gamma c} \text{ and } [\mathbf{k}\times\mathbf{L}]^2 \approx \left(\frac{\omega}{c\gamma}\right)^2 k_z^2$$

for $k_y = 0$ and $\gamma \gg 1$, we can obtain the following expression for the distribution of diffraction radiation in the ultrarelativistic case for $k_y = 0$, i.e., in the plane xz:

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2\omega^2k_z^2}{c}|n_0\alpha(\omega)|^2\frac{(\xi+\eta)^2\exp(-2a\omega/\gamma c)}{[(k_x\xi + k_z - \omega/v)^2 + (h\xi + h\eta + \xi\omega/\gamma c)^2][(k_x\eta - k_z + \omega/v)^2 + (\eta\omega/\gamma c)^2]}. \quad (15)$$

In the ultrarelativistic case, the following relations hold for small angles ϑ between the wave vector $\mathbf{k}$ of the radiated wave and the particle velocity $\mathbf{v}$:

$$k_x \sim k\vartheta, \quad k_z - \omega/v \sim (k/2)(\vartheta^2 + 1/\gamma^2);$$

therefore, expression (15) can be rewritten as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = e^2c^2k_z^2|n_0\alpha(\omega)|^2\frac{(\xi+\eta)^2\exp(-2a\omega/\gamma c)}{\{[\vartheta\xi + (\vartheta^2+\gamma^{-2})/2]^2 + (u\xi + u\eta + \xi/\gamma)^2\}\{[\vartheta\eta - (\vartheta^2+\gamma^{-2})/2]^2 + (\eta/\gamma)^2\}}, \quad (16)$$

where $u = h/k$. Note that the condition $n_x/n_z < \eta$, which guarantees the applicability of formula (4), remains valid for arbitrary values of the coefficient ξ in the equation for the first face of the wedge $x = \xi z$. It is interesting to compare the radiation in the limiting cases $\xi = 0$ (the first face of the wedge is parallel to the particle velocity) and $\xi = \infty$ (the first face of the wedge is perpendicular to the particle velocity).

For $\xi = 0$, the distribution of radiation energy (16) is expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = e^2c^2k_z^2|n_0\alpha(\omega)|^2\frac{\eta^2\exp(-2a\omega/\gamma c)}{\{(u\eta)^2 + (\vartheta^2+\gamma^{-2})^2/4\}\{[\vartheta\eta - (\vartheta^2+\gamma^{-2})/2]^2 + (\eta/\gamma)^2\}}. \quad (17)$$

In the limiting case $\xi = \infty$, the radiation distribution (16) takes the form

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = e^2c^2k_z^2|n_0\alpha(\omega)|^2\frac{\exp(-2a\omega/\gamma c)}{\{\vartheta^2 + (u+1/\gamma)^2\}\{[\vartheta\eta - (\vartheta^2+\gamma^{-2})/2]^2 + (\eta/\gamma)^2\}}. \quad (18)$$

The condition $n_x/n_z < \eta$, which guarantees the applicability of formula (4), remains valid for $\eta \longrightarrow \infty$ when the second face of the wedge is perpendicular to the particle velocity. In this case, we have to take into account that

$$h = g/2(\eta n_z + n_x),$$

so that, in the limit as

$$\eta \longrightarrow \infty,$$

we have

$$u\eta \longrightarrow g/2k_z.$$

In this case, the distribution of radiated energy is expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = e^2c^2k_z^2|n_0\alpha(\omega)|^2\frac{\exp(-2a\omega/\gamma c)}{\{[\vartheta\xi + (\vartheta^2+\gamma^{-2})/2]^2 + [u\xi + g/2k + \xi/\gamma]^2\}\{\vartheta^2 + (1/\gamma)^2\}}. \quad (19)$$

The factor $\{\vartheta^2 + (1/\gamma)^2\}$ in the denominator of (19) leads to stronger dependence of the radiation intensity on the particle energy. It can easily be seen that, in the limiting case of large u, i.e., for small absorption length of the transverse resonance waves in the substance, distribution (19) becomes independent of ξ, i.e., of the slope of the first face of the wedge. This is associated with the fact that, for small absorption length, the radiation is mainly emitted by the molecules situated near the second face of the wedge.

A comparison of (19) and (16) shows that the slope of the second face of the wedge with respect to the particle velocity weakens the dependence of the radiation intensity on the Lorentz factor of the particle.

The simplest limiting case corresponds to $\xi = 0$ and $\eta = \infty$, when the substance occupies the domain of negative x and negative z. Such a configuration of a substance can be called a step rather than a wedge. In this case, the distribution of diffraction radiation from an ultrarelativistic particle can easily be represented as

$$\frac{d^2E(\mathbf{n}, \omega)}{d\omega d\Omega} = e^2c^2k_z^2|n_0\alpha(\omega)|^2 \times \frac{\exp(-2a\omega/\gamma c)}{\{(\vartheta^2 + \gamma^{-2})^2/4 + (g/2k)^2\}\{\vartheta^2 + (1/\gamma)^2\}}. \quad (20)$$

3. DIFFRACTION RADIATION FROM AN INHOMOGENEOUS SURFACE

Now, consider another example of diffraction radiation at a frequency close to a natural frequency of the medium in the case when a particle moves by the law

$$x = a, \quad y = 0, \quad z = vt$$

past a homogeneous medium with the density of molecules n_0 and the volume of the medium is bounded by the condition $X_a < \zeta(Y_a, Z_a)$. It should be noted that no radiation is emitted when $\zeta = \text{const}$, i.e., when a charged particle moves uniformly parallel to the surface of a homogeneous medium. The reason lies in the fact that the energy and momentum conservation laws for the radiation hold only for the transfer of longitudinal momentum (along the direction of the particle velocity) to the medium; however, for such an orientation of the surface of a homogeneous medium, a transfer of longitudinal momentum is impossible due to the homogeneity of the conditions of the problem along the z coordinate. This means that the radiation field is zero for $\zeta = \text{const}$; i.e., radiation is emitted due to the irregularities of the surface, which makes possible the transfer of longitudinal momentum to the medium.

Thus, it is the polarization currents near the irregularities of the surface that are responsible for the diffraction radiation from a homogeneous medium, whereas the polarization currents in the bulk of the medium do not contribute to the diffraction radiation. Let us find a point on the surface $X = \zeta(Y, Z)$ of the medium with the minimal value of the coordinate X and choose the coordinate axes so that the plane $X = 0$ passes through this point. As pointed out above, polarization currents in the domain $X < 0$ do not contribute to the diffraction radiation. This fact allows one to take into account only the current density in the layer enclosed between the surface $X = \zeta(Y, Z)$ and the plane $X = 0$ when calculating the Fourier image of the density of polarization current in (4).

To avoid nonessential complications, we restrict the consideration to the diffraction radiation from a medium with the surface of the form $X = \zeta(Z)$. The amplitude of the transverse resonance wave radiated by a molecule situated at the point $\mathbf{R}_a$ exponentially decays in the medium due to inelastic scattering. The direction to the observation point is specified by a unit vector $\mathbf{n}$; this direction intersects the surface of the medium at a point $\mathbf{R}_a + \mathbf{R}'(X', Y', Z')$ whose z coordinate

$$Z_a + Z' \equiv Z_a + Z'(\mathbf{R}_a)$$

can be determined from the equation

$$\xi(Z_a + Z') = X_a + Z'(n_x/n_z). \quad (21)$$

The distance between the emitting molecule and the above intersection point is given by

$$|\mathbf{R}'(X', Y', Z')| = |Z'(\mathbf{R}_a)/n_z|.$$

A falloff in the amplitude of the resonance wave along the whole path in the homogeneous medium can be taken into account with the use of expression (7) for the Fourier image of the mean polarization current. Introducing the notation

$$S(\mathbf{p}) = (2\pi)^{-3}\int_{-\infty}^{\infty} dY \int_{-\infty}^{\infty} dZ \int_0^{\zeta(Y,Z)} dX \times \exp\{-i\mathbf{p}\cdot\mathbf{R} - gZ'(\mathbf{R})/2n_z\}, \quad (22)$$

we can write

$$\left\langle \sum_a \exp\{-i\mathbf{p}\cdot\mathbf{R}_a - gZ'(\mathbf{R}_a)/2n_z\} \right\rangle = n_0(2\pi)^3 S(\mathbf{p}). \quad (23)$$

Equation (22) shows that the domain of large values of the variable X does not make an appreciable contribution to the integral for the following two reasons: due to the decreasing exponential function $\exp\{-gZ'(\mathbf{R})/2n_z\}$ and due to the upper limit. Therefore, two qualitatively different cases are possible. In the first case, the thickness of the effective layer of irregularities is less than the absorption length of the resonance wave; in this case, a key role is played by the upper limit, while the decreasing exponential function varies slightly within the integration interval. Therefore, one can neglect the decreasing exponential function, which is equivalent to using the current in the form (6) instead of (7). In this case, $S(\mathbf{p})$ takes the form

$$S(\mathbf{p}) = (2\pi)^{-3}\int_{-\infty}^{\infty} dY \int_{-\infty}^{\infty} dZ \int_0^{\zeta(Y,Z)} dX \exp\{-i\mathbf{p}\cdot\mathbf{R}\} \quad (24)$$

instead of (22).

In the second case, the absorption length of a resonance wave is small compared with the thickness of the effective layer of irregularities. In this case, the upper limit of integration with respect to X in (22) plays a small role, and one can integrate with respect to X in (22) up to infinity. The effect of the surface shape manifests itself in the fact that the function $Z'(\mathbf{R}_a)$ is determined from Eq. (21), which is completely defined by the shape of the surface. Then, formula (22) is rewritten as

$$S(\mathbf{p}) = (2\pi)^{-3}\int_{-\infty}^{\infty}dY\int_{-\infty}^{\infty}dZ\int_{0}^{\infty}dX \times \exp\{-i\mathbf{p}\cdot\mathbf{R} - gZ'(\mathbf{R})/2n_z\}. \quad (25)$$

An estimate for (25) can be obtained using the properties of the function $Z'(\mathbf{R})$, i.e., using the specific shape of the surface.

Consider the first case, when the thickness of the layer of irregularities is less than the absorption length of the resonance wave. Then, we can neglect the amplitude decay due to scattering of the radiated wave in the substance. In this case, the distribution of the radiated energy is given by

$$\left\langle\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega}\right\rangle = c\omega^2 n_0^2(2\pi)^6 \times\left|\alpha(\omega)\int d^3p[\mathbf{k}\times\mathbf{E}_0(\mathbf{p})]S(\mathbf{p}-\mathbf{k})\delta(\omega-p_z v)\right|^2, \quad (26)$$

where $S(\mathbf{p}-\mathbf{k})$ is defined in (24). It is worth noting that, in the particular case $\zeta(Y, Z) = \zeta(Y)$, when the surface shape is independent of z, $S(\mathbf{p}-\mathbf{k})$ is proportional to $\delta(p_z - k_z)$, and $\delta(\omega - p_z v)\delta(p_z - k_z) = 0$ if there is no Cherenkov radiation in the medium. Thus, there is no diffraction radiation when $\zeta(Y, Z) = \zeta(Y)$ because a longitudinal momentum cannot be transferred to such a medium.

Now, consider the diffraction radiation from a medium whose surface is given by $X = \zeta(Z)$. In this case, the integral with respect to $\mathbf{p}$ in (26) contains a product of the delta functions $\delta(p_y - k_y)$ and $\delta(\omega - p_z v)$, which are used for integrating with respect to p_y and p_z. As a result, (26) is rewritten as

$$\left\langle\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega}\right\rangle = \frac{\omega^2 n_0^2(2\pi)^2}{c}\left|\alpha(\omega)\int_{-\infty}^{\infty}dZ \times\exp\{-i(k_z-\omega/v)Z\}\int_{0}^{\zeta(Z)}dX\exp(ik_xX) \times\int_{-\infty}^{\infty}dp_x[\mathbf{k}\times\mathbf{E}_0(p_x,k_y,\omega/v)]\exp(-ip_xX)\right|^2. \quad (27)$$

The integral with respect p_x contained in (27) is given in (10). Thus, upon integration, formula (27) is expressed as

$$\left\langle\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega}\right\rangle = e^2\frac{\omega^2 n_0^2}{cQ^2} \times\exp(-2aQ)|\alpha(\omega)[\mathbf{k}\times\mathbf{L}(\mathbf{k})]|^2 \times\left|\frac{1}{Q-ik_x}\int_{-\infty}^{\infty}dZ\exp\{-i(k_z-\omega/v)Z\} \times\exp\{(Q+ik_x)\zeta(Z)\}\right|^2. \quad (28)$$

In the particular case when the wavelength is much greater than the thickness of the layer of surface irregularities, $k_x\zeta \ll 1$, $Q\zeta \ll 1$, and expression (28) can be simplified to

$$\left\langle\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega}\right\rangle = e^2\frac{\omega^2 n_0^2}{cQ^2} \times\exp(-2aQ)|\alpha(\omega)[\mathbf{k}\times\mathbf{L}(\mathbf{k})]|^2 \times\left|(Q-ik_x)\int_{-\infty}^{\infty}dZ\zeta(Z)\exp\{-i(k_z-\omega/v)Z\}\right|^2. \quad (29)$$

In this case, the radiation distribution is determined by the Fourier image of the surface shape $\zeta(k_z - \omega/v)$.

4. DIFFRACTION RADIATION FROM A MEDIUM WITH PERIODIC SURFACE

Now, consider the diffraction radiation at a resonance frequency from a homogeneous substance with periodic, in the Z coordinate, surface of the form $X = 2b\{1 + \cos(gZ)\}$. Applying the well-known expansion of the exponential function in terms of Bessel functions,

$$\exp(iu\cos\Phi) = \sum_s i^s J_s(u)\exp(is\Phi), \quad (30)$$

we can reduce the integral with respect to Z in (28) to

$$M(\mathbf{k}) = \int_{-\infty}^{\infty}dZ\exp\{-i(k_z-\omega/v)Z\} \times\exp\{(Q-ik_x)2b(1+\cos gZ)\} = 2\pi\exp\{2b(Q-ik_x)\}\sum_s i^s J_s(2bQ-2ibk_x) \times\delta(sg+k_z-\omega/v). \quad (31)$$

The terms in the sum in (31) take nonzero values only

for those irradiation angles ϑ for which the argument of the delta function passes through zero, i.e., for

$$[1 - (v/c)\cos\vartheta] = sg(v/\omega). \quad (32)$$

This condition stems from the fact that, for fixed values of ω, ϑ, and v, the value of the longitudinal momentum transferred from the field to the medium is specified by the conservation law, while the periodically inhomogeneous surface can accept only a certain fixed momentum. Diffraction radiation from a periodically inhomogeneous surface was considered in various approximations by many authors, and expression (32) was often called a resonance condition. Under these conditions, diffraction radiation was also referred to as resonance radiation, although the frequency of this radiation was not related to the natural frequencies of the medium. Following this terminology, we can call the diffraction radiation *resonance radiation at resonance frequencies.* The energy distribution of this radiation is given by

$$\left\langle \frac{d^2E(\mathbf{n}, \omega)}{d\omega d\Omega} \right\rangle = 2T\pi e^2 \frac{\omega^2 n_0^2}{cQ^2}$$

$$\times \exp\{-2(a-b)Q\}|\alpha(\omega)\mathbf{L}(\mathbf{k})|^2 \quad (33)$$

$$\times \frac{1}{Q^2 + k_x^2} \sum_s |J_s(2bQ - 2ibk_x)|^2 \delta(sg + k_z - \omega/v),$$

where T is the total flight time.

5. DISCUSSION OF THE RESULTS

The results obtained above can be applied under the condition that the frequency of diffraction radiation is close to a certain resonance frequency $\omega_n = E_n/\hbar$, where E_n is the energy of an exciton. Transverse waves with such a frequency are strongly absorbed in the substance; therefore, the only transverse resonance waves that may leave the substance are those radiated by molecules lying close to the surface of the medium. Due to the small thickness of the layer of radiating molecules, one can restrict analysis to the case of single scattering of the field of a fast particle in the substance, which is accompanied by the radiation of a transverse resonance wave. This method is a generalization of the Fermi method for analyzing the reflection and scattering of transverse resonance waves [1]. The single-scattering approximation allows one to consider diffraction radiation microscopically as a transformation of the Fourier component of the particle field into a transverse resonance wave due to the interaction with a molecule. In this approximation, the radiation energy is a simple function of the coordinates of the molecules of the medium; this makes the analysis of diffraction radiation from complex surfaces simpler compared with the procedure involving boundary conditions in macroscopic electrodynamics. The method considered does not allow one to calculate the intensity of diffraction radiation for arbitrary frequencies. However, as applied to the problem of diffraction radiation from a dielectric wedge considered below, this method makes it possible to relatively easily investigate the radiation energy as a function of the wedge angle. In the problem of diffraction radiation from a nonplanar surface of a homogeneous medium, this method allows one to calculate rather easily the radiation energy as a function of the surface shape of the medium.

The analysis of diffraction radiation at a resonance frequency without involving macroscopic boundary conditions can be applied to the experimental investigation of the structure of complex surfaces. To obtain more detailed information, one can measure radiation energy at different values of the particle velocity and at several frequency points close to different natural frequencies of the medium.

REFERENCES

1. E. Fermi, in *Scientific Works* (Nauka, Moscow, 1971), Vol. 1, p. 150 [in Russian].
2. B. M. Bolotovskiĭ and G. V. Voskresenskiĭ, Usp. Fiz. Nauk **94**, 377 (1968) [Sov. Phys. Usp. **11**, 143 (1968)].
3. Yu. N. Dnestrovskiĭ and D. P. Kostomarov, Dokl. Akad. Nauk SSSR **116**, 377 (1957) [Sov. Phys. Dokl. **2**, 442 (1957)]; **124**, 1026 (1959) [**4**, 158 (1959)].
4. A. P. Kazantsev and G. I. Surdutovich, Dokl. Akad. Nauk SSSR **147**, 74 (1962) [Sov. Phys. Dokl. **7**, 990 (1963)].
5. V. P. Shestopalov, *Diffraction Electronics* (Khark. Gos. Univ., Kharkov, 1976) [in Russian].
6. J. H. Brownell, G. Doucas, and J. Walsh, Phys. Rev. E **57**, 1075 (1998).
7. A. P. Potylitsin, Nucl. Instrum. Methods Phys. Res. B **145**, 169 (1998).
8. A. P. Potylitsin, Phys. Lett. A **238**, 112 (1998).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 475–486.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 536–550.

ATOMS, MOLECULES, OPTICS

Coherent Control of Quantum Correlations in Atomic Systems

A. M. Basharov*, A. A. Bashkeev, and É. A. Manykin*****
Russian Research Center Kurchatov Institute, Moscow, 123182 Russia
*e-mail: *bash@online.ru; **_brilliant_@rambler.ru; ***edmany@isssph.kiae.ru*
Received July 20, 2004

Abstract—A new mechanism of development of quantum correlations due to the combined effects of a common heat bath and a resonant classical electromagnetic field is analyzed for a system of two two-level atoms coupled by dipole–dipole interaction. Conditions are found under which arbitrary steady-state entanglement can be created in the absence of measurement in an entangled basis. The analysis of dynamics of entanglement is focused on the effects of dipole–dipole interaction, mean number of thermal photons, degree of squeezing of thermal radiation, classical field strength, and presence of additional decay channels. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Entangled states of quantum objects involve correlations that preclude description in terms of a factorized wavefunction or density matrix. In view of the measurement postulate, the correlations lead to the paradoxes of quantum theory formulated by Einstein, Podolsky, and Rosen [1] and Schrödinger [2]. Present-day interest in entangled states is motivated by their prospective applications in quantum cryptography, quantum computing, and quantum teleportation [3, 4], which motivate extensive studies of conditions and methods for preparation of entangled states, as well as their properties and possible utility [5].

Currently, entangled states are created in the form of two-mode squeezed light, which is widely used in quantum optics. Preparation and control of entangled states of atoms, including many-particle ones, presents a more difficult problem, which is the subject of ongoing theoretical and experimental studies. The possibility of creation of two entangled macroscopic ensembles of atoms in free space has been demonstrated in a recent experiment [6]. Analysis of many-particle states and methods for creation and control of entangled states is largely stimulated by the model of a one-way quantum computer based on quantum measurement [7]. Progress has also recently been made toward teleportation of atomic states [8].

In this paper, we analyze entanglement of atomic states and conditions for entanglement control using electromagnetic field. Our results suggest a new mechanism of entanglement and demonstrate a possibility for entanglement control. For simplicity, we consider a system of two two-level atoms. A two-state quantum system is called a qubit in quantum information theory.

Entanglement of atomic states can be created by three principal mechanisms: (1) dynamic (e.g., dipole–dipole) interaction between subsystems, (2) interaction with a common heat bath, and (3) measurement of independent atomic systems (in an entangled basis).

The first mechanism does not seem to call for any comment, being directly responsible for entanglement of states of atomic subsystems. However, the analysis presented in this paper shows that its role in steady-state entanglement may vary substantially, depending on the relative contributions of interactions with a common heat bath and independent heat baths specific to individual atoms to the overall relaxation process.

In the nontrivial model of entanglement by interaction of atoms with a common heat bath considered in [9–12], the state space of the system is decomposed into two subspaces with different particle permutation symmetries. In this model, the coupling to a common heat bath makes it possible to sustain different types of dynamics in these subspaces: the evolution of states in the decoherence-free subspace is unitary [13], whereas the Dicke states (which span the other subspace) are characterized by nonunitary (dissipative) dynamics. Owing to the difference in dynamics, initially disentangled atomic states that do not lie entirely in either subspace become entangled. In superradiance theory, the states belonging to these subspaces are called subradiant and superradiant, respectively [14, 15], and entanglement can also be created if the dissipative dynamics of subradiant states are different from those of superradiant states. Note that these states are responsible for the experimentally observed dependence of spontaneous emission time on interatomic distance [16, 17] (cf. the theory in [18]).

Measurements of atomic systems usually involve detection of photons emitted by atoms. If the atom that has emitted a detected photon cannot be identified, then the state of an initially doubly excited two-atom system

immediately after the click of the detector will be a superposition of the ground state of the atom that has emitted the photon and the excited state of the other atom. Thus, an Einstein–Podolsky–Rosen (EPR) pair, i.e., a maximally entangled pair of qubits, is created. Note that the resulting wavefunction characterizes the so-called posterior state of the atoms [19]. A measurement is the projection of the measured state onto the subspace spanned by the entangled states. The entanglement of the subsystem thus selected is exactly due to a nonselective measurement, which is related to the indistinguishability of quantum subsystems with respect to some observable. The entanglement is totally unrelated to any collective process, being entirely due to an effective measurement in an entangled basis. A specific analogy between the preparation of entangled states of a system of noninteracting subsystems by state reduction (measurement) and quantum state evolution was considered in [20].

To control the creation of entangled states, one must be able to change the system's parameters. The simplest way to control the dynamics of an atomic system is coherent control, i.e., coupling to a coherent field. However, interaction between a system and a coherent field, as well as any classical field, does not necessarily give rise to quantum correlations between its noninteracting subsystems. For this reason, the present analysis of the dynamics of creation of entangled states, including steady states, of a system driven by a resonant coherent field takes into account both dipole–dipole coupling and decay by interaction with common and independent thermal reservoirs.

The criterion for entanglement is a separate problem. The existence of quantum correlations in a system can be checked by applying Bell's inequality [21]. However, no general operational criterion for entanglement of arbitrary systems is available to this day. The concurrence introduced by Wootters as an entanglement measure [22] cannot be used to formulate necessary and sufficient conditions. In this study, we use the Peres–Horodecki criterion, which apparently is the only available necessary and sufficient condition for entanglement (inseparability) of bipartite mixed states [23, 24]. According to this criterion, a state is entangled (inseparable) if at least one eigenvalue of the (partial) transpose of the density matrix with respect to the variables associated with one atom is negative. (The transpose is henceforth called *the Peres–Horodecki matrix*.) Then, there exist quantum correlations in the system, and the sum of the negative eigenvalues of the Peres–Horodecki matrix provides a measure for the correlations. The present analysis is illustrated by graphs of the absolute value of the lowest eigenvalue of the Peres–Horodecki matrix.

The results of our study of the dynamics of a resonantly driven, spontaneously decaying two-atom system coupled to a common heat bath suggest that a classical field offers an effective tool for coherent control based on a new mechanism of entanglement. It has been shown that entanglement does not arise when the doubly excited state of the system decays via spontaneous emission [9]. However, their interaction with a coherent resonant pulse can lead to both time-dependent and steady-state entanglement. We find conditions for quantum correlation corresponding to a specific value of the Peres–Horodecki measure between –0.5 (for an EPR pair) and zero. We demonstrate a difference in the role played by dipole–dipole interaction in the creation of a steady entangled state between the case of decay by interaction with a common heat bath and the case when the decaying atoms interact with independent heat baths in the absence of collective relaxation. With increasing contribution of independent decay channels, the mechanism of steady-state entanglement by interaction with a common heat bath transforms into a mechanism dominated by dipole–dipole interaction, while the interactions with a classical field and independent heat baths only guarantee relaxation to a steady state. The results obtained here can be summarized as follows: a qubit with a required entanglement measure can be prepared via quantum state evolution (without performing any measurement-induced projection) by coupling an atomic system to electromagnetic fields. This is a promising possibility for quantum communication, because the resulting entangled state has a longer lifetime.

The paper is organized as follows. In Section 2, we formulate a master equation for a two-atom density matrix describing the processes of interest here. In Section 3, we analyze the quantum correlations arising between the atoms when the resonantly driven system is coupled to a common heat bath at zero temperature, taking into account dipole–dipole interaction. In Section 4, we discuss the case when the common heat bath is in the maximally squeezed state at a finite temperature. In Section 5, we present the numerical results obtained for a system coupled to an arbitrary common heat bath and independent heat baths specific to individual atoms. We analyze the change in characteristics of coherent control caused by deviation from optimal parameters.

Preliminary results of this study have been reported at several conferences [25].

2. DESCRIPTION OF THE MODEL AND GOVERNING EQUATIONS

The two qubits considered here can be impurity atoms (or ions) in a photonic crystal or single-mode fiber, or atoms in a magnetic trap or microcavity, or atoms in other man-made environments. All of these systems can be accurately described by models of collective decay by interaction with a common heat bath, which can be reduced to models of atom dynamics in one-dimensional resonant fields. We consider a classi-

cal electromagnetic wave propagating in the positive z direction and represent the electric field of the wave as

$$\mathbf{E} = \mathscr{E}\exp[-i(\omega t - kz)] + \mathscr{E}^*\exp[i(\omega t - kz)].$$

In the dipole approximation, the density matrix $\mathscr{R}$ of the two-atom system obeys the master equation

$$\frac{d\mathscr{R}}{dt} = \frac{i}{\hbar}[\mathscr{R}, \mathscr{H}] - \hat{\Gamma}_0\mathscr{R}. \quad (1)$$

Here, the Hamiltonian $\mathscr{H}$ of the two-atom system is the sum of the Hamiltonians $\mathscr{H}_A$ of isolated atoms, an operator $\mathscr{V}_{cl}$ representing the dipole–field interaction, and the dipole–dipole interaction operator $\mathscr{V}_{d-d}$,

$$\mathscr{H} = \mathscr{H}_A + \mathscr{V}_{cl} + \mathscr{V}_{d-d},$$

where

$$\begin{aligned}\mathscr{H}_A &= \frac{1}{2}\hbar\omega_0(C_3 \otimes 1 + 1 \otimes C_3),\\ \mathscr{V}_{cl} &= -(\mathbf{E}\cdot\mathbf{d}\otimes\hat{1} + \hat{1}\otimes\mathbf{E}\cdot\mathbf{d}),\\ \mathscr{V}_{d-d} &= \frac{\mathbf{d}\otimes\mathbf{d}}{R^3},\\ C_3 &= |1\rangle\langle 1| - |0\rangle\langle 0|,\end{aligned} \quad (2)$$

R denotes the distance between the qubits; in the case of Markovian evolution of the system coupled to a common heat bath, the relaxation superoperator Γ_0 has the form (see [9])

$$\begin{aligned}&\hat{\Gamma}_0\mathscr{R}\\ &= \frac{\chi|d_{10}|^2}{2}(N+1)(\mathscr{R}\mathscr{C}_+\mathscr{C}_- + \mathscr{C}_+\mathscr{C}_-\mathscr{R} - 2\mathscr{C}_-\mathscr{R}\mathscr{C}_+)\\ &+ \frac{\chi|d_{10}|^2}{2}N(\mathscr{R}\mathscr{C}_-\mathscr{C}_+ + \mathscr{C}_-\mathscr{C}_+\mathscr{R} - 2\mathscr{C}_+\mathscr{R}\mathscr{C}_-)\\ &+ \frac{\chi d_{10}^2}{2}(2\mathscr{C}_+\mathscr{R}\mathscr{C}_+ - \mathscr{R}\mathscr{C}_+\mathscr{C}_+ - \mathscr{C}_+\mathscr{C}_+\mathscr{R})M\\ &+ \frac{\chi d_{10}^{*2}}{2}(2\mathscr{C}_-\mathscr{R}\mathscr{C}_- - \mathscr{R}\mathscr{C}_-\mathscr{C}_- - \mathscr{C}_-\mathscr{C}_-\mathscr{R})M^*,\end{aligned} \quad (3)$$

with

$$\begin{aligned}\mathscr{C}_\pm &= C_\pm\otimes\hat{1} + e^{\mp i\theta}\hat{1}\otimes C_\pm,\\ C_- &= |0\rangle\langle 1|, \quad C_+ = |1\rangle\langle 0|;\end{aligned}$$

the state vectors $|0\rangle$ and $|1\rangle$ correspond to the ground and excited energy levels; $\hbar\omega_0$ is the energy difference between the levels; $\mathbf{d}$ is the atomic dipole moment operator; d_{10} is a matrix element of the dipole moment operator; and θ is the phase gained by the wave as it travels the distance between the atoms.

The common heat bath is a quantized one-dimensional electromagnetic field whose initial state $|\Phi_0\rangle$ is delta-correlated in time:

$$\begin{aligned}\langle\Phi_0|b_\omega^\dagger b_{\omega'}|\Phi_0\rangle &= N(\omega)\delta(\omega-\omega'),\\ \langle\Phi_0|b_\omega b_{\omega'}^\dagger|\Phi_0\rangle &= [1+N(\omega)]\delta(\omega-\omega'),\\ \langle\Phi_0|b_\omega b_{\omega'}|\Phi_0\rangle &= M(\omega)\delta(2\Omega_\Gamma-\omega-\omega'),\\ \langle\Phi_0|b_\omega^\dagger b_{\omega'}^\dagger|\Phi_0\rangle &= M^*(\omega)\delta(2\Omega_\Gamma-\omega-\omega'),\end{aligned}$$

where b_ω and $b_\omega^\dagger$ denote, respectively, the annihilation and creation operators for thermal photons with wavenumber $\mathbf{q} = q\mathbf{e}_z = \omega\mathbf{e}_z/c$ and $N(\omega)$ is the spectral density of the bath. The common heat bath is in a squeezed state characterized by a central frequency Ω_Γ and a degree of squeezing

$$M(\omega) = |M(\omega)|\exp(-2i\Omega_\Gamma t),$$

$$M(\omega) \le \sqrt{N(\omega)[N(\omega)+1]}.$$

The evolution of atoms coupled to the heat bath is described in the Markovian approximation. The coupling constant χ, the parameters $N(\omega) = N(\omega_0) = N$ and $M(\omega) = M(\omega_0) = M$, and the exponential factor

$$\begin{aligned}&\exp[i\omega(\mathbf{r}_2-\mathbf{r}_1)\cdot\mathbf{e}_z/c]\\ &= \exp[i\omega_0(\mathbf{r}_2-\mathbf{r}_1)\cdot\mathbf{e}_z/c] = \exp(i\theta)\end{aligned}$$

representing the phase gain are independent of frequency ($\mathbf{r}_1$ and $\mathbf{r}_2$ are the locations of the atoms, $|\mathbf{r} - \mathbf{r}_2| = R$). The polarization states of thermal photons are ignored.

In the limit case of $M = N = 0$, the heat bath defined above reduces to a one-dimensional vacuum field. In other cases, it provides a good model of the electromagnetic field produced by a degenerate optical parametric oscillator [26].

The recoil effects due to photon emission and absorption are neglected in deriving both the model of qubits driven by a classical field and the relaxation superoperator. Polarization effects are neglected and the boldface vector notation is not used, since the electric field and dipole moment are assumed to be parallel.

Superoperators having the form of (3) with $\theta = 2\pi n$ ($n = 0, \pm1, \ldots$) describe interactions with one- and three-dimensional heat baths.

In the one-dimensional case, electromagnetic modes propagating in both directions are assumed to have similar characteristics. To allow for difference in parameters between thermal modes and describe the case of $\theta \ne 2\pi n$ ($n = 0, \pm1, \ldots$), one should replace the superoperator $\hat{\Gamma}_0$ with $\hat{\Gamma}_0 + \hat{\Gamma}_0'$, where the superoperator $\hat{\Gamma}_0'$ is obtained by substituting primed quantities (in

particular, changing $\mathscr{C}_\pm$ to $\mathscr{C}'_\pm = C_\pm \otimes \hat{1} + e^{\pm i\theta}\hat{1} \otimes C_\pm)$ in $\hat{\Gamma}_0$.

In the case of a three-dimensional heat bath, the atoms must be localized in a region much smaller than the resonant wavelength. However, the separation between the atoms must still be taken into account [27] by introducing relaxation superoperators $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ corresponding to independent decay channels:

$$\begin{aligned}\hat{\Gamma}_j\mathscr{R} &= \frac{\chi^{(j)}|d_{10}|^2}{2}(N^{(j)}+1)\\ &\times(\mathscr{R}\mathscr{C}_+^{(j)}\mathscr{C}_-^{(j)} + \mathscr{C}_+^{(j)}\mathscr{C}_-^{(j)}\mathscr{R} - 2\mathscr{C}_-^{(j)}\mathscr{R}\mathscr{C}_+^{(j)})\\ &+ \frac{\chi^{(j)}|d_{10}|^2}{2}N^{(j)}\\ &\times(\mathscr{R}\mathscr{C}_-^{(j)}\mathscr{C}_+^{(j)} + \mathscr{C}_-^{(j)}\mathscr{C}_+^{(j)}\mathscr{R} - 2\mathscr{C}_+^{(j)}\mathscr{R}\mathscr{C}_-^{(j)})\\ &+ \frac{\chi^{(j)}d_{10}^2}{2}M^{(j)}\\ &\times(2\mathscr{C}_+^{(j)}\mathscr{R}\mathscr{C}_+^{(j)} - \mathscr{R}\mathscr{C}_+^{(j)}\mathscr{C}_+^{(j)} - \mathscr{C}_+^{(j)}\mathscr{C}_+^{(j)}\mathscr{R})\\ &+ \frac{\chi^{(j)}d_{10}^{*2}}{2}M^{(j)*}\\ &\times(2\mathscr{C}_-^{(j)}\mathscr{R}\mathscr{C}_-^{(j)} - \mathscr{R}\mathscr{C}_-^{(j)}\mathscr{C}_-^{(j)} - \mathscr{C}_-^{(j)}\mathscr{C}_-^{(j)}\mathscr{R}),\end{aligned} \tag{4}$$

where

$$\mathscr{C}_\pm^{(1)} = C_\pm \otimes \hat{1}, \quad \mathscr{C}_\pm^{(2)} = \hat{1} \otimes C_\pm. \tag{5}$$

In other words, the operator $\hat{\Gamma}_0$ must be replaced by $\hat{\Gamma}_0 + \hat{\Gamma}_1 + \hat{\Gamma}_2$ in (1).

The Lindblad equations with relaxation superoperators having the form of (3) or (4) provide a general formalism for describing decay processes in the Markovian approximation [28]. For this reason, we believe that system (1)–(3) written for two-level atoms, with Eq. (1) in some cases replaced by

$$\frac{d\mathscr{R}}{dt} = \frac{i}{\hbar}[\mathscr{R}, \mathscr{H}] - \hat{\Gamma}_0\mathscr{R} - \hat{\Gamma}'_0\mathscr{R} - \hat{\Gamma}_1\mathscr{R} - \hat{\Gamma}'_2\mathscr{R}, \tag{6}$$

can be applied to describe the dynamics of two qubits coupled to a resonant classical and a broad-bandwidth quantized electromagnetic fields.

Equations (1) and (6) entail equations for slowly varying density-matrix elements (as compared to $\exp(\pm i\omega_0 t)$). Assuming (for simplicity) that the atomic transition frequency, the driving-field carrier frequency, and the central frequencies of all heat baths are equal, we find that the slowly varying density matrix is identical to the density matrix in the interaction representation,

$$\begin{aligned}\tilde{\mathscr{R}} &= \exp(i\mathscr{H}_A t/\hbar)\mathscr{R}\exp(-i\mathscr{H}_A t/\hbar),\\ \dot{\tilde{\mathscr{R}}} &= \frac{i}{\hbar}[\tilde{\mathscr{R}}, \tilde{\mathscr{V}}] - \hat{\Gamma}\tilde{\mathscr{R}},\\ \tilde{\mathscr{V}} &= \exp(i\mathscr{H}_A t/\hbar)\mathscr{V}\exp(-i\mathscr{H}_A t/\hbar).\end{aligned} \tag{7}$$

Averaging Eq. (7) over fast oscillations, we obtain a similar equation in which the nonzero matrix elements represent dipole–dipole interaction: $\tilde{\mathscr{V}}_{\mathrm{fl}} = d_{10}d_{01}R^{-3} = \tilde{\mathscr{V}}^*_{\mathrm{lf}}$. We use the notation for the state space basis vectors introduced in [9]:

$$\begin{aligned}|\mathrm{g}\rangle &= |0\rangle\otimes|0\rangle, \quad |\mathrm{e}\rangle = |1\rangle\otimes|1\rangle,\\ |\mathrm{f}\rangle &= |1\rangle\otimes|0\rangle, \quad |\mathrm{l}\rangle = |0\rangle\otimes|1\rangle.\end{aligned} \tag{8}$$

Here, $|\mathrm{f}\rangle$ represents the singly excited state of a two-atom system in which the first atom is excited, the second one is in the ground state, and so on. Frequently, the following partially entangled basis should be used instead of (8):

$$\begin{aligned}|\mathrm{g}\rangle, |\mathrm{a}\rangle &= \frac{1}{\sqrt{2}}(|\mathrm{f}\rangle - e^{-i\theta}|\mathrm{l}\rangle),\\ |\mathrm{s}\rangle &= \frac{1}{\sqrt{2}}(|\mathrm{f}\rangle + e^{-i\theta}|\mathrm{l}\rangle), |\mathrm{e}\rangle.\end{aligned} \tag{9}$$

The matrix elements calculated in these basis sets are related as follows:

$$\begin{aligned}\mathscr{R}_{\mathrm{ff}} &= \frac{1}{2}\{\mathscr{R}_{\mathrm{ss}} + \mathscr{R}_{\mathrm{aa}} + \mathscr{R}_{\mathrm{as}} + \mathscr{R}_{\mathrm{sa}}\},\\ \mathscr{R}_{\mathrm{ll}} &= \frac{1}{2}\{\mathscr{R}_{\mathrm{ss}} + \mathscr{R}_{\mathrm{aa}} - \mathscr{R}_{\mathrm{as}} - \mathscr{R}_{\mathrm{sa}}\},\\ \mathscr{R}_{\mathrm{el}} &= \frac{1}{\sqrt{2}}e^{i\theta}\{\mathscr{R}_{\mathrm{es}} - \mathscr{R}_{\mathrm{ea}}\},\\ \mathscr{R}_{\mathrm{ef}} &= \frac{1}{\sqrt{2}}\{\mathscr{R}_{\mathrm{ea}} + \mathscr{R}_{\mathrm{es}}\},\\ \mathscr{R}_{\mathrm{gl}} &= \frac{1}{\sqrt{2}}e^{i\theta}\{\mathscr{R}_{\mathrm{gs}} - \mathscr{R}_{\mathrm{ga}}\},\\ \mathscr{R}_{\mathrm{gf}} &= \frac{1}{\sqrt{2}}\{\mathscr{R}_{\mathrm{ga}} + \mathscr{R}_{\mathrm{gs}}\},\\ \mathscr{R}_{\mathrm{fl}} &= \frac{1}{2}e^{i\theta}\{\mathscr{R}_{\mathrm{as}} + \mathscr{R}_{\mathrm{ss}} - \mathscr{R}_{\mathrm{aa}} - \mathscr{R}_{\mathrm{sa}}\}.\end{aligned}$$

When the carrier frequency of the classical field is different from the atomic transition frequency,

another definition of slowly varying amplitudes is more suitable:

$$\mathcal{R}_{ge} = \tilde{\tilde{\mathcal{R}}}_{ge}e^{2i\omega t}, \quad \mathcal{R}_{fg} = \tilde{\tilde{\mathcal{R}}}_{fg}e^{-i\omega t},$$

$$\mathcal{R}_{fl} = \tilde{\tilde{\mathcal{R}}}_{fl}, \quad \mathcal{R}_{ee} = \tilde{\tilde{\mathcal{R}}}_{ee}, \ldots .$$

Taking into account the detuning from resonance [29], we use a standard method to derive the corresponding equations for slowly varying amplitudes $\tilde{\tilde{\mathcal{R}}}$ (dropping the double tilde in $\tilde{\tilde{\mathcal{R}}}$):

$$\left(\frac{d}{dt} + \Gamma\right)\mathcal{R}_{gg} = \frac{id_{10}^*\mathcal{E}^*}{\hbar}(\mathcal{R}_{fg} + e^{-i\theta}\mathcal{R}_{lg}) - \frac{id_{10}\mathcal{E}}{\hbar}(\mathcal{R}_{gf} + e^{i\theta}\mathcal{R}_{gl}),$$

$$\left(\frac{d}{dt} + \Gamma\right)\mathcal{R}_{ff} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(\mathcal{R}_{fg} - e^{-i\theta}\mathcal{R}_{ef}) - \frac{id_{10}\mathcal{E}}{\hbar}(e^{i\theta}\mathcal{R}_{fe} - \mathcal{R}_{gf}) + \frac{i|d_{10}|^2}{\hbar R^3}(\mathcal{R}_{fl} - \mathcal{R}_{lf}),$$

$$\left(\frac{d}{dt} + \Gamma\right)\mathcal{R}_{ll} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(e^{-i\theta}\mathcal{R}_{lg} - \mathcal{R}_{el}) - \frac{id_{10}\mathcal{E}}{\hbar}(\mathcal{R}_{le} - e^{i\theta}\mathcal{R}_{gl}) + \frac{i|d_{10}|^2}{\hbar R^3}(\mathcal{R}_{lf} - \mathcal{R}_{fl}),$$

$$\left(\frac{d}{dt} + \Gamma\right)\mathcal{R}_{ee} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(e^{-i\theta}\mathcal{R}_{ef} + \mathcal{R}_{el}) + \frac{id_{10}\mathcal{E}}{\hbar}(\mathcal{R}_{le} + e^{i\theta}\mathcal{R}_{fe}),$$

$$\left[\frac{d}{dt} + i(\omega - \omega_0) + \Gamma\right]\mathcal{R}_{gf} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(\mathcal{R}_{gg} - \mathcal{R}_{ff} - e^{-i\theta}\mathcal{R}_{lf}) - \frac{id_{10}\mathcal{E}}{\hbar}e^{i\theta}\mathcal{R}_{ge} + \frac{i|d_{10}|^2}{\hbar R^3}\mathcal{R}_{gl},$$

$$\left[\frac{d}{dt} + i(\omega - \omega_0) + \Gamma\right]\mathcal{R}_{gl} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(e^{-i\theta}\mathcal{R}_{gg} - e^{-i\theta}\mathcal{R}_{ll} - \mathcal{R}_{fl}) - \frac{id_{10}\mathcal{E}}{\hbar}e^{i\theta}\mathcal{R}_{ge} + \frac{i|d_{10}|^2}{\hbar R^3}\mathcal{R}_{gf},$$

$$\left[\frac{d}{dt} + 2i(\omega - \omega_0) + \Gamma\right]\mathcal{R}_{ge} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(\mathcal{R}_{gl} - \mathcal{R}_{fe} + e^{-i\theta}\mathcal{R}_{gf} - e^{-i\theta}\mathcal{R}_{le}),$$

$$\left(\frac{d}{dt} + \Gamma\right)\mathcal{R}_{fl} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}e^{-i\theta}(\mathcal{R}_{fg} - \mathcal{R}_{el}) - \frac{id_{10}\mathcal{E}}{\hbar}(\mathcal{R}_{fe} - \mathcal{R}_{gl}) + \frac{i|d_{10}|^2}{\hbar R^3}(\mathcal{R}_{ff} - \mathcal{R}_{ll}),$$

$$\left[\frac{d}{dt} + i(\omega - \omega_0) + \Gamma\right]\mathcal{R}_{fe} = -\frac{id_{10}^*\mathcal{E}^*}{\hbar}(\mathcal{R}_{fl} + e^{-i\theta}\mathcal{R}_{ff} - e^{-i\theta}\mathcal{R}_{ee}) + \frac{id_{10}\mathcal{E}}{\hbar}\mathcal{R}_{ge} - \frac{i|d_{10}|^2}{\hbar R^3}\mathcal{R}_{le},$$

$$\left[\frac{d}{dt} - i(\omega - \omega_0) + \Gamma\right]\mathcal{R}_{el} = -\frac{id_{10}\mathcal{E}}{\hbar}(\mathcal{R}_{ee} - \mathcal{R}_{ll} - e^{i\theta}\mathcal{R}_{fl}) - \frac{id_{10}^*\mathcal{E}^*}{\hbar}e^{-i\theta}\mathcal{R}_{eg} + \frac{i|d_{10}|^2}{\hbar R^3}\mathcal{R}_{ef}. \quad (10)$$

The matrix elements of the relaxation superoperator representing unidirectional decay by interaction with a common heat bath are expressed as

$$\Gamma\mathcal{R}_{gg} = -\chi|d_{10}|^2(N+1) \times (\mathcal{R}_{ll} + e^{-i\theta}\mathcal{R}_{fl} + e^{i\theta}\mathcal{R}_{lf} + \mathcal{R}_{ff}) + 2\chi|d_{10}|^2N\mathcal{R}_{gg} - \chi(d_{10})^2e^{-i\theta}\mathcal{R}_{ge}M - \chi(d_{10}^*)^2e^{i\theta}\mathcal{R}_{eg}M^*,$$

$$\Gamma\mathcal{R}_{ff} = \frac{\chi|d_{10}|^2}{2}(N+1) \times (2\mathcal{R}_{ff} + e^{-i\theta}\mathcal{R}_{fl} + e^{i\theta}\mathcal{R}_{lf} - 2\mathcal{R}_{ee}) + \frac{\chi|d_{10}|^2}{2}N(e^{-i\theta}\mathcal{R}_{fl} + e^{i\theta}\mathcal{R}_{lf} + 2\mathcal{R}_{ff} - 2\mathcal{R}_{gg}) + \chi d_{10}^2\mathcal{R}_{ge}e^{-i\theta}M + \chi d_{10}^*e^{i\theta}\mathcal{R}_{eg}M^*,$$

$$\Gamma\mathcal{R}_{ll} = \frac{\chi|d_{10}|^2}{2}(N+1) \times (\mathcal{R}_{lf}e^{i\theta} + 2\mathcal{R}_{ll} + e^{-i\theta}\mathcal{R}_{fl} - 2\mathcal{R}_{ee}) + \frac{\chi|d_{10}|^2}{2}N(2\mathcal{R}_{ll} + e^{i\theta}\mathcal{R}_{lf} + e^{-i\theta}\mathcal{R}_{fl} - 2\mathcal{R}_{gg})$$

$$+\chi d_{10}^2 e^{-i\theta}\mathcal{R}_{ge}M + \chi d_{10}^{*}\mathcal{R}_{eg}e^{i\theta}M^{*},$$

$$\Gamma\mathcal{R}_{ee} = 2\chi|d_{10}|^2(N+1)\mathcal{R}_{ee}$$

$$-\chi|d_{10}|^2 N(\mathcal{R}_{ll}+\mathcal{R}_{ff}+e^{-i\theta}\mathcal{R}_{fl}+e^{i\theta}\mathcal{R}_{lf})$$

$$-\chi(d_{10})^2 e^{-i\theta}\mathcal{R}_{ge}M - \chi(d_{10}^{*})^2\mathcal{R}_{eg}e^{i\theta}M^{*},$$

$$\Gamma\mathcal{R}_{gf} = \frac{\chi|d_{10}|^2}{2}(N+1)$$

$$\times(\mathcal{R}_{gf}+e^{-i\theta}\mathcal{R}_{gl}-2\mathcal{R}_{fe}-2e^{i\theta}\mathcal{R}_{le})$$

$$+\frac{\chi|d_{10}|^2}{2}N(e^{-i\theta}\mathcal{R}_{gl}+3\mathcal{R}_{gf})$$

$$+\chi d_{10}^{*2}(\mathcal{R}_{fg}+e^{i\theta}\mathcal{R}_{lg}-e^{i\theta}\mathcal{R}_{ef})M^{*},$$

$$\Gamma\mathcal{R}_{gl} = \frac{\chi|d_{10}|^2}{2}(N+1) \qquad (11)$$

$$\times(e^{i\theta}\mathcal{R}_{gf}+\mathcal{R}_{gl}-2\mathcal{R}_{fe}-2e^{i\theta}\mathcal{R}_{le})$$

$$+\frac{\chi|d_{10}|^2}{2}N(3\mathcal{R}_{gl}+e^{i\theta}\mathcal{R}_{gf})$$

$$+\chi d_{10}^{*2}(\mathcal{R}_{fg}e^{i\theta}+\mathcal{R}_{lg}-e^{i\theta}\mathcal{R}_{el})M^{*},$$

$$\Gamma\mathcal{R}_{ge} = \chi|d_{10}|^2(N+1)\mathcal{R}_{gg}+\chi|d_{10}|^2 N\mathcal{R}_{ge}$$

$$+\chi d_{10}^{*2}(e^{i\theta}\mathcal{R}_{ll}+e^{i\theta}\mathcal{R}_{ff}+\mathcal{R}_{fl}$$

$$+e^{2i\theta}\mathcal{R}_{lf}-\mathcal{R}_{gg}e^{i\theta}-e^{i\theta}\mathcal{R}_{ee})M^{*},$$

$$\Gamma\mathcal{R}_{fl} = \frac{\chi|d_{10}|^2}{2}(N+1)$$

$$\times(e^{i\theta}\mathcal{R}_{ff}+2\mathcal{R}_{fl}+e^{i\theta}\mathcal{R}_{ll}-2e^{i\theta}\mathcal{R}_{ee})$$

$$+\frac{\chi|d_{10}|^2}{2}N(2\mathcal{R}_{fl}+e^{i\theta}\mathcal{R}_{ff}+e^{i\theta}\mathcal{R}_{ll}-2e^{i\theta}\mathcal{R}_{gg})$$

$$+\chi d_{10}^2\mathcal{R}_{ge}M + \chi d_{10}^{*2}e^{2i\theta}\mathcal{R}_{eg}M^{*},$$

$$\Gamma\mathcal{R}_{fe} = \frac{\chi|d_{10}|^2}{2}(N+1)(3\mathcal{R}_{fe}+e^{i\theta}\mathcal{R}_{le})$$

$$+\frac{\chi|d_{10}|^2}{2}N(e^{i\theta}\mathcal{R}_{le}+\mathcal{R}_{fe}-2\mathcal{R}_{gl}-2e^{i\theta}\mathcal{R}_{gf})$$

$$+\chi d_{10}^{*2}(e^{i\theta}\mathcal{R}_{el}+e^{2i\theta}\mathcal{R}_{ef}-e^{i\theta}\mathcal{R}_{fg})M^{*},$$

$$\Gamma\mathcal{R}_{le} = \frac{\chi|d_{10}|^2}{2}(N+1)(3\mathcal{R}_{le}+e^{-i\theta}\mathcal{R}_{fe})$$

$$+\frac{\chi|d_{10}|^2}{2}N(\mathcal{R}_{le}+e^{-i\theta}\mathcal{R}_{fe}-2e^{-i\theta}\mathcal{R}_{gl}-2\mathcal{R}_{gf})$$

$$+\chi d_{10}^{*2}(\mathcal{R}_{el}+e^{i\theta}\mathcal{R}_{ef}-e^{i\theta}\mathcal{R}_{lg})M^{*},$$

where time-dependent squeezing is characterized by

$$M = |M|\exp(-2i\Omega_{\Gamma}t + 2i\omega t).$$

We assume that the atomic transition frequency, driving-field carrier frequency, and the central frequency of the heat bath are equal, so that M is independent of time.

3. ENTANGLEMENT BY INTERACTION WITH A HEAT BATH AT ZERO TEMPERATURE

First, we consider the simplest case of unidirectional collective decay in the case of $N = M = 0$, $\hat{\Gamma}_0' = \hat{\Gamma}_1 = \hat{\Gamma}_2 = 0$, and $\theta = 2\pi n$ ($n = 0, \pm 1, \ldots$).

Let us discuss the dependence of entanglement on the quantities $\beta = \chi|d_{10}|^2/2$ and $\delta = |d_{10}|^2/R^3\hbar$ characterizing interaction with the heat bath and dipole–dipole coupling, respectively, and the Rabi frequency $\xi = |\mathcal{E}d_{10}|/\hbar$.

When the only effective entanglement mechanism is the interaction with the common heat bath, steady-state quantum correlations are observed and the degree of entanglement can amount to –0.17, depending on initial conditions [9].

When only the dipole–dipole interaction is effective, no steady-state quantum correlations are created, because the degree of entanglement oscillates between zero and –0.5. The classical field does not create any entanglement at all, and the corresponding eigenvalues of the Peres–Horodecki matrix are [1, 0, 0, 0] at any instant.

Any combination of these mechanisms leads to a different entanglement behavior. However, all of the resulting steady states depend on initial conditions, as in the case of collective decay by interaction with a common heat bath.

We introduce dimensionless parameters characterizing the relative magnitudes of dipole–dipole coupling and Rabi frequency as compared to interaction with the heat bath:

$$\kappa = \delta/\beta = 2/R^3\hbar\chi,$$
$$c = \xi/\beta = 2|\mathcal{E}|/\hbar\chi|d_{10}|.$$

Dipole–dipole interaction in a two-atom system coupled to a common heat bath either does not affect steady-state entanglement or destroys it. In the case of a doubly excited (Dicke) initial state, no entanglement is created in the system. In the case of a singly excited state,

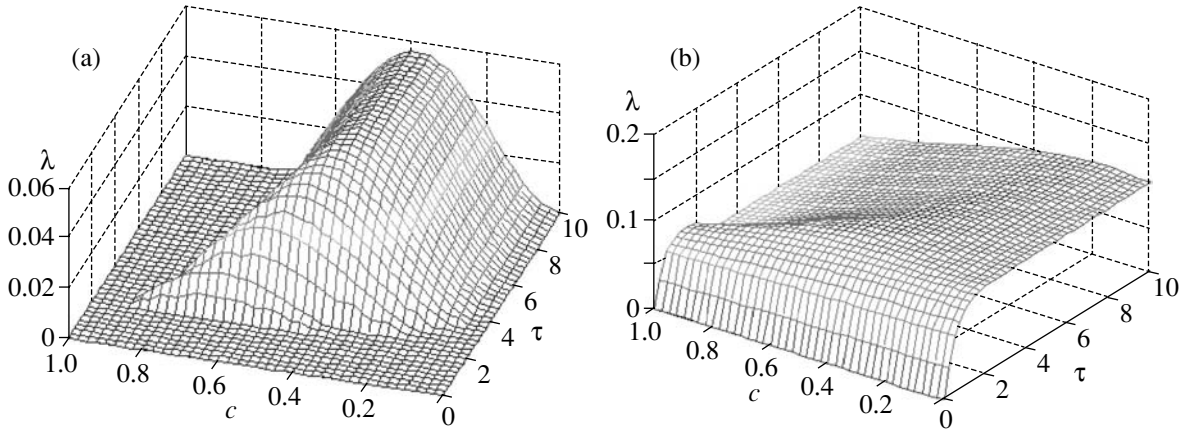

Fig. 1. Time dependence of the lowest eigenvalue of the Peres–Horodecki matrix parameterized by the ratio of the Rabi frequency to the constant of coupling to a common heat bath with zero photon number in the absence of dipole–dipole interaction: (a) doubly excited system; (b) singly excited system.

steady-state quantum correlations are created, and the lowest eigenvalue of the Peres–Horodecki matrix,

$$\lambda = \frac{1}{4}[1 - e^{-2\beta t} - \sqrt{2(1 - e^{-2\beta t} + e^{-4\beta t} - e^{-2\beta t}\cos\kappa t)}],$$

tends to a limit value of –0.104. Since this limit is independent of the dipole–dipole interaction strength, the corresponding steady-state entanglement is similar to the entanglement of noninteracting atoms by interaction with a common heat bath [9]. This is clear from the fact that the dipole–dipole interaction does not mix subspaces of states with different particle permutation symmetries if $\theta = 2\pi n$ ($n = 0, \pm 1, \ldots$). When the phase gain is different from these values, the dipole–dipole interaction mixes such subspaces and can destroy steady-state entanglement if κ is sufficiently large.

A totally different behavior is observed when the atoms are driven by a resonant classical field. In this case, steady-state entanglement is created when the system is initially in the doubly excited state, ground state, or singly excited state. The dependence of entanglement on the initial state of the system is illustrated by Figs. 1a and 1b, which correspond to doubly and singly excited states, respectively.

As the classical-field strength increases, the development of quantum correlations is followed by their destruction; i.e., there exists an optimal field strength for maximal entanglement. Note also that entanglement of this kind is enhanced by increasing dipole–dipole interaction.

Both the operator of interaction with a classical field and the dipole–dipole interaction operator are symmetric under particle permutation in the case of $\theta = 2\pi n$ ($n = 0, \pm 1, \ldots$) considered here. However, the interaction with a classical field leads to different types of steady-state entanglement, since the corresponding operator is reducible; i.e., the state space of the system is decomposed into irreducible subspaces. Due to the difference in dynamics between the subspaces, both time-dependent and steady-state entanglements are created by change in the representation of any state of the system in an entangled or disentangled basis set. In general, each particular operator contributing to the system's dynamics is associated with a specific decomposition into irreducible subspaces, whose properties determine the dynamics of entanglement.

In the case of a singly excited state (Fig. 1b), the interaction with a classical field destroys steady-state entanglement, while the dipole–dipole interaction only induces oscillations, while its effect on steady-state quantum correlations is negligible in a wide range of parameters.

If the three interactions are effective simultaneously, then the nonzero elements of the time-independent density matrix are

$$R_{gg} = (1 + 2c^2 + 4c^4 + k^2)\Delta^{-1},$$

$$R_{gs} = c\sqrt{2}(k - 2ic^2 - i)\Delta^{-1},$$

$$R_{ge} = -2c^2(1 + ik)\Delta^{-1},$$

$$R_{ss} = 2c^2(2c^2 + 1)\Delta^{-1}, \quad R_{se} = -2ic^3\sqrt{2}\Delta^{-1}, \quad (12)$$

$$R_{ee} = 4c^4\Delta^{-1}, \quad R_{aa} = R_{aa},$$

$$\Delta^{-1} = (1 - R_{aa})(1 + 4c^2 + 12c^4 + k^2)^{-1}.$$

This general solution is reduced to a particularly simple form by retaining only the first-order terms with respect to the constant of coupling to the classical field (in the partially entangled basis):

$$\mathcal{R} = \begin{pmatrix} 1-\mathcal{R}_{aa} & 0 & \frac{\kappa - i}{\kappa^2+1}c\sqrt{2}(1-\mathcal{R}_{aa}) & 0 \\ 0 & \mathcal{R}_{aa} & 0 & 0 \\ \frac{\kappa + i}{\kappa^2+1}c\sqrt{2}(1-\mathcal{R}_{aa}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The corresponding Peres–Horodecki matrix is

$$\mathcal{R}_{st}^{P-H} = \begin{pmatrix} 1-\mathcal{R}_{aa} & \frac{\kappa + i}{\kappa^2+1}c(1-\mathcal{R}_{aa}) & \frac{\kappa - i}{\kappa^2+1}c(1-\mathcal{R}_{aa}) & \frac{1}{2}\mathcal{R}_{aa} \\ \frac{\kappa - i}{\kappa^2+1}c(1-\mathcal{R}_{aa}) & \frac{1}{2}\mathcal{R}_{aa} & 0 & 0 \\ \frac{\kappa + i}{\kappa^2+1}c(1-\mathcal{R}_{aa}) & 0 & \frac{1}{2}\mathcal{R}_{aa} & 0 \\ \frac{1}{2}\mathcal{R}_{aa} & 0 & 0 & 0 \end{pmatrix}. \quad (13)$$

In these expressions, the time-independent element $\mathcal{R}_{aa}$ of the density matrix is determined by the initial conditions and the coupling constants. The corresponding expression is very cumbersome. For this reason, our analysis presented above is restricted to numerical simulations.

If the dipole–dipole interaction is neglected, then the lowest eigenvalue of the Peres–Horodecki matrix has the simple form

$$\lambda = \frac{1}{2}\frac{R_{aa} + 6R_{aa}c^2 + 8R_{aa}c^4 - 2c^2 + 4c^4}{1 + 4c^2 + 12c^4}. \quad (14)$$

Hence, when $\mathcal{R}_{aa} = 0$, the highest degree of squeezing is characterized by $\lambda \approx -0.06$, which corresponds to $c = \xi/\beta = 2|\mathcal{E}|/\hbar\chi|d_{10}| = 0.38$.

4. MAXIMALLY SQUEEZED HEAT BATH

Now, we consider the most interesting case, when the common heat bath is in the maximally squeezed state at a finite temperature ($|M| = \sqrt{N(N+1)}$), retaining the condition $\theta = 2\pi n$ ($n = 0, \pm 1, \ldots$). In this case, the evolution of initially pure states is characterized by a two-atom wavefunction.

The Peres–Horodecki matrix is

$$\mathcal{R}_{st}^{P-H} = \begin{pmatrix} \frac{N+1}{2N+1}(1-\mathcal{R}_{aa}) & 0 & 0 & -\frac{1}{2}\mathcal{R}_{aa} \\ 0 & \frac{1}{2}\mathcal{R}_{aa} & -\frac{\sqrt{N(N+1)}}{2N+1}(1-\mathcal{R}_{aa}) & 0 \\ 0 & -\frac{\sqrt{N(N+1)}}{2N+1}(1-\mathcal{R}_{aa}) & \frac{1}{2}\mathcal{R}_{aa} & 0 \\ -\frac{1}{2}\mathcal{R}_{aa} & 0 & 0 & \frac{N}{2N+1}(1-\mathcal{R}_{aa}) \end{pmatrix}.$$

When $\mathcal{R}_{aa} = 0$, the atoms are in identical (excited or ground) initial states. In the absence of dipole–dipole interaction, the only negative eigenvalue of the Peres–Horodecki matrix is $\lambda = -\sqrt{N(N+1)}/(2N+1)$, which tends to its lowest limit, $\lambda = -0.5$, with increasing photon number.

Thus, the collective interaction between two excited or ground states and a maximally squeezed high-intensity broad-bandwidth electromagnetic wave gives rise to steady-state entanglement analogous to that in an EPR pair. This is explained by the high degree of anisotropy of angular-momentum relaxation in a

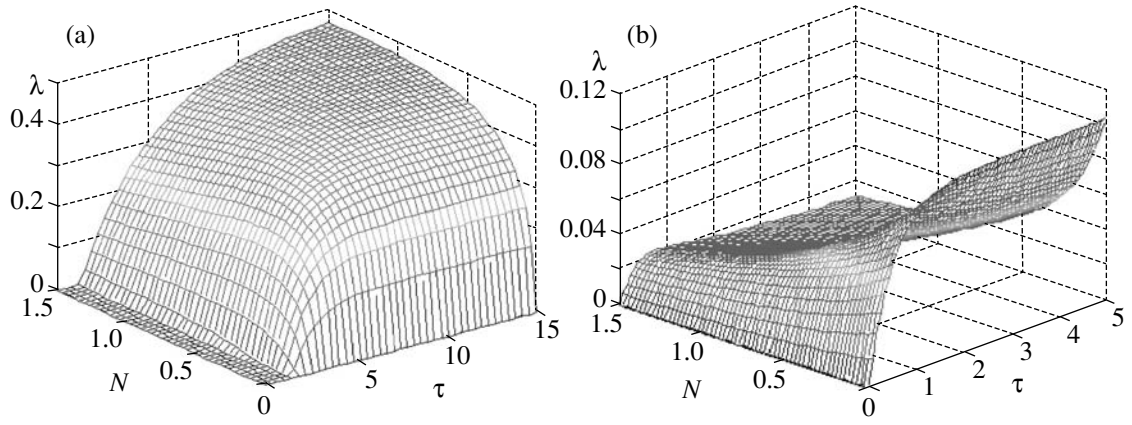

Fig. 2. Time dependence of the lowest eigenvalue of the Peres–Horodecki matrix parameterized by thermal photon number for a squeezed heat bath: (a) doubly excited system; (b) singly excited system.

squeezed heat bath (see Appendix 1 in [29]), which can be interpreted as transfer of correlations from the squeezed heat bath to angular-momentum states. The relationship between entanglement and spin squeezing was discussed in [30]. Since the evolution of Dicke states (symmetrized under permutation of the states of a two-level system) can be described in terms of angular-momentum dynamics [14, 15], the steady state coupled to a maximally squeezed heat bath is the following pure one (see [31]):

$$\frac{\sqrt{N+1}|g\rangle + \sqrt{N}|e\rangle}{\sqrt{2N+1}}.$$

Figure 2a illustrates the dependence of the time-dependent lowest eigenvalue of the Peres–Horodecki matrix on the thermal photon number for a doubly excited two-atom system. Here, the dipole–dipole interaction has no effect on the steady-state entanglement.

Thus, since initially identical atomic states interacting with a maximally squeezed heat bath evolve toward an entangled state analogous to an EPR pair, any degree of entanglement can be obtained if the system is controlled so that entanglement is not destroyed. The desired control is a classical driving field. The resonantly driven system evolves into a mixed steady state, and the corresponding time-independent density matrix is

$$\begin{aligned}
R_{gg} &= \Delta^{-1}(1-R_{aa}) + (8N^2 + 5N - 2c^2N \\
&+ 4N^3 + 8c^4|M| - 6|M|N - 4N^2|M| + 8c^4N \\
&+ 4c^4 - 2|M| - 8c^2|M| + 1 + u), \\
R_{gs} &= -i\sqrt{2}c\Delta^{-1}(1-R_{aa})(u + N - |M| + 1), \\
R_{ge} &= -\Delta^{-1}(1-R_{aa})(u - 2c^2|M| - 4N^2|M| \\
&- 4N|M| + 4N^3 + 6N^2 + 2N - |M|), \\
R_{ss} &= 2c^2\Delta^{-1}(1-R_{aa})(2N + 1 + u), \\
R_{se} &= -\sqrt{2}c\Delta^{-1}(1-R_{aa})(u + N - |M|), \qquad (15) \\
R_{ee} &= \Delta^{-1}(1-R_{aa}) \\
&\times (4N^2 + N - 4N^2|M| + 2c^2N + 4N^3 \\
&- 4c^2|M| + 8c^4N - 2|M|N + 8c^4|M| + 4c^4), \\
R_{ga} &= R_{as} = R_{ea} = 0, \\
u &= 4c^2|M| + 4c^2N + 2c^2, \\
\Delta &= 1 + 12N^2 + 6N - 8N^2|M| + 24c^4|M| \\
&- 8N|M| - 8c^2|M| + 8c^2N + 8N^3 - 2|M| \\
&+ 4c^2 + 24c^4N + 12c^4.
\end{aligned}$$

An analysis of the lowest eigenvalue of the Peres–Horodecki matrix shows that no entanglement is created in the system by strong or weak classical field if $\mathcal{R}_{aa} = 0$. Figure 3 also demonstrates that classical driving field generally destroys steady-state entanglement. However, the degree of steady-state entanglement varies between zero and that in an EPR pair within a certain range of classical-field strength. This observation is most important in terms of coherent control, because it implies that any required degree of entanglement can be created in a resonantly driven system, depending on the strength of coupling to the classical field.

In the case of a singly excited system, the degree of entanglement decreases with increasing thermal photon number (see Fig. 2b). When driven by a classical field, the system exhibits Rabi oscillations and the degree of

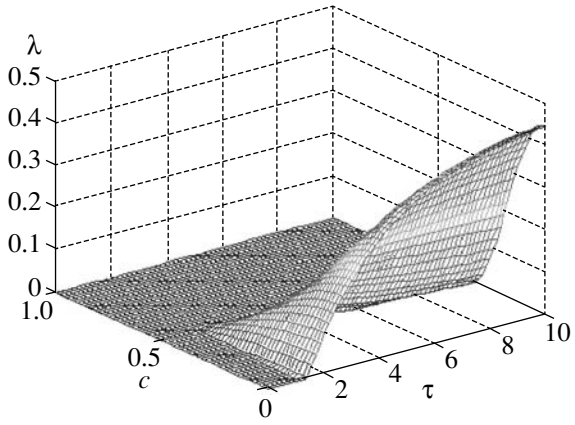

Fig. 3. Time dependence of the lowest eigenvalue of the Peres–Horodecki matrix parameterized by the ratio of the Rabi frequency to the constant of coupling to a maximally squeezed common heat bath in the case of a doubly excited system.

steady-state entanglement is independent of both photon number and dipole–dipole interaction.

5. ENTANGLEMENT BY INTERACTION WITH AN ARBITRARY HEAT BATH

Let us now discuss the dependence of entanglement on detuning from optimal parameters. Practical implementation of maximal squeezing presents certain difficulties in the case of broad-bandwidth radiation. This leads to a certain difference between the actual and highest possible degrees of entanglement. The deviation from maximal squeezing reduces the degrees of both time-dependent and steady-state entanglement, as illustrated by Fig. 4.

Another important mechanism responsible for deviation from optimal steady-state entanglement in the present model is the presence of additional decay channels, which can be modeled by introducing interactions with independent heat baths into Eq. (6). For simplicity, we assume that the independent heat baths are not squeezed, have equal parameters, and do not contain any photons. When the dipole–dipole interaction between the atoms is neglected, the nonzero elements of the time-independent density matrix obtained by solving Eq. (6) (in the partially entangled basis):

$$\begin{aligned} R_{gg} &= [w + 8c^2(2c^2 + 1 + 2b + b^2)]\Delta^{-1}, \\ R_{ss} &= 8c^2(2c^2 + 1 + 2b + b^2), \\ R_{aa} &= R_{ee} = 16c^4\Delta^{-1}, \\ R_{gs} &= -2i\sqrt{2}c(4b^2 + b^3 + 2 + 5b + 4c^2 + 4bc^2)\Delta^{-1}, \\ R_{ge} &= -4c^2(b^2 + 3b + 2)\Delta^{-1}, \\ R_{se} &= -8i(1 + b)\sqrt{2}c^3. \end{aligned} \tag{16}$$

Here, c denotes the ratio of the frequency $\xi = |\mathcal{E}d_{10}|/\hbar$ characterizing the interaction between an atom and a classical field to the constant $\beta = \chi|d_{10}|^2/2$ of coupling to an independent heat bath, $c = \xi/\beta = 2|\mathcal{E}|/\hbar\chi|d_{10}|$ and $b = \chi_i/\chi$ is the ratio of the constants of coupling to the independent and common heat baths. The remaining parameters are defined as follows:

$$\Delta = 64c^4 + 16c^2(2c^2 + 1 + 2b + b^2) + w,$$

$$w = b^4 + 6b^3 + 13b^2 + 12b + 4.$$

The interactions with independent heat baths destroy the entanglement created by a classical wave in a system coupled to a common heat bath. Figure 5 illus-

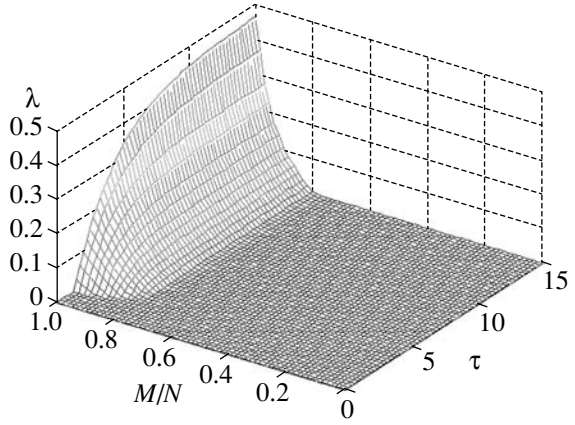

Fig. 4. Time dependence of the lowest eigenvalue of the Peres–Horodecki matrix parameterized by the degree of squeezing for a common heat bath in the case of a doubly excited system.

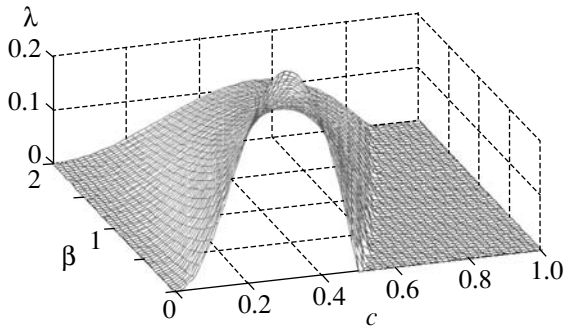

Fig. 5. The lowest eigenvalue of the time-independent Peres–Horodecki matrix. Here, β and c are, respectively, the ratios of the constant of coupling to independent heat baths and the Rabi frequency to the constant of coupling to a common heat bath with zero photon number in the case of a doubly excited system.

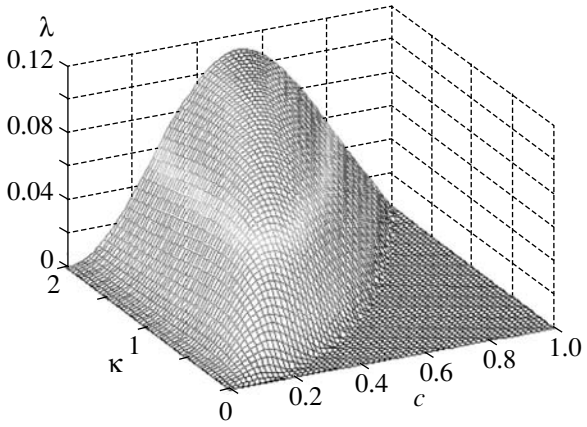

Fig. 6. The lowest eigenvalue of the time-independent Peres–Horodecki matrix. Here, κ and *c* are, respectively the ratios of the dipole–dipole interaction strength and the Rabi frequency to the constant of coupling to independent heat baths with zero photon numbers in the case of a doubly excited system.

trates the entanglement behavior under these conditions.

Finally, we consider the limit case when the atoms are coupled only to independent heat baths. Then, time-dependent entanglement can be created only by dipole–dipole interaction (under appropriate initial conditions), whereas no steady-state entanglement is created in the absence of resonant classical field. The corresponding diagonal time-independent density matrix,

$$R_{gg} = \frac{(N+1)^2}{(2N+1)^2}, \quad R_{ff} = R_{ll} = \frac{(N+1)N}{(2N+1)^2},$$

$$R_{ee} = \frac{N^2}{(2N+1)^2},$$

is independent of dipole–dipole interaction and possible squeezing by independent heat baths. Here, N denotes the spectral density of an independent heat bath.

For a system driven by a resonant classical wave, the time-independent solution describes an entangled state depending on the dipole–dipole interaction strength:

$$R_{gg} = (16c^4 + 4k^2 + 8c^2 + 1)\Delta^{-1},$$

$$R_{gf} = 2c(4ic^2 + 2k + i)\Delta^{-1},$$

$$R_{gl} = 2c(-4ic^2 + 2k - i)\Delta^{-1}, \quad R_{ge} = 4c^2\Delta^{-1},$$

$$R_{ff} = R_{ll} = 4c^2(4c^2 + 1)\Delta^{-1}, \qquad (17)$$

$$R_{ee} = 16c^4\Delta^{-1}, \quad R_{fl} = -4c^2(2ik + 1)\Delta^{-1},$$

$$R_{fe} = -R_{le} = -8ic^3\Delta^{-1},$$

$$\Delta = 64c^4 + 4k^2 + 16c^2 + 1.$$

Here, the photon number in the independent heat baths is set to zero for simplicity. Figure 6 illustrates the entanglement behavior under these conditions.

6. CONCLUSIONS

The entanglement mechanism considered here is very difficult to analyze, because it is determined by at least eight groups of variable parameters: the photon number and degree of squeezing of the common heat bath, analogous parameters of independent heat baths, the constants of coupling to the heat baths, the dipole–dipole coupling parameters, the Rabi frequency, the detuning from resonance with the classical field, the detunings from the central frequencies of the heat baths, and the phase gain. Moreover, the steady state of a two-atom system coupled to a common heat bath depends on the initial states of the atoms. Our discussion and analysis are focused on situations that, on the one hand, highlight the specific features of the new entanglement mechanism and, on the other hand, demonstrate the possibility of entanglement control by variation of the classical-field strength. Since a quantitative criterion for entanglement is obtained by solving fourth-order algebraic equations, analytical description of entanglement cannot be developed in most cases, and we had to rely on numerical simulations.

One special feature of the discussed mechanism of development of quantum correlations is that steady-state entanglement is independent of the direct dipole–dipole interaction when the system is initially in a singly excited state, being determined primarily by the characteristics of the common heat bath and the Rabi frequency (for $\theta = 2\pi n$, where $n = 0, \pm 1, \ldots$). The steady-state entanglement behavior changes only when the decay via independent interactions with specific heat baths plays a substantial role in the dynamics of each atom. In the absence of collective decay, the entanglement is mainly determined by the dipole–dipole interaction and the Rabi frequency. Basically, this is explained by the properties of the irreducible representations of the total dipole moment operator and the operator of dipole–dipole interaction between the atoms. Furthermore, this observation suggests two essentially different mechanisms of entanglement between atoms driven by a resonant classical electromagnetic wave. One of these is based on collective decay by interaction with a common heat bath; the other, on direct (in the present case, dipole–dipole) interaction between the subsystems.

The numerous studies of entanglement of atomic states coupled to a common heat bath reported in recent papers (e.g., see [12, 30, 32, 33]) included analyses of entanglement of atoms driven by a classical electromagnetic field. In [34], a Dicke-type model was invoked to analyze the steady-state entanglement of a coherently driven system coupled to a common heat bath in the absence of both dipole–dipole interaction and squeezing, and use was made of the Wootters

entanglement measure (concurrence). Our results agree with those obtained in [34] when restricted to Dicke states and $\mathcal{R}_{aa} = 0$. However, entanglement in the case of an unsqueezed common heat bath is due specifically to $\mathcal{R}_{aa} \neq 0$. Moreover, general analysis of entanglement of symmetrized states of indistinguishable particles (including Dicke states) leads to ambiguous results [35] and must be taken further. The relationship between entanglement of Dicke states and atomic-state squeezing was considered in [30]. In a recent paper [36], the entanglement created by a classical driving field was analyzed for spontaneously decaying atomic systems coupled by dipole–dipole interaction. The results reported therein are consistent with those obtained here in the absence of coupling to a common heat bath. Note, however, that the present analysis of entanglement is performed by taking into account both symmetrized and antisymmetrized (with respect to particle permutation) quantum states.

ACKNOWLEDGMENTS

We thank V.N. Gorbachev for helpful discussions.

REFERENCES

1. A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
2. E. Schrödinger, Naturwissenschaften **23**, 807, 823, 844 (1935).
3. *Physics of the Quantum Information*, Ed. by D. Bouwmeester, A. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
4. G. Alber, T. Beth, M. Horodecki, P. Horodecki, R. Horodecki, M. Rotteler, H. Weinfurter, R. Werner, and A. Zeilinger, *Quantum Information. An Introduction to Basic Theoretical Concepts and Experiments* (Springer, Berlin, 2001).
5. V. N. Gorbachev, A. I. Zhiliba, A. A. Rodichkina, and A. I. Trubilko, Phys. Lett. A **323**, 339 (2004); quant-ph/0402147.
6. B. Julsgaard, A. Kozhekin, and E. S. Polzik, Nature **413**, 400 (2001); quant-ph/0106057.
7. H. J. Briegel and R. Raussendorf, Phys. Rev. Lett. **86**, 5188 (2001).
8. H. J. Kimble and S. J. Van Enk, Nature **429**, 712 (2004).
9. A. M. Basharov, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 151 (2002) [JETP Lett. **75**, 123 (2002)]; Zh. Éksp. Teor. Fiz. **121**, 1249 (2002) [JETP **94**, 1070 (2002)].
10. L. Jak'obczyk, J. Phys. A **35**, 6383 (2002).
11. H. T. Dung, S. Scheel, D.-G. Welsch, and L. Knoll, J. Opt. B: Quantum Semiclassic. Opt. **4**, 169 (2002).
12. R. Tanas and Z. Ficek, J. Opt. B: Quantum Semiclassic. Opt. **6**, S90 (2004); quant-ph/0309195.
13. D. A. Lidar, I. L. Chuang, and K. B. Whaley, Phys. Rev. Lett. **81**, 2594 (1998).
14. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988) [in Russian].
15. M. G. Benedict, A. M. Ermolaev, V. A. Malyshev, I. V. Sokolov, and E. D. Trifonov, *Superradiance: Multiatomic Coherent Emission* (IOP, Bristol and Philadelphia, 1996).
16. R. G. DeVoe and R. G. Brewer, Phys. Rev. Lett. **76**, 2049 (1996).
17. J. Eschner, Ch. Raab, F. Schmidt-Kaler, and R. Blatt, Nature **413**, 495 (2001).
18. A. A. Makarov and V. S. Letokhov, Zh. Éksp. Teor. Fiz. **124**, 766 (2003) [JETP **97**, 688 (2003)].
19. A. Barchielli and V. P. Belavkin, J. Phys. A **24**, 1495 (1991).
20. A. M. Basharov and É. A. Manykin, Opt. Spektrosk. **96**, 91 (2004) [Opt. Spectrosc. **96**, 81 (2004)].
21. J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).
22. W. K. Wootters, Phys. Rev. Lett. **80**, 2245 (1998); C. H. Bennett, D. P. Di Vincenco, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
23. A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
24. M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).
25. A. M. Basharov, E. V. Moreva, and É. A. Manykin, Opt. Spektrosk. **96**, 724 (2004) [Opt. Spectrosc. **96**, 658 (2004)]; A. M. Basharov and A. A. Bashkeev, Opt. Spektrosk. **96**, 716 (2004) [Opt. Spectrosc. **96**, 650 (2004)]; A. M. Basharov, A. A. Bashkeev, and É. A. Manykin, in *Proceedings of 10th International Conference on Quantum Optics (ICQO-2004)* (Minsk, 2004), p. 43.
26. M. J. Collet and C. W. Gardiner, Phys. Rev. A **30**, 1386 (1984).
27. G. Kurizki and A. Ben-Reuven, Phys. Rev. A **36**, 90 (1987).
28. G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
29. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer Academic, Dordrecht, 1999).
30. A. Messikh, Z. Ficek, and M. R. B. Wahiddin, J. Opt. B: Quantum Semiclassic. Opt. **5**, L1 (2003).
31. G. M. Palma and P. L. Knight, Phys. Rev. A **39**, 1962 (1989).
32. F. Benatti, R. Floreanini, and M. Piani, Phys. Rev. Lett. **91**, 070402 (2003).
33. A. Messikh, M. R. B. Wahiddin, C. H. Pah, and Z. Ficek, J. Opt. B: Quantum Semiclassic. Opt. **6**, 289 (2004).
34. S. Schneider and G. J. Milburn, Phys. Rev. A **65**, 042107 (2002).
35. G. C. Ghirardi and L. Marinatto, quant-ph/0401065.
36. O. Cakir, A. A. Kliachko, and A. S. Shumovsky, quant-ph/0406081.

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 487–504.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 551–569.

ATOMS, MOLECULES, OPTICS

Probabilities of Radiative Transitions between Stark States in Orthohelium

A. A. Kamenskiĭ and V. D. Ovsyannikov*
Physics Department, Voronezh State University, Voronezh, 394006 Russia
**e-mail: ovd@phys.vsu.ru*
Received September 13, 2004

Abstract—The wavefunctions, matrix elements, and probabilities of radiative transitions between Stark sublevels of atomic multiplets are calculated as a function of the strength of a dc electric field. The general expressions for the wavefunction of a multiplet state in a field obtained by perturbation theory for close-lying levels with the use of completely reduced Green's function allow one to determine the field dependence of both the dipole-allowed and dipole-forbidden radiation amplitudes. A decomposition of the second-order amplitude for the transition between fine-structure sublevels of two levels of equal parity into irreducible components is obtained. Numerical calculations of the probabilities of radiative transitions between triplet states of helium show the possibility of experimental observation of the emergence and vanishing of Stark lines of radiative transitions in the vicinity of anticrossing fields. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Radiative transitions between bound states are the basic processes that characterize the quantum structure of an atom. The probabilities of these transitions determine the possibility of observing and measuring the corresponding intra-atomic characteristics. The probabilities of radiative transitions change under an external field. A quantitative description of this phenomenon allows one not only to predict the results of experimental investigations but also to control the process of radiation by a dc field, which gives rise to new (forbidden) lines in the emission and absorption spectra and leads to the vanishing of a number of lines that exist in the spectrum of a free atom.

A helium atom is of special interest because it represents the simplest, after a hydrogen atom, quantum system that consists of three particles. A specific feature of the Stark effect for a helium atom is the anticrossing of levels with equal nonzero angular momenta l of an excited electron (equal to the total angular momentum of the atom $l = L$). The anticrossing may occur both between the singlet $1snl^1L_J$ (a total spin of $S = 0$ and a total momentum of $J = L$) and the triplet $1snl^3L_J$ (a total spin of $S = 1$) sublevels [1, 2], as well as between sublevels of a triplet [3] that differ by the value of the total momentum $J = L, L \pm 1$ ($\mathbf{J} = \mathbf{L} + \mathbf{S}$).

The anticrossing of levels in a dc electric field $\mathbf{F}_0 = F_0\mathbf{e}_0$ is caused by the divergence (repulsion) of sublevels as F_0 increases in a strong field $F_0 > F_a$; the difference between the polarizabilities of the lower and upper sublevels is such that, for $F_0 < F_a$, these sublevels approach each other as F_0 increases. At the anticrossing point $F_0 = F_a$, which corresponds to maximally close sublevels, the mixing of the initial states leads to a significant change in the properties of diverging sublevels. In particular, an admixture state with a different momentum J may remove the constraint on the dipole radiation or, conversely, may nullify the radiation matrix element due to the destructive superposition of nonzero matrix elements of dipole-allowed transitions. Thus, in addition to the shift and the splitting of levels, a dc electric field may lead either to the emergence of forbidden or to the vanishing of allowed lines in the atomic spectrum.

Since the discovery of the Stark effect, extensive experimental research has been devoted to studying the behavior of an atomic spectrum, including studying the variation in the intensity of atomic lines in an electric field (see the survey in [4]). However, a consistent theoretical analysis of the wavefunctions and the probabilities and intensities of radiative transitions in atoms as a function of a dc electric field has been carried out only for a hydrogen atom: (i) for highly excited Rydberg states with principal quantum numbers of $n = 10$, 30, 40 [5] by diagonalizing the Hamiltonian of the interaction of an atom with the field in the restricted basis of states of n' shells with $|n' - n| \le 3$ and (ii) for arbitrary states by perturbation theory with the use of the reduced Coulomb Green function both in the first nonvanishing terms [6] and in higher order terms in the field strength F_0 [7]. The calculation of quadrupole corrections to the probabilities of radiative transitions in a hydrogen-like atom has shown that the contribution of a multipole interaction with radiation is comparable to the effects of the spin–orbit interaction (fine structure) and becomes negligible compared with the Stark correc-

tions when the energy of interaction with a dc field is greater than the energy of the spin–orbit interaction [8].

Theoretical calculations of the Stark effect in multielectron atoms have been primarily restricted to the determination of the shift and the splitting of energy levels as a function of the electric-field strength. A consistent description of the field dependence of the energy at which the anticrossing of the n^3P_J triplet sublevels of helium occurs has been performed by perturbation theory for close-lying levels on the basis of high-precision relativistic calculations of polarizabilities [9], as well as with regard to the fourth-order corrections (expressed in terms of hyperpolarizabilities) both to the diagonal [10] and off-diagonal [11] matrix elements. Higher order perturbation theory for almost degenerate states developed in this way allows one to calculate not only the eigenvalues but also the eigenfunctions of an atom in a field, which are primarily required to determine the intensities of multiplet emission and absorption lines of electromagnetic waves by atoms [12] in a dc electric field. In the present paper, we apply the method developed in [11] to calculate the wavefunctions, matrix elements, and probabilities of radiative transitions between the triplet states of helium.

2. THE WAVEFUNCTION OF A MULTIPLET STATE IN A DC ELECTRIC FIELD

The integral Schrödinger equation

$$\Psi(\mathbf{r}) = \sum_{i=1}^{k} a_{J_iM}\phi_{nLJ_iM}(\mathbf{r}) - G'_{\bar{E}}(\mathbf{r},\mathbf{r}')(\hat{V}(\mathbf{r}')-\Delta E)|\Psi(\mathbf{r}')\rangle, \quad (1)$$

expressed in terms of the completely reduced Green function

$$G'_{\bar{E}}(\mathbf{r},\mathbf{r}') = G_{\bar{E}}(\mathbf{r},\mathbf{r}') - \sum_{i=1}^{k}\frac{\phi_{nLJ_iM}(\mathbf{r})\phi^*_{nLJ_iM}(\mathbf{r}')}{E_{nLJ_i}-\bar{E}}, \quad (2)$$

with the mean energy of an atomic multiplet

$$\bar{E} = \sum_{i=1}^{k}\frac{E_{nLJ_i}}{k}$$

(k is the number of interacting sublevels of the multiplet with energies E_{nLJ_i}) for the wavefunction Ψ, which becomes the wavefunction $\phi_{nLJM}(\mathbf{r})$ of one of the sublevels when the field is switched off, can be reduced to a set of homogeneous algebraic equations for the superposition coefficients $a_{JM} = \langle\phi_{nLJM}|\Psi\rangle$ [11]:

$$\sum_{i=1}^{k} a_{J_iM}[W_{JJ_i} + (\epsilon_{J_i}-\Delta E)\delta_{JJ_i}] = 0, \quad J = J_1, J_2, \ldots, J_k. \quad (3)$$

Here, $\epsilon_{J_i} = E_{nLJ_i} - \bar{E}$ and

$$W_{JJ_i} = \langle\phi_{nLJM}|\hat{V}[1 + G'_{\bar{E}}(\hat{V}-\Delta E)]^{-1}|\phi_{nLJ_iM}\rangle \quad (4)$$

is the matrix of the operator $\hat{V} = \mathbf{F}_0\cdot\mathbf{r}$ that takes into account the interaction of an atom with a dc field $\mathbf{F}_0 = F_0\mathbf{e}_0$ in all orders in the field amplitude F_0 [13, 14] (here, we use the atomic system of units, $\mathbf{r}$ being the radius vector of an optical electron).

The energy correction $\Delta E = E - \bar{E}$ is determined by diagonalizing this matrix, which is equivalent to solving the secular equation

$$\det\|W_{JJ_i} + (\epsilon_{J_i}-\Delta E)\delta_{JJ_i}\| = 0. \quad (5)$$

Here and in (3), δ_{JJ_i} is the Kronecker delta.

Note that, in contrast to the first term in Eq. (1) with fixed principal n, orbital L, and magnetic M quantum numbers, in the second term, which is orthogonal to the first term, only the projection of the total momentum M onto the direction of the dc field is fixed. Thus, the orbital and total momenta of an atom in state Ψ are not fixed; this may lead to violation of the selection rules for radiative transitions and, hence, may give rise to lines of dipole-forbidden transitions, which are missing in the spectrum of a free atom and whose intensity increases with the field strength F_0.

In contrast to degenerate states, for which $\epsilon_{J_i} \equiv 0$, only one coefficient a_{J_jM} may be nonzero for $F_0 = 0$ in (1); as follows from (3) and (5), this coefficient corresponds to the state with energy $\Delta E = \epsilon_{J_j}$. However, for a field such that $|W_{JJ_i}| \sim \epsilon_{J_i}$, $i = 1, 2, \ldots, k$, all coefficients in Eqs. (1) and (3) are close to each other in absolute value and weakly depend on the field. Obviously, the accuracy of calculating the energy ΔE from Eq. (5) depends on the accuracy of calculating a matrix element, which can be calculated by (4) up to an arbitrary order in the field.

2.1. Pairwise Interacting States

Consider, for simplicity, the interaction between two close-lying sublevels ($k = 2$, $\epsilon_{J_{1(2)}} = {}_{(+)}^{-}\delta/2$, where $\delta = E_{nLJ_2} - E_{nLJ_1}$ is the splitting energy in zero field) that are realized in the triplet states of helium with $M = 0$ and $J = L \pm 1$, as well as in the states with $M = \pm L$ and $J = L, L + 1$. In this case, the set of equations (3) and secular equation (5) take the form

$$\begin{aligned}(W_{J_1J_1} - \delta/2 - \Delta E)a_{J_1M} + W_{J_1J_2}a_{J_2M} &= 0,\\ W_{J_2J_1}a_{J_1M} + (W_{J_2J_2} + \delta/2 - \Delta E)a_{J_2M} &= 0,\end{aligned} \tag{6}$$

$$\begin{aligned}(W_{J_1J_1} - \delta/2 - \Delta E)(W_{J_2J_2} + \delta/2 - \Delta E)\\ = W_{J_1J_2}W_{J_2J_1}.\end{aligned} \tag{7}$$

The solution of system (6) with regard to the normalization condition

$$\langle\Psi|\Psi\rangle = \sum_{i=1}^{k}|a_{J_iM}|^2 - \langle\Psi|G'_{\bar{E}}(\hat{V} - \Delta E)|\Psi\rangle = 1$$

leads to equations for the superposition coefficients a_{JM} that make it possible to apply an iterative procedure to calculate corrections to the wavefunction of arbitrary order in F_0:

$$\begin{aligned}|a_{J_1M}|^2 &= A\frac{W_{J_2J_2} + \delta/2 - \Delta E}{W_{J_1J_1} + W_{J_2J_2} - 2\Delta E},\\ |a_{J_2M}|^2 &= A\frac{W_{J_1J_1} - \delta/2 - \Delta E}{W_{J_1J_1} + W_{J_2J_2} - 2\Delta E},\end{aligned} \tag{8}$$

where

$$A = 1 + \langle\Psi|G'_{\bar{E}}(\hat{V}(\mathbf{r}') - \Delta E)|\Psi\rangle.$$

The substitution of the solutions to secular equation (7) for ΔE allows one to express the superposition coefficients as

$$\begin{aligned}a^{\pm}_{J_1M} &= \mp\mathrm{sgn}(W_{J_1J_2})a^{\mp}_{J_2M}\\ &= \left\{\frac{A}{2}\left(1 \mp \frac{(\delta + W_{J_2J_2} - W_{J_1J_1})}{\delta(F_0)}\right)\right\}^{1/2}.\end{aligned} \tag{9}$$

Here,

$$\delta(F_0) = [(\delta + W_{J_2J_2} - W_{J_1J_1})^2 + 4(W_{J_1J_2})^2]^{1/2},$$

$\mathrm{sgn}x = x/|x|$ is the sign function, and the signs $\pm$ are chosen with regard to the following limit conditions for $F_0 \longrightarrow 0$ ($W_{JJ'} \longrightarrow 0$):

$$\begin{aligned}a^{-}_{J_1M} &= \mathrm{sgn}(W_{J_1J_2})a^{+}_{J_2M} \longrightarrow 1,\\ a^{+}_{J_1M} &= -\mathrm{sgn}(W_{J_1J_2})a^{-}_{J_2M} \longrightarrow 0.\end{aligned}$$

To calculate the matrix element (4), it is convenient to introduce a system of coordinates with the axis $Z \parallel \mathbf{e}_0$. Then, the operator of interaction with a dc field $F_0\mathbf{e}_0$ can be represented as

$$\hat{V} = F_0 z.$$

Let us derive perturbation theory series for the wavefunction and the energies of multiplet states in a dc electric field $F_0 \ll 1$. To this end, we expand all field-depending quantities in Eqs. (3) in powers of the field [11]:

$$\begin{aligned}W_{JJ'} &= W^{(2)}_{JJ'} + W^{(4)}_{JJ'} + \ldots,\\ \Delta E &= \Delta E^{(2)} + \Delta E^{(4)} + \ldots,\\ a_{JM} &= a^{(0)}_{JM} + a^{(2)}_{JM} + \ldots,\end{aligned} \tag{10}$$

where

$$\begin{aligned}W^{(2)}_{JJ'} &= F_0^2 w^{(2)}_{JJ'},\\ W^{(4)}_{JJ'} &= F_0^4 w^{(4)}_{JJ'} + F_0^2 u_{JJ'}\Delta E^{(2)}.\end{aligned} \tag{11}$$

The field-independent coefficients in these expressions can be expressed as second-, third-, and fourth-order matrix elements,

$$\begin{aligned}w^{(2)}_{JJ'} &= -\langle\phi_{nLJM}|zG'_{\bar{E}}z|\phi_{nLJ'M}\rangle,\\ u_{JJ'} &= -\langle\phi_{nLJM}|zG'_{\bar{E}}G'_{\bar{E}}z|\phi_{nLJ'M}\rangle,\\ w^{(4)}_{JJ'} &= -\langle\phi_{nLJM}|zG'_{\bar{E}}zG'_{\bar{E}}zG'_{\bar{E}}z|\phi_{nLJ'M}\rangle.\end{aligned} \tag{12}$$

Note that, in contrast to the corrections to the wavefunctions of Stark states of a hydrogen atom [7], expansions (10) contain only even-order corrections. In the vicinity of the minimum of the splitting of the multiplet sublevels (in the neighborhood of the anticrossing $F_0 \approx F_a$), corrections to the matrix elements and to the energy are comparable to the splitting ($W_{JJ'} \sim \alpha_{nL}F_0^2 \sim \delta$, where α_{nL} is the polarizability). Therefore, we assume that the splitting δ is on the order of F_0^2. Only the first nonvanishing correction to a matrix element is strictly quadratic in the field, $W^{(2)}_{JJ'} \sim F_0^2$. For other quantities, the order number does not imply that the relevant quantity is proportional to the corresponding power of the field but just indicates the relative contribution of these quantities to expansion (10), whose terms form a descending sequence; in particular, $|a^{(0)}_{JM}| \gg |a^{(2)}_{JM}| \gg |a^{(4)}_{JM}| \gg \ldots$.

A correction to the energy of the first nonvanishing term of perturbation theory series contains the fine-structure splitting energy δ and the components of matrix elements W that are quadratic in the field:

$$\Delta E^{(2)}_{\pm} = \frac{1}{2}\{W^{(2)}_{J_1J_1} + W^{(2)}_{J_2J_2} \pm \delta^{(2)}(F_0)\}, \tag{13}$$

where

$$\delta^{(2)}(F_0) = E_+^{(2)} - E_-^{(2)}$$
$$= [(\delta + W_{J_2J_2}^{(2)} - W_{J_1J_1}^{(2)})^2 + 4(W_{J_1J_2}^{(2)})^2]^{1/2}$$

is the fine-structure splitting energy in the field. This expression is not strictly proportional to F_0^2) and only formally describes the quadratic Stark effect of doublet sublevels. Moreover, for $|W_{JJ'}^{(2)}| \ll \delta$, the off-diagonal matrix element $W_{J_1J_2}^{(2)}$ determines the leading (resonance) contribution to the fourth-order correction [10, 15]. Correction (13) becomes proportional to the square of the field strength, $\Delta E^{(2)} \propto F_0^2$, only for $|W_{JJ'}^{(2)}| \gg \delta$.

To determine the fourth-order correction $\Delta E^{(4)}$ together with nonresonance terms, one should take into account the fourth-order corrections $W_{JJ'}^{(4)}$ to the matrix elements in Eq. (7) [10, 11]. In this case, neither $\Delta E^{(4)}$ nor $W_{JJ'}^{(4)}$ for $|W_{JJ'}^{(2)}| \ll \delta$ are strictly proportional to F_0^4 because the expression for $W_{JJ'}^{(4)}$ contains $\Delta E^{(2)}$. Formula (9) in the zeroth-order approximation gives the coefficients $a_{JM}^{(0)}$, which are functions of the field strength:

$$a_{J_1M}^{(0)\pm} = \mp \mathrm{sgn}(W_{J_1J_2}^{(2)}) a_{J_2M}^{(0)\mp}$$
$$= \left\{\frac{1}{2}\left(1 \mp \frac{(\delta + W_{J_2J_2}^{(2)} - W_{J_1J_1}^{(2)})}{\delta^{(2)}(F_0)}\right)\right\}^{1/2}. \quad (14)$$

For $|W_{J_1J_2}^{(2)}| \gg \delta$, the coefficients $a_{JM}^{(0)}$ ($J = L, L \pm 1$) become comparable to each other and are determined by the field-independent ratio of combinations of second-order matrix elements. The corrections $a_{JM}^{(2)}$ can be determined either from system (3) or from Eqs. (9), taking into account the fourth-order corrections to the matrix elements and to the normalization factor:

$$a_{JM}^{(2)\pm} = a_{JM}^{(0)\pm}\left\{\frac{A^{(2)}}{2} + [1 - (a_{JM}^{(0)\pm})^2]\right.$$
$$\times (1 - 2a_{JM}^{(0)\pm})\frac{W_{J_1J_2}^{(4)}}{W_{J_1J_2}^{(2)}} \pm (a_{JM}^{(0)\pm})^2 \quad (15)$$
$$\left.\times [1 - (a_{JM}^{(0)\pm})^2]^2 \frac{W_{J_2J_2}^{(4)} - W_{J_1J_1}^{(4)}}{(W_{J_1J_2}^{(2)})^2}\delta^{(2)}(F_0)\right\},$$

where

$$A^{(2)} = F_0^2(a_{J_1M}^{(0)2}u_{J_1J_1} + a_{J_2M}^{(0)2}u_{J_2J_2} + 2a_{J_1M}^{(0)}a_{J_2M}^{(0)}u_{J_1J_2}).$$

The substitution of the obtained expressions for energy ΔE and the superposition coefficients a_{JM} into Eq. (1) allows one to obtain an expansion in powers of the field for the wavefunction:

$$\Psi = \Psi^{(0)} + \Psi^{(1)} + \Psi^{(2)} + \ldots. \quad (16)$$

Up to second-order terms, we have

$$\Psi^{(0)} = \sum_{i=1}^{k} a_{J_iM}^{(0)}|\phi_{nLJ_iM}\rangle, \quad (17)$$

$$\Psi^{(1)} = -F_0\sum_{i=1}^{k} a_{J_iM}^{(0)}G'_{\bar{E}}z|\phi_{nLJ_iM}\rangle,$$
$$\Psi^{(2)} = \sum_{i=1}^{k}[a_{J_iM}^{(2)} + F_0^2 a_{J_iM}^{(0)}G'_{\bar{E}}zG'_{\bar{E}}z]|\phi_{nLJ_iM}\rangle. \quad (18)$$

These expressions are sufficient to determine the field-dependence of the amplitudes and the probabilities and intensities of radiative transitions up to the second-order corrections in F_0. Formulas (10)–(13) make it possible to determine the energy of the states up to the fourth-order corrections. Here, one should keep in mind that the field dependence of the superposition coefficients a_{JM} may significantly change the probabilities even in the zeroth-order approximation, especially in the neighborhood of the anticrossing field $F_0 \approx F_a$, where $a_{LM}^{(0)}$ and $a_{L\pm1M}^{(0)}$ become comparable to each other in magnitude.

Of greatest interest are the transitions in the helium spectrum whose initial and/or final states are pairwise interacting triplet states n^3P_J whose wavefunctions in the zeroth-order approximation are given by superpositions (17) with $k = 2$. The spin–orbit splitting of these states is rather large and can be resolved in the emission and absorption spectra [3]. For sublevels with $J_1 = 2$, $J_2 = 0$, and zero projection of the momentum onto the direction of the field ($M = 0$), the effect of anticrossing occurs in the field

$$F_0 = F_a \approx \sqrt{\frac{2\delta_{nP}}{9\alpha_{nP}^t}} \neq 0, \quad (19)$$

in contrast to the states with $J = 1, 2$ and $M = \pm1$, for which $F_a = 0$ and the repulsion of levels occurs for any F_0 (anticrossing in zero field). For definiteness, we will supply the parameter δ with the index of a multiplet, nL. The tensor polarizability α_{nP}^t, which is given in (22) below, is a universal characteristic for all sublevels of the triplet; in particular, it determines the matrix elements of the operator of interaction with field (20). Figures 1a and 1b show the field dependence of the superposition coefficients $a_{JM}^{(0)-}$ of the triplet sublevels 3^3P for $M = 0$ and $M = \pm1$, respectively. One can see that,

for $M = 0$, in an anticrossing field of $F_a \approx 145$ kV/cm, the contribution of the upper sublevel to the wavefunction of the state with energy ΔE_- amounts to about 3% and becomes equal to the contribution of the lower level for $F_0 \approx 3F_a$. The fine-structure splitting δ for the states with $M = \pm 1$ ($J = 1, 2$) is an order of magnitude smaller than that for the states with $M = 0$ ($J = 0, 2$); therefore, the coefficients $a_{21}^{(0)-}$ and $a_{11}^{(0)-}$ become virtually equal even in a field of $F_0 \approx F_a$ (Fig. 1b). Numerical calculations for the n^3P states with $n = 2, 4, 5, 6$ show similar field dependence for the superposition coefficients with regard to the renormalization of the anticrossing field of the levels with $M = 0$ ($J = 2, 0$) according to formula (19) (see Table 1).

The explicit expressions for $w_{JJ'}^{(2)}$ [11],

$$w_{JJ'}^{(2)} = -\frac{1}{2}\sum_{j=0,2} C_{J'Mj0}^{JM}(-1)^{J'+L+S} \times \left[\frac{(2L+1)_{j+1}(2J'+1)}{(2L+1-j)_j}\right]^{1/2} \begin{Bmatrix} L & L & j \\ J' & J & S \end{Bmatrix} \alpha_{nLS}^{(j)}, \quad (20)$$

show that the second-order matrix elements (11) are determined by two irreducible parameters of the $|nLS\rangle$ state: the scalar

$$\alpha_{nL}^{s} = \alpha_{nLS}^{(0)} = \frac{2}{3(2L+1)} \times \langle \phi_{nLJM} | r[Lg_{L-1}^{(n)} + (L+1)g_{L+1}^{(n)}]r | \phi_{nLJM} \rangle \quad (21)$$

and the tensor

$$\alpha_{nL}^{t} = \alpha_{nLS}^{(2)} = -\frac{2L}{3(2L+1)} \times \left\langle \phi_{nLJM} \middle| r\left[g_{L-1}^{(n)} + \frac{2L-1}{2L+3}g_{L+1}^{(n)}\right]r \middle| \phi_{nLJM} \right\rangle \quad (22)$$

polarizabilities. Here, $g_{L'}^{(n)}$ is a reduced radial Green function in a subspace of states with orbital momentum L' [13]. Formula (20) involves standard notation for the Clebsch–Gordan coefficients and the 6j symbols [16]. The numerical values of the scalar and tensor polarizabilities calculated by the model potential method [13] for the P, D, and F states with the principal quantum number $n \le 6$ of orthohelium are presented in Table 1, together with the values of the fine-structure splitting taken from [17] and the anticrossing field (19). The table shows that the effect of anticrossing may occur not only for the n^3P_{2-0} states but also for the 3^3D_{3-1} states with $M = 0$. However, the anticrossing field of the 3^3D_{3-1} sublevels,

$$F_a(3^3D_{3-1}) = \sqrt{\frac{4\delta_{3D}}{5\alpha_{3D}^{t}}},$$

is about an order of magnitude less than that of the 3^3P_{2-0} sublevels.

When calculating the probability of the radiative transition $n^3P \longrightarrow n'^3P$ in the vicinity of the anticross-

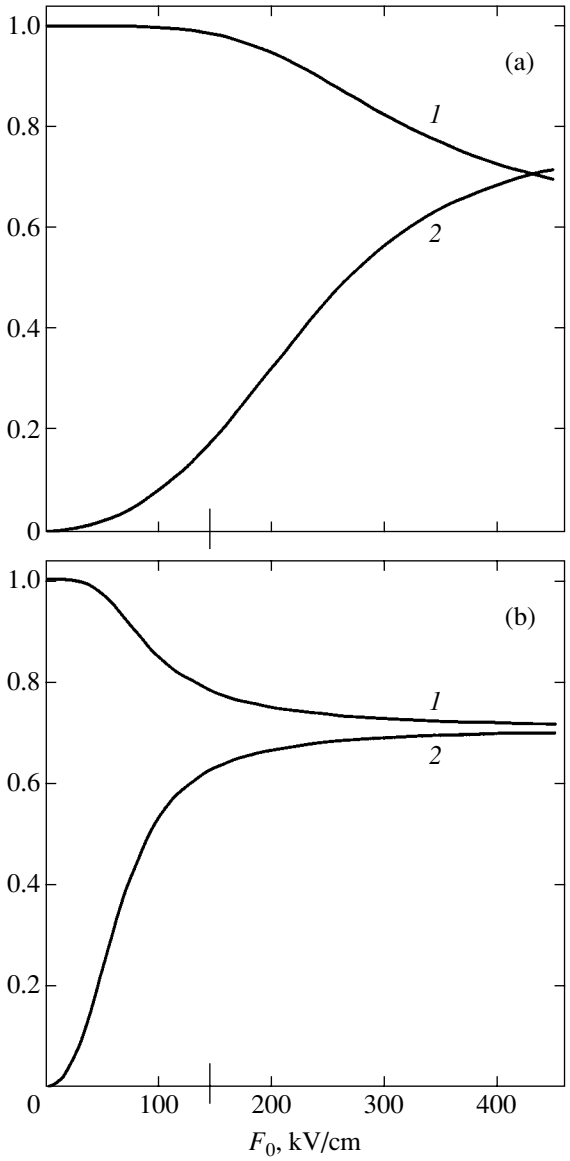

Fig. 1. Superposition coefficients $a_{JM}^{(0)-}$ of the sublevels of the triplet 3^3P; (a) $M = 0$: (*1*) $a_{20}^{(0)-} = -a_{00}^{(0)+}$ and (*2*) $a_{00}^{(0)-} = a_{20}^{(0)+}$; (b) $M = \pm 1$: (*1*) $a_{2\pm1}^{(0)-} = \mp a_{1\pm1}^{(0)+}$ and (*2*) $\pm a_{1\pm1}^{(0)-} = a_{2\pm1}^{(0)+}$. The vertical line on the horizontal axis corresponds to the anticrossing field $F_a = 145$ kV/cm.

Table 1. Numerical values of the spin–orbit splitting $\delta_{nL} = E_{nL_{L-1}} - E_{nL_{L+1}}$, the anticrossing field F_a, and the scalar α^s_{nL} and tensor α^t_{nL} polarizabilities of triplet states of a helium atom

n	δ_{nP}, MHz	δ_{nD}, MHz	F_a, kV/cm	α^s_{nP}, au	α^t_{nP}, au	α^s_{nD}, au	α^t_{nD}, au	α^s_{nF}, au	α^t_{nF}, au
2	31908	–	617	48.0	74.8	–	–	–	–
3	8772	1401	145	1.72(4)	374	–9.32(3)	1.07(4)	–	–
4	3576	591.2	43.9	1.70(5)	1.66(3)	2.89(6)	–7.46(5)	–2.13(6)	2.14(6)
5	1798	303.4	14.9	8.98(5)	7.26(3)	1.94(7)	–5.11(6)	7.10(7)	–2.13(7)
6	1028	175.4	5.55	3.40(6)	2.78(4)	7.95(7)	–2.11(7)	3.93(8)	–1.30(8)

Note: The number in parentheses indicates a decimal exponent: $x(n) = x \times 10^n$.

ing field of the lower n'^3P state ($n' < n$), one may assume that all fine-structure components of the upper level n^3P are completely mixed. This assumption is especially valid for the n^3D and n^3F states, which have nearly an order of magnitude smaller spin–orbit splitting; moreover, the tensor polarizability of these states is from 2 to 3 orders of magnitude greater in absolute value than that of the n^3P states, as is shown in Table 1. Thus, the condition $\left|W^{(2)}_{J_1J_2}\right| \sim F_0^2\left|\alpha^t_{nL}\right| \gg \delta_{nL}$ is satisfied for $L \geq 2$ virtually over the entire range of the field considered. In this case, one may neglect the ϵ_{J_i} in Eqs. (3) for the upper level. The scalar polarizability, which enters the diagonal matrix elements and energy, cancels out in the coefficients $W_{JJ} - \Delta E$ of these equations and vanishes. Hence, in the first nonvanishing order, all coefficients of the set of equations (3) are proportional to α^t_{nL}. Since these equations are homogeneous, one can eliminate the factor α^t_{nL} from all terms, so that, ultimately, the coefficients of the set of equations are determined only by a combination of the coefficients of the vector sum in (20) and depend only on the set of spin–orbit quantum numbers. Thus, using (20) and (11), we can obtain the following limit (for $\delta_{nL} \ll \left|\alpha^t_{nL}\right|F_0^2$) expressions for energy (13) and superposition coefficients (14) of pairwise interacting states.

1. For nL states with $M = 0$ and $J = L + 1, L - 1$:

$$\Delta E_\pm = -\frac{F_0^2}{2}\left[\alpha^s_{nL} - \frac{(2L^2 + 2L - 3) \pm 3\,\mathrm{sgn}\,\alpha^t_{nL}}{2L(2L-1)}\alpha^t_{nL}\right], \quad (23)$$

$$a^{(0)\pm}_{L+1\,0} = \pm a^{(0)\mp}_{L-1\,0}\,\mathrm{sgn}\,\alpha^t_{nL} = \sqrt{\frac{2L + 1 \pm \mathrm{sgn}\,\alpha^t_{nL}}{2(2L+1)}}. \quad (24)$$

2. For nL states with $M = \pm L$ and $J = L, L + 1$,

$$\Delta E_\pm = -\frac{F_0^2}{2}\left[\alpha^s_{nL} + \frac{(2L-3) \mp 3\,\mathrm{sgn}\,\alpha^t_{nL}}{2L}\alpha^t_{nL}\right], \quad (25)$$

$$a^{(0)\pm}_{L+1M} = \pm a^{(0)\mp}_{L\,M}\,\mathrm{sgn}(M\alpha^t_{nL}) = \sqrt{\frac{L + 1 \pm (L-1)\,\mathrm{sgn}\,\alpha^t_{nL}}{2(L+1)}}. \quad (26)$$

Table 2 presents the limit values of the superposition coefficients $a^{(0)-}_{JM}$ for the pairwise interacting states $|n^3L_{J_1J_2}M\rangle$, $L = 1, 2, 3$ that are calculated by the above formulas. When using these values of the coefficients, one should make sure that higher order corrections cannot significantly affect them. To this end, we can evaluate the contribution of the second-order corrections (15). Numerical calculations show that the inequality $\left|a^{(2)}_{JM}\right| \ll \left|a^{(0)}_{JM}\right|$ is valid for any level nL up to field strengths of $F_0 \approx 3F_a$, where F_a is the anticrossing field of the n^3P_J sublevels with $M = 0$.

Table 2. Limit values of the superposition coefficients $a^{(0)-}_{JM}$ for pairwise interacting $|n^3L_{J_1J_2}M\rangle$ states with $\alpha^t_{nL} > 0$. For $\alpha^t_{nL} < 0$, this table presents the coefficients $a^{(0)+}_{JM}$

	$\lvert n^3P_{20}\,0\rangle$	$\lvert n^3P_{21}\pm1\rangle$	$\lvert n^3D_{31}\,0\rangle$	$\lvert n^3D_{32}\pm2\rangle$	$\lvert n^3F_{42}\,0\rangle$	$\lvert n^3F_{43}\pm3\rangle$
$a^{(0)-}_{J_1M}$	$\sqrt{1/3}$	$\sqrt{1/2}$	$\sqrt{2/5}$	$\sqrt{1/3}$	$\sqrt{3/7}$	1/2
$a^{(0)-}_{J_2M}$	$\sqrt{2/3}$	$\pm\sqrt{1/2}$	$\sqrt{3/5}$	$\pm\sqrt{2/3}$	$\sqrt{4/7}$	$\pm\sqrt{3/4}$

2.2. Isolated and Threefold Degenerate States

In the general case of triplet states with arbitrary orbital momentum L, along with the above-described pairwise interacting sublevels that are defined by the wavefunction of the form

$$\Psi^{(0)\pm}_{nLM} = a^{(0)\pm}_{J_1M}\phi_{nLJ_1M} + a^{(0)\pm}_{J_2M}\phi_{nLJ_2M}, \quad (27)$$

in the zeroth-order approximation there exist isolated states with $M = 0$ and $J = L$ and with $|M| = J = L + 1$ that are not mixed with other sublevels in a field. All off-diagonal matrix elements in Eqs. (3) are identically zero for these states, and the first nonvanishing energy correction is strictly quadratic (see (11), (20)):

$$\Delta E^{(2)} = W^{(2)}_{LL} = -\frac{F_0^2}{2}\left(\alpha^s_{nL} + \frac{3 - L(L+1)}{L(2L-1)}\alpha^t_{nL}\right), \quad M = 0, \quad J = L, \quad (28)$$

$$\Delta E^{(2)} = W^{(2)}_{L+1\,L+1} = -\frac{F_0^2}{2}(\alpha^s_{nL} + \alpha^t_{nL}), \quad |M| = J = L + 1, \quad (29)$$

$a_{JM} \equiv 1$, and the wavefunction in the zeroth-order approximation coincides with the wavefunction of an unperturbed atom: $\Psi^{(0)}_{nLJM} = \phi_{nLJM}$. These states include, in particular, all three $|n^3S_1M\rangle$ states with orbital momentum $L = 0$, which differ by the value of the spin projection onto the electric-field direction, $M = 0, \pm 1$ (where $\alpha^t_{nS} \equiv 0$).

In addition, for $L \geq 2$, there are $2(L-1)$ sets of threefold degenerate states, which represent a superposition of all three interacting sublevels of the triplet $|nLJ_iM\rangle$ ($i = 1, 2, 3$) with $J_1 = L + 1$, $J_2 = L$, and $J_3 = L - 1$. Each set corresponds to a certain value of the magnetic quantum number M satisfying the condition $1 \leq |M| \leq L - 1$ and is described by the wavefunctions of zeroth-order approximation (17) with $k = 3$:

$$\Psi^{(0)}_{nL\lambda M} = a^{(0)}_{J_1\lambda M}\phi_{nLJ_1M} + a^{(0)}_{J_2\lambda M}\phi_{nLJ_2M} + a^{(0)}_{J_3\lambda M}\phi_{nLJ_3M}. \quad (30)$$

Here, $\lambda = 1, 2, 3$ is a quantum number that enumerates three solutions of Eqs. (3) that correspond to three multiplet states of an atom in a field. These superpositions have a sufficiently large orbital momentum L; therefore, for these superpositions, the energy of Stark splitting in the range of fields F_0 considered is much greater than the fine-structure splitting δ_{nL}, so that one can set $\epsilon_{J_i} = 0$ in Eqs. (3) and (5).

Calculating the matrix elements $W^{(2)}_{JJ'}$ and substituting them into Eqs. (3) and (5), we obtain three sets of solutions for each M. In the limit case $\epsilon_{J_i} \equiv 0$, these solutions can be expressed in the analytic form both for the energy $\Delta E^{(2)}_{nL\lambda}$ and for the corresponding superposition coefficients $a^{(0)}_{J\lambda M}$. In this case, in contrast to the coefficients $a^{(0)}_{J\lambda M}$, the energy is independent of the sign of the magnetic quantum number. In particular:

I. For $|M| = L - 1 \geq 1$:

$$(1)\ \Delta E^{(2)}_{nL1} = -\frac{F_0^2}{2}(\alpha^s_{nL} + \alpha^t_{nL}),$$
$$a^{(0)}_{L+11M} = \frac{1}{\sqrt{(L+1)(2L+1)}},$$
$$a^{(0)}_{L\,1\,M} = \mathrm{sgn}(M)\frac{1}{\sqrt{L+1}},$$
$$a^{(0)}_{L-11M} = \sqrt{\frac{2L-1}{2L+1}}, \quad (31)$$

$$(2)\ \Delta E^{(2)}_{nL2} = -\frac{F_0^2}{2}\left(\alpha^s_{nL} + \frac{L-3}{L}\alpha^t_{nL}\right),$$
$$a^{(0)}_{L+12\,M} = 2\sqrt{\frac{L}{(L+1)(2L+1)}},$$
$$a^{(0)}_{L\,2\,M} = \mathrm{sgn}(M)\frac{L-1}{\sqrt{L(L+1)}},$$
$$a^{(0)}_{L-12M} = -\sqrt{\frac{2L-1}{L(2L+1)}}. \quad (32)$$

$$(3)\ \Delta E^{(2)}_{nL3} = -\frac{F_0^2}{2}\left(\alpha^s_{nL} + \frac{2L^2 - 13L + 12}{L(2L-1)}\alpha^t_{nL}\right),$$
$$a^{(0)}_{L+13M} = \sqrt{\frac{L(2L-1)}{(L+1)(2L+1)}},$$
$$a^{(0)}_{L\,3\,M} = -\mathrm{sgn}(M)\sqrt{\frac{2L-1}{L(L+1)}},$$
$$a^{(0)}_{L-13M} = \frac{1}{\sqrt{L(2L+1)}}. \quad (33)$$

II. For $|M| = L - 2 \geq 1$:

$$(1)\ \Delta E^{(2)}_{nL1} = -\frac{F_0^2}{2}\left(\alpha^{(s)}_{nL} + \frac{L-3}{L}\alpha^t_{nL}\right),$$
$$a^{(0)}_{L+11M} = \sqrt{\frac{3}{(L+1)(2L+1)}},$$
$$a^{(0)}_{L\,1\,M} = \mathrm{sgn}(M)\sqrt{\frac{2L-1}{L(L+1)}},$$
$$a^{(0)}_{L-11M} = \sqrt{\frac{(L-1)(2L-1)}{L(2L+1)}}, \quad (34)$$

$$(2)\ \Delta E^{(2)}_{nL2} = -\frac{F_0^2}{2}\left(\alpha^s_{nL} + \frac{2L^2-13L+12}{L(2L-1)}\alpha^t_{nL}\right),$$

$$a^{(0)}_{L+12M} = \sqrt{\frac{3(2L-1)}{(L+1)(2L+1)}},$$
$$a^{(0)}_{L2M} = \mathrm{sgn}(M)\frac{L-2}{\sqrt{L(L+1)}}, \quad (35)$$
$$a^{(0)}_{L-12M} = -2\sqrt{\frac{L-1}{L(2L+1)}},$$

$$(3)\ \Delta E^{(2)}_{nL3} = -\frac{F_0^2}{2}\left(\alpha^s_{nL} + \frac{2L^2-19L+27}{L(2L-1)}\alpha^t_{nL}\right),$$

$$a^{(0)}_{L+13M} = \sqrt{\frac{(L-1)(2L-1)}{(L+1)(2L+1)}},$$
$$a^{(0)}_{L3M} = -\mathrm{sgn}(M)\sqrt{\frac{3(L-1)}{L(L+1)}}, \quad (36)$$
$$a^{(0)}_{L-13M} = \sqrt{\frac{3}{L(2L+1)}}.$$

The disposition of solutions is chosen so that, for positive α^t_{nL}, an increase in the quantum number λ corresponds to an increase in energy. All three coefficients $a_{J\lambda M}$ that correspond to the same solution (to energy $E_{nL\lambda}$) may simultaneously change their signs; this is equivalent to the change of the phase of wavefunction (30).

Table 3. Limit values of the energy of the Stark splitting $\Delta E^{(2)}_{nL\lambda}$ and the superposition coefficients $a^{(0)}_{J\lambda M}$ for the threefold degenerate $|n^3L_JM\rangle$ states

L	M	λ	$-2\Delta E^{(2)}_{nL\lambda}/F_0^2$	$a^{(0)}_{L+1\lambda M}$	$a^{(0)}_{L\lambda M}$	$a^{(0)}_{L-1\lambda M}$
2	±1	1	$\alpha^s_{nD}+\alpha^t_{nD}$	$1/\sqrt{15}$	$\pm1/\sqrt{3}$	$\sqrt{3/5}$
		2	$\alpha^s_{nD}-\alpha^t_{nD}/2$	$\sqrt{8/15}$	$\pm1/\sqrt{6}$	$-\sqrt{3/10}$
		3	$\alpha^s_{nD}-\alpha^t_{nD}$	$\sqrt{2/5}$	$\mp1/\sqrt{2}$	$1/\sqrt{10}$
3	±1	1	α^s_{nF}	$\sqrt{3/28}$	$\pm\sqrt{5/12}$	$\sqrt{10/21}$
		2	$\alpha^s_{nF}-3\alpha^t_{nF}/5$	$\sqrt{15/28}$	$\pm1/\sqrt{12}$	$-\sqrt{8/21}$
		3	$\alpha^s_{nF}-4\alpha^t_{nF}/5$	$\sqrt{5/14}$	$\mp1/\sqrt{2}$	$1/\sqrt{7}$
3	±2	1	$\alpha^s_{nF}+\alpha^t_{nF}$	$1/\sqrt{28}$	±1/2	$\sqrt{5/7}$
		2	α^s_{nF}	$\sqrt{3/7}$	$\pm1/\sqrt{3}$	$-\sqrt{5/21}$
		3	$\alpha^s_{nF}-3\alpha^t_{nF}/5$	$\sqrt{15/28}$	$\mp\sqrt{5/12}$	$1/\sqrt{21}$

Table 3 shows the sets of the limit values of energy and the corresponding superposition coefficients for the triplet states D and F with $M = \pm1, \pm2$. For the states with negative α^t_{nL}, the sequence order of solutions is reversed, which corresponds either to the replacement $\lambda = 1 \Leftrightarrow \lambda = 3$ or to the permutation of rows in the table that correspond to $\lambda = 1$ and $\lambda = 3$.

The data of Tables 1–3 can be used to calculate the frequencies and the probabilities of radiative transitions in orthohelium as a function of a dc electric field. Below, we present the calculation of the transition probabilities between triplet levels with momenta $L = 0$, 1, 2, 3: the dipole-allowed transitions $n^3P \longrightarrow n'^3S$ and $n^3D \longrightarrow n'^3P$ and the dipole-forbidden transitions $n^3S \longrightarrow n'^3S$, $n^3D \longrightarrow n'^3S$, $n^3P \longrightarrow n'^3P$, and $n^3F \longrightarrow n'^3P$.

3. PROBABILITY OF A RADIATIVE TRANSITION IN AN ELECTRIC FIELD

The basic quantities that determine the probability of a forced or a spontaneous transition between bound states, $\Psi_i \longrightarrow \Psi_f$, are the transition frequency $\omega_{fi} = E_f - E_i$ and the matrix element of the operator $\hat{V}$ ($\hat{V}^\dagger$) of interaction between an atom and the field of an absorbed (emitted) photon:

$$\mathcal{M}_{fi} = \langle\Psi_f|\hat{V}|\Psi_i\rangle. \quad (37)$$

In the one-electron dipole approximation, the operator of interaction with radiation can be represented as

$$\hat{V} = \frac{F}{2}(\mathbf{e}\cdot\mathbf{r}) = \frac{F}{2}r_\mu,$$

where F is the electric-field strength of the radiation and $r_\mu = r\mathbf{C}_{1\mu}(\theta, \varphi)$. Here, $\mathbf{C}_{1\mu}(\theta, \varphi)$ is a modified spherical function [16]; $\mu = 0$ for π radiation (the polarization vector is parallel to the dc field, $\mathbf{e} \parallel \mathbf{e}_0$), and $\mu = \pm1$ for σ radiation (the polarization vector is perpendicular to the dc field, $\mathbf{e} \perp \mathbf{e}_0$) [18].

The field dependence of the transition frequency is related to the Stark effect for energies E_i and E_f. For isolated states near the point $F_0 \approx F_a$, this dependence is quite accurately described by quadratic corrections that are determined by the polarizabilities from Table 1. For the fine-structure components, higher order corrections also become significant [15]; they can be calculated by perturbation theory for close-lying levels [11, 14]. Simple estimates show that, for $F_0 \le F_a$, where F_a is the anticrossing field of the lower state, the relative variation in the frequency of radiative transitions of optical range is no greater than 0.1%. Therefore, in calculating the field dependence of the probabilities of the transi-

tions considered below, we neglect the contribution of the Stark corrections to the energies E_i and E_f.

The field dependence of a matrix element that may vanish in zero field due to the dipole selection rules is more important. Under the influence of a dc field $\mathbf{F}_0$, the selection rules are significantly changed; this, in particular, gives rise to dipole-forbidden radiation lines. In turn, the probabilities of dipole-allowed transitions may also be significantly changed. This effect is especially clearly manifested in the ratio of the transition intensities within multiplets; it leads to a decrease (up to vanishing) of certain components and the emergence or the enhancement of other fine-structure components of atomic lines. These effects are mainly related to the variation in the wavefunctions and, hence, in the matrix elements in a field.

In this section, we consider the variation in the transition probabilities of the most intense lines n^3P–n'^3S and n^3D–n'^3P of orthohelium for the field strengths in the vicinity of the anticrossing of the states $^3P_{2-0}$. We assume that the expansion coefficients of the D states reach their asymptotic values and neglect higher order corrections.

3.1. Probabilities of Dipole-Allowed Transitions as a Function of an Electric Field

Due to the field dependence of the superposition coefficients (14), the dependence of $\mathcal{M}_{fi}$ on the field $\mathbf{F}_0$ for dipole-allowed transitions between multiplet sublevels manifests itself even in the zeroth-order approximation $\Psi_{i(f)} = \Psi^{(0)}_{i(f)}$:

$$\begin{aligned}\mathcal{M}^{(0)}_{fi}(F_0) &= \langle\Psi^{(0)}_f|\hat{V}|\Psi^{(0)}_i\rangle \\ &= \sum_{j'=1}^{k'}\sum_{j=1}^{k} a^{(0)}_{J'_j\lambda'M'}a^{(0)}_{J_j\lambda M}\langle\phi_{n'L'J'_jM'}|\hat{V}|\phi_{nLJ_jM}\rangle,\end{aligned} \tag{38}$$

where the primed quantities denote the quantum numbers of the final state, the unprimed quantities denote the quantum numbers of the initial state, and the quantum number λ enumerates the solutions to Eqs. (3) (the $\pm$ solutions in (13) and (14)). Neglecting the variation of the transition frequency and the radial matrix element in the field, we can represent the corresponding field dependence of the probability of a radiative transition as

$$\mathcal{W}_{fi}(F_0) = R_{\lambda'\lambda}(F_0)\mathcal{W}_{fi}(0), \tag{39}$$

where $\mathcal{W}_{fi}(0)$ is the integrated probability of transitions for $F_0 = 0$ between the fine-structure sublevels of the initial $|nLJM\rangle$ and final $|n'L'J'M'\rangle$ states with given M and M'. Thus, we can obtain the following relation for the ratio of probabilities $R_{\lambda'\lambda} = \mathcal{W}_{fi}(F_0)/\mathcal{W}_{fi}(0)$:

$$\begin{aligned}R_{\lambda'\lambda}(F_0) &= \frac{\left|\mathcal{M}^{(0)}_{fi}(F_0)\right|^2}{\sum_{J'J}\left|\langle\phi_{n'L'J'M'}|\hat{V}|\phi_{nLJM}\rangle\right|^2} \\ &= \left|\sum_J a^{(0)}_{J\lambda M}(-1)^J\sqrt{2J+1}\sum_{J'}a^{(0)}_{J'\lambda'M'}C^{J'M'}_{JM1\mu}\begin{Bmatrix} L & 1 & J \\ J' & 1 & L'\end{Bmatrix}\right|^2 \\ &\quad\times\left\{\sum_J(2J+1)\sum_{J'}\left(C^{J'M'}_{JM1\mu}\right)^2\right. \\ &\quad\left.\times\left\{\begin{matrix} L & 1 & J \\ J' & 1 & L'\end{matrix}\right\}^2\right\}^{-1}.\end{aligned} \tag{40}$$

Here, just as in (38), the summation over $J(J')$ includes only the momenta of interacting states from a basis of dimension $k(k')$. Thus, the field dependence of the probability of a radiative transition, which is allowed in the dipole approximation, between the fine-structure sublevels of an atom in the lowest order term in F_0 is determined by the linear combination of the superposition coefficients $a^{(0)}$ in (38) that satisfy the conditions

$$\left|a^{(0)}_{J'\lambda'M'}\right| \xrightarrow[F_0\to 0]{} \delta_{J'J'_0}, \quad \left|a^{(0)}_{J\lambda M}\right| \xrightarrow[F_0\to 0]{} \delta_{JJ_0},$$

and guarantee the equality

$$\sum_{\lambda'\lambda} R_{\lambda'\lambda}(F_0) = 1.$$

3.2. Probabilities of the $n^3P \longrightarrow n'^3S$ Transitions

For a transition between the $|n'^3S_1M'\rangle$ state, which has no fine structure, and the $|n^3P_\lambda M\rangle$ state, which is superposition (27) of two states of a triplet, matrix element (38) is represented as a superposition of the matrix elements of P–S transitions from two interacting sublevels $|n^3PJM\rangle$:

$$\begin{aligned}\mathcal{M}^{(0)}_{n'n} &= a^{(0)}_{J_1M}\langle\phi_{n'SM'}|\hat{V}|\phi_{nPJ_1M}\rangle \\ &\quad + a^{(0)}_{J_2M}\langle\phi_{n'SM'}|\hat{V}|\phi_{nPJ_2M}\rangle.\end{aligned}$$

Each of these matrix elements can be expressed in terms of the same radial matrix element $\langle\phi_{n'S}|r|\phi_{nP}\rangle$ and the Clebsch–Gordan coefficients:

$$\begin{aligned}&\langle\phi_{n'SM'}|\hat{V}|\phi_{nPJM}\rangle \\ &= -\frac{F}{6}\sqrt{2J+1}C^{1M'}_{JM1\mu}\langle\phi_{n'S}|r|\phi_{nP}\rangle.\end{aligned}$$

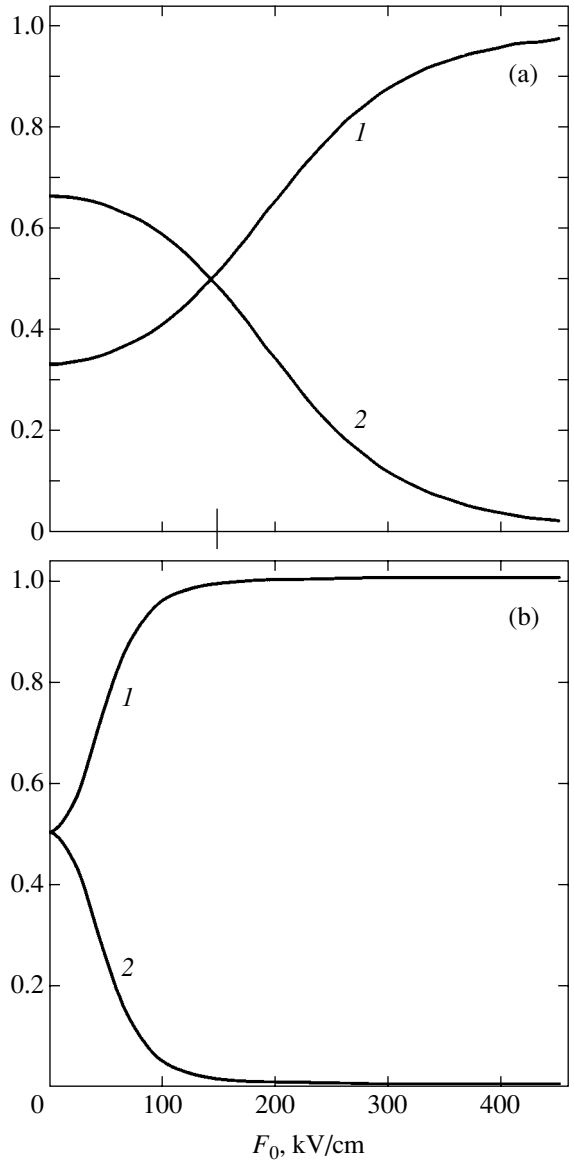

Fig. 2. Probabilities R_λ of radiative transitions from the $|3^3P_\pm M\rangle$ to the $|2^3S_1M'\rangle$ states: (a) $M = 0$ and (b) $M = \pm 1$; (*1*) $R_{+(-)}$ ($3^3P_{+(-)}$, $M \xrightarrow{\pi(\sigma)} n'^3S_1$, M') and (*2*) $R_{-(+)}$ ($3^3P_{-(+)}$, $M \xrightarrow{\pi(\sigma)} n'^3S_1$, M').

Thus, among the dipole-allowed n^3P–n'^3S lines with fixed n and n', we have four components

$$R(n^3P_2 \pm 2 \xrightarrow{\sigma} n'^3S_1 \pm 1) = R(n^3P_1 0 \xrightarrow{\sigma} n'^3S_1 \pm 1) = 1$$

that are independent of the dc field F_0 and six lines whose matrix elements depend on the field even in the zero order in F_0. This dependence is determined by the superposition coefficients (the index $\lambda' = J' = 1$ of the final state, which is the same for all transitions, is omitted):

$$R_\lambda(n^3P_\lambda 0 \xrightarrow{\pi} n'^3S_1 0) = \frac{1}{3}(\sqrt{2}a^{(0)}_{2\lambda 0} - a^{(0)}_{0\lambda 0})^2, \quad (41)$$

$$R_\lambda(n^3P_\lambda 0 \xrightarrow{\sigma} n'^3S_1 \pm 1) = \frac{1}{3}(a^{(0)}_{2\lambda 0} + \sqrt{2}a^{(0)}_{0\lambda 0})^2, \quad (42)$$

$$R_\lambda(n^3P_\lambda \pm 1 \xrightarrow{\pi} n'^3S_1 \pm 1) = \frac{1}{2}(a^{(0)}_{2\lambda\pm 1} \mp a^{(0)}_{1\lambda\pm 1})^2, \quad (43)$$

$$R_\lambda(n^3P_\lambda \pm 1 \xrightarrow{\sigma} n'^3S_1 0) = \frac{1}{2}(a^{(0)}_{2\lambda\pm 1} \pm a^{(0)}_{1\lambda\pm 1})^2. \quad (44)$$

The explicit dependence of Eqs. (43) and (44) on the sign of the magnetic quantum number does not influence the transition probabilities since the relative sign of the coefficients $a^{(0)}_{J\lambda\pm 1}$ also depends on M.

The probabilities of radiative π and σ transitions from state $|3^3P_\pm M\rangle$ with $M = 0$ to state $|2^3S_1M'\rangle$ versus a dc electric field are shown in Fig. 2a. One can see that the probabilities of the π transitions $R_+(3^3P_+0 \xrightarrow{\pi} n'^3S_1 0)$ and $R_-(3^3P_-0 \xrightarrow{\pi} n'^3S_1 0)$ (curves *1* and *2*), which correspond to a radiative decay of the $|3^3P_J 0\rangle$ levels with $J = 0$ and $J = 2$ and differ by a factor of 2 at $F_0 = 0$, become equal in a field of $F_0 = F_a \approx 145$ kV/cm, which corresponds to the anticrossing field (indicated by a long vertical line on the horizontal axis). For $F_0 >$ 400 kV/cm, the integrated probability, which corresponds to $R^\pi_\Sigma = R^{(\pi)}_+ + R^{(\pi)}_-$ and remains invariant in the approximation considered, accumulates at the transition $|3^3P_+0\rangle \xrightarrow{\pi} |n'^3S_1 0\rangle$ (curve *1*), while the probability of the transition $|3^3P_-0\rangle \xrightarrow{\pi} |n'^3S_1 0\rangle$ becomes negligible (curve *2*). A similar situation occurs for σ transitions. The probabilities of these transitions also differ by a factor of 2 at $F_0 = 0$ and become equal at $F_0 = F_a$; the integrated probability $R^\sigma_\Sigma = R^{(\sigma)}_+ + R^{(\sigma)}_-$ is accumulated at the transition $|3^3P_-0\rangle \xrightarrow{\sigma} |n'^3S_1 \pm 1\rangle$, while the probability of the σ radiation of the line $|3^3P_+0\rangle \xrightarrow{\sigma} |n'^3S_1 \pm 1\rangle$ becomes negligible as F_0 increases.

Figure 2b represents the relative probabilities of transitions from the $|3^3P_\lambda M\rangle$states with $M = \pm 1$. In this case, significant variations in the probabilities occur at lower fields of $F_0 \le 100$ kV/cm. For $F_0 \ge 100$ kV/cm, the probabilities with $\lambda = +$ for the π transitions and with

$\lambda = -$ for the σ transitions remain nonzero (curve *1*). The probabilities of the remaining two lines vanish (curve *2*).

3.3. Probabilities of the $n^3D \longrightarrow n'^3P$ Transitions

Among the $|n^3D_JM\rangle$ sublevels of the D multiplet are isolated sublevels $|n^3D_20\rangle$ and $|n^3D_3 \pm 3\rangle$, pairwise interacting sublevels $|n^3D_10\rangle$ and $|n^3D_30\rangle$, and threefold degenerate sublevels $|n^3D_J \pm 1\rangle$, which prove to be completely mixed at a field strength F_0 in the vicinity of the anticrossing of the n'^3P state. In such fields, one can take the values corresponding to the limit for $\left|W^{(2)}_{J_1J_2}\right| \gg \delta$ from Tables 1 and 2 as the superposition coefficients $a^{(0)}_{J\lambda M}$ of the initial $|n^3D_\lambda M\rangle$ state.

In the set of all transitions $|n^3D\rangle \longrightarrow |n'^3P\rangle$ between the multiplet sublevels with fixed n and n', there is a nonzero probability of observing three π transitions that are independent of F_0 in the approximation considered (despite the degeneracy of the momentum projection with respect to its sign, transitions between the states with opposite signs of M are assumed to be different),

$$R(n^3D_20 \xrightarrow{\pi} n'^3P_10) = 1, \tag{45}$$

$$R_\lambda(n^3D_\lambda \pm 2 \xrightarrow{\pi} n'^3P_2 \pm 2) = \begin{cases} 0, & \lambda = 1, \\ 1, & \lambda = 2; \end{cases} \tag{46}$$

and eight σ transitions

$$\begin{aligned} R(n^3D_3 \pm 3 &\xrightarrow{\sigma} n'^3P_2 \pm 2) = 1, \\ R_\lambda(n^3D_\lambda \pm 1 &\xrightarrow{\sigma} n'^3P_2 \pm 2) \\ &= \begin{cases} 0, & \lambda = 1, 2, \\ 1, & \lambda = 3, \end{cases} \end{aligned} \tag{47}$$

$$R_\lambda(n^3D_\lambda \pm 1 \xrightarrow{\sigma} n'^3P_10) = \begin{cases} \dfrac{12}{7\lambda(\lambda+1)}, & \lambda = 1, 3, \\ 0, & \lambda = 2. \end{cases} \tag{48}$$

Here, for the sake of uniformity, we use numerical notation for the $\pm$ states of a doubly degenerate basis, similar to a threefold degenerate basis, starting from the state with the minimal energy (i.e., $\lambda = 1$ for the $-$ state and $\lambda = 2$ for the $+$ state). In addition, there exist transitions whose field dependence is determined by the superposition coefficients of the $|n'^3P_\lambda 0\rangle$ states even in the zeroth-order approximation. They include four π transitions

$$R_{\lambda'\lambda}(n^3D_\lambda 0 \xrightarrow{\pi} n'^3P_{\lambda'}0) = \begin{cases} \dfrac{1}{7}(a^{(0)}_{2\lambda'0} + \sqrt{2}a^{(0)}_{0\lambda'0})^2, & \lambda = 1, \\ \dfrac{4}{21}(\sqrt{2}a^{(0)}_{2\lambda'0} - a^{(0)}_{0\lambda'0})^2, & \lambda = 2, \end{cases} \tag{49}$$

and 12 σ transitions

$$R_{\lambda'\lambda}(n^3D_\lambda \pm 1 \xrightarrow{\sigma} n'^3P_{\lambda'}0) = \begin{cases} \dfrac{4}{13\lambda(\lambda+1)}(a^{(0)}_{2\lambda'0} + \sqrt{2}a^{(0)}_{0\lambda'0})^2, & \lambda = 1, 3, \\ \dfrac{2}{13}(\sqrt{2}a^{(0)}_{2\lambda'0} - a^{(0)}_{0\lambda'0})^2, & \lambda = 2. \end{cases} \tag{50}$$

Among 28 transitions to the states $|n'^3P_{\lambda'} \pm 1\rangle$ whose probability is different from zero and depends on a field via the coefficients $a^{(0)}_{J'\lambda'\pm1}$, there are 8 π transitions with the probabilities

$$R_{\lambda'\lambda}(n^3D_\lambda \pm 1 \xrightarrow{\pi} n'^3P_{\lambda'} \pm 1) = \begin{cases} 0, & \lambda = 1, \\ \dfrac{3}{14}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 2, \\ \dfrac{2}{7}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 3, \end{cases} \tag{51}$$

and 20 σ transitions

$$R_{\lambda'}(n^3D_20 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \frac{1}{2}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1})^2, \tag{52}$$

$$R_{\lambda'}(n^3D_\lambda 0 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \begin{cases} \dfrac{3}{10}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 1, \\ \dfrac{1}{5}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 2, \end{cases} \tag{53}$$

$$R_{\lambda'\lambda}(n^3D_\lambda \pm 2 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \begin{cases} \dfrac{1}{3}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 1, \\ \dfrac{1}{6}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1})^2, & \lambda = 2. \end{cases} \tag{54}$$

Here, the sign ($\pm$) in parentheses corresponds to the sign of the magnetic quantum number $M' = \pm1$.

The right-hand sides of expressions (49)–(54) for the probabilities coincide with those of (41)–(44); therefore, the field dependence of these probabilities, just as for the probabilities of the transitions $n^3P \longrightarrow n'^3S$, is represented by the curves shown in Fig. 2. These probabilities satisfy the following properties, which were pointed out above:

(1) In the anticrossing field of the $n'^3P_{\lambda'}0$ levels, the transition probabilities (49) and (50) take equal values for different λ'.

(2) The amplitudes of half of the transitions to the states $n'^3P_{\lambda'}M'$ vanish as the field F_0 increases.

4. PROBABILITIES OF RADIATIVE TRANSITIONS INDUCED BY A DC ELECTRIC FIELD

The effect of inducing, by an electric field, radiative transitions that are missing in the spectrum of a free atom due to the selection rules for dipole transitions is even more important than the variation in the probabilities of dipole-allowed transitions. The most interesting representatives of such transitions are the nS–$n'S$ transitions, which are forbidden in the nonrelativistic approximation in all orders of multipolarity, and the $n^3P_{0,1}$–n'^3P_0 transitions (the transitions $n^3P_{1,2}$–$n'^3P_{1,2}$ and n^3P_2–n'^3P_0 are allowed in electroquadrupole approximation). The corresponding field dependence of the probability $\mathcal{W}_{\rm fi}(F_0)$ is determined by the squared matrix element of a dipole transition in the first order of perturbation theory $\mathcal{M}_{\rm fi}^{(1)}(F_0)$.

A zeroth-order approximation for wavefunction (1) is not sufficient to calculate radiative matrix elements of such transitions. Here one should take into consideration the term that contains the Green function with the operator of interaction with a dc field $\hat{V}$ and corresponds to the correction to wavefunction (18) in the first order in F_0. Then, the first nonvanishing matrix element is

$$\begin{aligned}\mathcal{M}^{(1)}_{n'\lambda'n\lambda} &= -\sum_J\sum_{J'} a^{(0)}_{J'\lambda'M'}a^{(0)}_{J\lambda M}\langle\phi_{n'L'J'M'}|\hat{v}G^{(n)}\hat{V} \\ &\quad + \hat{V}G^{(n')}\hat{v}|\phi_{nLJM}\rangle \\ &= -\frac{FF_0}{2}\sum_J\sum_{J'} a^{(0)}_{J'\lambda'M'}a^{(0)}_{J\lambda M}U^{(\mu)}_{n'L'J'M'nLJM}.\end{aligned} \tag{55}$$

After integrating with respect to the angular variables with the use of standard methods of the angular momentum algebra [16], the matrix element

$$U^{(\mu)}_{n'L'J'M'nLJM} = \langle\phi_{n'L'J'M'}|r_\mu G^{(n)}r_0 + r_0G^{(n')}r_\mu|\phi_{nLJM}\rangle$$

can be represented as a decomposition into irreducible parts, similar to the decomposition of the amplitude of a two-photon dipole transition on a free atom [13, 19, 20]. Neglecting the effect of the fine structure of intermediate states on the second-order radial matrix element, we obtain

$$\begin{aligned}U^{(\mu)}_{n'L'J'M'nLJM} &= (-1)^{J+L+S} \\ &\times \sum_{p=0,1,2} C^{J'M'}_{JMp\mu}\left[\frac{(2L+1)_{p+1}(2J+1)}{(2L+1-p)_p}\right]^{1/2} \\ &\times \begin{Bmatrix} L'Lp \\ JJ'S\end{Bmatrix}\beta^{(p\mu)}_{n'n},\end{aligned} \tag{56}$$

where

$$\begin{aligned}\beta^{(p\mu)}_{n'n} &= -C^{p\mu}_{1\mu\,10}\left[\frac{(2L+1-p)_p}{(2L+2)_p}\right. \\ &\left.\times(2L'+1)(2p+1)\right]^{1/2} \\ &\times\sum_l C^{l0}_{L'010}C^{l0}_{L010}\begin{Bmatrix} L' & L & p \\ 1 & 1 & l\end{Bmatrix}\sigma^{(p)}_l.\end{aligned} \tag{57}$$

Here, the radial matrix elements with the reduced Green function in a subspace of states with orbital momentum l and the energies of the initial $(g_l^{(n)})$ and final $(g_l^{(n')})$ states are grouped into combinations

$$\sigma^{(p)}_l = \langle n'L'|rg^{(n)}_l r|nL\rangle + (-1)^p\langle n'L'|rg^{(n')}_l r|nL\rangle.$$

The sum over p in decomposition (56) may contain one, two, or all three terms, depending on the values of the orbital quantum numbers L and L', as well as on the type of radiation (on μ) because the summation index p must satisfy the inequalities

$$|\mu| \le p \le 2, \quad |L-L'| \le p \le L+L', \quad |J-J'| \le p \le J+J',$$

which follow from the general properties of the Clebsch–Gordan coefficients and the 6j symbols. Thus, the term with $p = 0$ gives a nonzero contribution to amplitude (56) only for π transitions ($\mu = 0$), and the term with $p = 1$ gives such a contribution only for σ transitions ($\mu = \pm1$). The field-induced transitions $n^3S_1 \longrightarrow n'^3S_1$ and $n^3P_0 \longrightarrow n'^3P_0$ are accompanied

only by π radiation (because the amplitude of σ radiation is $U^{(\pm 1)}_{n'01M'n01M} = U^{(\pm 1)}_{n'100n100} \equiv 0$ in this case).

Note that the quantities $\beta^{(p\mu)}_{n'n}$ are the same for all sublevels of the initial and final multiplets because the dependence of amplitude (56) on the total momenta J and J' and the spin S is determined only by the coefficients of vector summation.

4.1. Transition Probabilities for L' = L

For $L = L' \geq 1$ and $J + J' \geq 2$, all three parameters $\beta^{(p\mu)}_{n'n}$ in (56) may have nonzero values; for these parameters, the μ dependence is expressed in explicit form:

$$\beta^{(0\mu)}_{n'n} = \delta_{\mu 0}\beta^{(0)}, \quad \beta^{(1\mu)}_{n'n} = \mu\beta^{(1)}, \quad \beta^{(2\mu)}_{n'n} = \frac{\sqrt{4-\mu^2}}{2}\beta^{(2)}. \tag{58}$$

The invariant quantities $\beta^{(p)}$ are represented as linear combinations of the radial matrix elements $\sigma^{(0)}_{L'} = \sigma^{(2)}_{L'} \equiv \sigma^{+}_{L'}$ and $\sigma^{(1)}_{L'} \equiv \sigma^{-}_{L'}$, where

$$\sigma^{\pm}_{L'} = \langle n'L|r[g^{(n)}_{L'} \pm g^{(n')}_{L'}]r|nL\rangle,$$

in analogy with the representation of the scalar, antisymmetric, and tensor components of polarizability [21] (see also formulas (21) and (22)):

$$\beta^{(0)} = \frac{1}{3(2L+1)}[L\sigma^{+}_{L-1} + (L+1)\sigma^{+}_{L+1}],$$
$$\beta^{(1)} = \frac{L}{2(2L+1)}[\sigma^{-}_{L-1} - \sigma^{-}_{L+1}], \tag{59}$$
$$\beta^{(2)} = -\frac{L}{3(2L+1)}\left[\sigma^{+}_{L-1} + \frac{2L-1}{2L+3}\sigma^{+}_{L+1}\right].$$

Formulas (58) and (59) imply the following symmetry property:

$$\beta^{(p\mu)}_{n'n} = \beta^{(p-\mu)}_{nn'},$$

which corresponds to opposite signs of the helicity μ for the photons emitted during decay or absorbed during excitation.

Thus, the matrix elements of transitions between states with identical momenta can be represented as linear combinations of the components of transition polarizabilities (59). In particular, the transition $n^3S_1 \longrightarrow n'^3S_1$ is determined by a single (scalar) component

$$\mathcal{M}^{(1)}_{n'n} = -\frac{FF_0}{2}U^{(0)}_{n'J_fnJ_i} = -\frac{FF_0}{2}\beta^{(0)}.$$

In this case, the probability is a strictly quadratic function of the field strength:

$$\mathcal{W}_{n'n} \sim \left|\mathcal{M}^{(1)}_{n'n}\right|^2 \sim F_0^2.$$

For the $n^3P_\lambda \longrightarrow n'^3P_{\lambda'}$ lines, there are both isolated and pairwise interacting states among the upper and lower sublevels. Accordingly, part of the matrix elements of the transitions are determined only by the polarizability components β and are strictly proportional to the dc field strength F_0,

$$\mathcal{M}^{(1)}(n^3P_10 \xrightarrow{\pi} n'^3P_10) = \mathcal{M}^{(1)}(n^3P_2 \pm 2 \xrightarrow{\pi} n'^3P_2 \pm 2) = -\frac{FF_0}{2}(\beta^{(0)} + \beta^{(2)}). \tag{60}$$

For the remaining transitions, the field dependence of the matrix elements is more complicated and includes the field dependence via the superposition coefficients $a^{(0)}_{J\lambda'M}$:

$$\mathcal{M}^{(1)}(n^3P_\lambda 0 \xrightarrow{\pi} n'^3P_{\lambda'}0) = -\frac{FF_0}{2}\beta^{(0)}(a^{(0)}_{0\lambda'0}a^{(0)}_{0\lambda0} + a^{(0)}_{2\lambda'0}a^{(0)}_{2\lambda0}) - \frac{FF_0}{2}\beta^{(2)}(\sqrt{2}a^{(0)}_{2\lambda'0}a^{(0)}_{0\lambda0} + \sqrt{2}a^{(0)}_{0\lambda'0}a^{(0)}_{2\lambda0} - a^{(0)}_{2\lambda'0}a^{(0)}_{2\lambda0}), \tag{61}$$

$$\mathcal{M}^{(1)}(n^3P_\lambda \pm 1 \xrightarrow{\pi} n'^3P_{\lambda'} \pm 1) = -\frac{FF_0}{2}\beta^{(0)}(a^{(0)}_{1\lambda'\pm1}a^{(0)}_{1\lambda\pm1} + a^{(0)}_{2\lambda'\pm1}a^{(0)}_{2\lambda\pm1}) - \frac{FF_0}{4}\beta^{(2)}(-a^{(0)}_{1\lambda'\pm1}a^{(0)}_{1\lambda\pm1} \pm 3a^{(0)}_{2\lambda'\pm1}a^{(0)}_{1\lambda\pm1} \pm 3a^{(0)}_{1\lambda'\pm1}a^{(0)}_{2\lambda\pm1} - a^{(0)}_{2\lambda'\pm1}a^{(0)}_{2\lambda\pm1}), \tag{62}$$

$$\mathcal{M}^{(1)}(n^3P_\lambda 0 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \mp\frac{FF_0\sqrt{3}}{12}\beta^{(1)}(-2\sqrt{2}a^{(0)}_{1\lambda'\pm1}a^{(0)}_{0\lambda0} + a^{(0)}_{1\lambda'\pm1}a^{(0)}_{2\lambda0} \pm 3a^{(0)}_{2\lambda'\pm1}a^{(0)}_{2\lambda0}) - \frac{FF_0\sqrt{3}}{8}\beta^{(2)}(2\sqrt{2}a^{(0)}_{2\lambda'\pm1}a^{(0)}_{0\lambda0} \mp 3a^{(0)}_{1\lambda'\pm1}a^{(0)}_{2\lambda0} - a^{(0)}_{2\lambda'\pm1}a^{(0)}_{2\lambda0}), \tag{63}$$

$$\mathcal{M}^{(1)}(n^3P_2 \pm 2 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \mp\sqrt{2}\mathcal{M}^{(1)}(n^3P_10 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \pm\frac{FF_0}{2\sqrt{2}}\left(\beta^{(1)} + \frac{3}{2}\beta^{(2)}\right)(a^{(0)}_{1\lambda'\pm1} \mp a^{(0)}_{2\lambda'\pm1}). \tag{64}$$

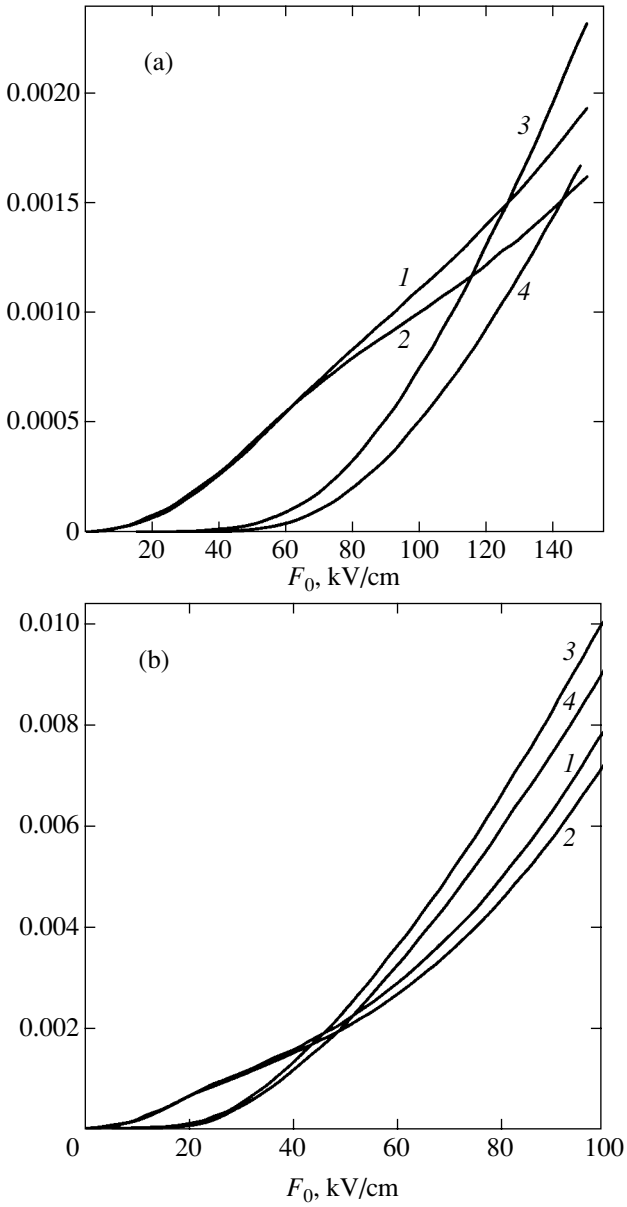

Fig. 3. Relative probability of the forbidden $n^3P_\lambda M$–$n'^3P_{\lambda'}M'$ transitions; (*1*) ($n^3P_+0 \xrightarrow{\pi} n'^3P_+0$), (*2*) ($n^3P\_0 \xrightarrow{\pi} n'^3P\_0$), (*3*) ($n^3P_+0 \xrightarrow{\pi} n'^3P\_0$), and (*4*) ($n^3P\_0 \xrightarrow{\pi} n'^3P_+0$); (a) $n = 4$ and $n' = 2$ and (b) $n = 5$ and $n' = 3$.

Taking into account the limit values of the coefficients $a^{(0)}_{J\lambda' M}$ (Table 2), one can verify that half of the forbidden components (61)–(64), which arise for $F_0 \neq 0$, vanish as F_0 increases. Moreover, if the condition $|W_{JJ'}| \gg \delta$ is fulfilled for the upper nP state in the vicinity of the anticrossing $F_0 \approx F'_a$ of the lower $n'P$ state, then the values $a^{(0)}_{J\lambda 0}$ can be replaced by their asymptotic values from Table 2. Then, similar to the case of allowed transitions, the above formulas are rewritten in terms of combinations of the form $a^{(0)}_{2\lambda' 0} + \sqrt{2}a^{(0)}_{0\lambda' 0}$ and $\sqrt{2}a^{(0)}_{2\lambda' 0} - a^{(0)}_{0\lambda' 0}$. This, in particular, implies again that the transition probabilities for the ± components coincide at $F_0 = F_a$ and that half of these components vanish as the field strength F_0 increases.

The probability of the forbidden nP–$n'P$ transitions can be characterized by the ratio of $\left|\mathcal{M}^{(1)}_{n'PnP}\right|^2$ to the sum of the squared matrix elements of allowed lines with close values of frequency, i.e., of the nD–$n'P$ transitions in zero field. As an example, Fig. 3 shows these ratios for matrix elements (61), where four components correspond to different states in the field for the lower and upper multiplets. The probabilities of dipole-forbidden transitions essentially depend on the principal quantum number of the upper state. The figures show that, in the vicinity of $F_0 = 3F_a$, where F_a is the anticrossing field of the upper nP level, these probabilities amount to about 0.2% of the integrated probability of the dipole-allowed transitions.

In a weak field, the probabilities of dipole-allowed transitions increase quadratically with F_0. However, in the neighborhood of the anticrossing field of the upper state, $F_0 \approx F_a$, the interference of sublevels gives rise to inflection points on the curves corresponding to the transitions $n^3P_\pm \xrightarrow{\pi} n'^3P_\pm$; after the inflection point, the probabilities depend linearly on the field strength up to $F_0 \approx 3F_a$ (curves *1* and *2* in Fig. 3). Note that a similar transition from quadratic to linear field dependence was observed experimentally in [22] for the probability of forbidden two-photon transitions to the Rydberg states of a cesium atom.

For curves *3* and *4*, which describe the probabilities of the transitions $n^3P_\pm \xrightarrow{\pi} n'^3P_\mp$, the inflection points lie near the anticrossing field of the lower state, which falls outside the applicability domain of the zeroth-order approximation for the coefficients a_{JM} of the upper level. One should also notice the intersection of curves *1* and *2* with curves *3* and *4*, which indicates that the probabilities of all transitions under consideration become equal for $F_0 \approx 3F_a$.

In calculating the probabilities, we used the numerical values of the polarizabilities $\beta^{(p)}$ that were obtained for the transitions n^3S_1–n'^3S_1 and n^3P–n'^3P by the model potential method [13] and are presented in Table 4 for n and n' ranging from 2 to 6. The table shows that the main contribution to the probabilities of π transitions is made by the scalar polarizability $\beta^{(0)}$, which has negative values and is from 1 to 2 orders of magnitude greater than the tensor part $\beta^{(2)}$ in absolute value. In this case, $|\beta^{(0)}|$ is a monotonically increasing function of the principal quantum numbers of the initial and final states, unlike the tensor part $\beta^{(2)}$, which is not only nonmonotonic but also may change sign as n and n' increase. The antisymmetric part $\beta^{(1)}$, which deter-

mines the amplitudes of σ transitions, is positive and virtually coincides with $|\beta^{(0)}|$. Hence, the amplitudes of σ and π transitions (63) and (61) for the same field strength are close in absolute value.

4.2. Transition Probabilities for L' = L – 2

For $L' = L - 2$, the sum over p in the correction $\mathcal{M}^{(1)}_{n'n}$ (56) contains a single term with $p = 2$:

$$U^{(\mu)}_{n'L-2J'M'nLJM} = (-1)^{J+L+S}\sqrt{\frac{(2J+1)L(L-1)(4-\mu^2)}{6(2L-1)}} \times C^{J'M'}_{JM2\mu}\begin{Bmatrix} L-2 & L & 2 \\ J & J' & S \end{Bmatrix}\sigma^{+}_{L-1}, \quad (65)$$

where

$$\sigma^{+}_{L-1} = \langle n'L-2|r[g^{(n')}_{L-1} + g^{(n)}_{L-1}]r|nL\rangle$$

is a second-order radial matrix element.

Note that the electric field changes the dipole selection rule for the orbital momentum without changing the selection rule for the magnetic quantum number $\Delta M = 0, \pm 1$. Therefore, a dipole transition from the state $|nLJM\rangle$ to the state $|n'L-2J'M'\rangle$ for $|M| = L + 1$ remains forbidden because $|M - M'| \geq 2$ in this case.

The minimal value of the orbital momentum $L = 2$ in the case under consideration corresponds to the transitions $n^3D \longrightarrow n'^3S$. The anticrossing field of the 3^3D_{1-3} states with $M = 0$ is approximately $F_a(3D) \approx 20$ kV/cm. Estimates show that, for such a field, the probability of a field-induced transition may amount to, at most, 10^{-5} of the probability of allowed transitions. Therefore, to observe the transition $3^3D \longrightarrow 2^3S$, it is necessary that the field strength F_0 be at least ten times greater than $F_a(3D)$. Here, the superposition coefficients of the 3^3D states reach their asymptotic values (see Tables 2 and 3), and the matrix elements of all electric-field-induced transitions are expressed in terms of the same quantity, which determines, for example, the amplitude of any σ transition:

$$\mathcal{M}_{DS} \equiv \frac{FF_0}{2\sqrt{15}}\sigma^{+}_1$$
$$= \mathcal{M}^{(1)}(n^3D_{\lambda=2}\pm 2 \xrightarrow{\sigma} n'^3S_1 \pm 1)$$
$$= \mathcal{M}^{(1)}(n^3D_{\lambda=2}\pm 1 \xrightarrow{\sigma} n'^3S_1 0)$$
$$= \sqrt{2}\mathcal{M}^{(1)}(n^3D_{\lambda=1}0 \xrightarrow{\sigma} n'^3S_1 \pm 1)$$
$$= \mp\sqrt{2}\mathcal{M}^{(1)}(n^3D_{J=2}0 \xrightarrow{\sigma} n'^3S_1 \pm 1).$$

Table 4. Polarizabilities $\beta^{(p)}$ for the transitions n^3S_1–n'^3S_1 and n^3P–n'^3P in orthohelium

Line		n^3S_1–n'^3S_1	n^3P–n'^3P		
n'	n	$\beta^{(0)}$	$\beta^{(0)}$	$\beta^{(1)}$	$\beta^{(2)}$
2	3	–3.48(2)	–3.69(3)	3.32(3)	1.30(2)
2	4	–9.02(2)	–6.95(3)	6.19(3)	2.95(2)
2	5	–1.96(3)	–1.32(4)	1.17(4)	5.80(2)
2	6	–3.73(3)	–2.32(4)	2.06(4)	1.04(3)
3	4	–3.55(3)	–2.79(4)	2.49(4)	9.80(1)
3	5	–5.33(3)	–3.70(4)	3.43(4)	6.17(2)
3	6	–8.90(3)	–5.69(4)	5.31(4)	1.23(3)
4	5	–2.19(4)	–1.35(5)	1.12(5)	–2.22(3)
4	6	–2.24(4)	–1.37(5)	1.27(5)	–4.57(2)
5	6	–9.60(4)	–4.94(5)	3.73(5)	–1.38(4)

In this case, the amplitude of a π transition differs from M_{DS} only by a numerical factor:

$$\mathcal{M}^{(1)}(n^3D_{\lambda=2}0 \xrightarrow{\pi} n'^3S_1 0)$$
$$= \mathcal{M}^{(1)}(n^3D_{\lambda=3}\pm 1 \xrightarrow{\pi} n'^3S_1 \pm 1) = -\frac{2}{\sqrt{3}}\mathcal{M}_{DS}.$$

Consider the field dependence of the probability of π and σ transitions from the states n^3F_λ and $n'^3P_{\lambda'}$. Here, like for the *P*–*P* transitions, the field dependence of a matrix element includes, together with the linear dependence, the dependence on the superposition coefficients of the $n'^3P_{J'}$ states (in this case, the superposition coefficients of the n^3F_J states virtually coincide with their asymptotic values).

Among the sublevels of the *F* multiplet, there are isolated sublevels and the sublevels that interact by pairs and by threes with wavefunctions (27) and (30), respectively. For $\left|W^{(2)}_{J_1J_2}\right| \gg \delta_{nF}$, we can assume that these sublevels are completely mixed, taking the limit values of the superposition coefficients $a^{(0)}_{J\lambda M}$ from Tables 2 and 3. In this case, there are three field-induced π transitions whose matrix elements are linear in the field:

$$\mathcal{M}^{(1)}(n^3F_30 \xrightarrow{\pi} n'^3P_10) = -\frac{FF_0}{5}\sqrt{\frac{2}{7}}\sigma^{+}_2 \equiv \mathcal{M}_{FP}, \quad (66)$$

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 2 \xrightarrow{\pi} n'^3P_2 \pm 2) = \begin{cases} 0, & \lambda = 1, 2, \\ \mathcal{M}_{FP}, & \lambda = 3, \end{cases} \quad (67)$$

and 12 π transitions with the matrix elements that include, together with $\mathcal{M}_{FP}$, the dependence on the superposition coefficients of the $|n'^3P_JM\rangle$ states. Among these transitions are four transitions between the states with $M = 0$,

$$\mathcal{M}^{(1)}(n^3F_\lambda 0 \xrightarrow{\pi} n'^3P_{\lambda'}0) = \begin{cases} \frac{\mathcal{M}_{FP}}{\sqrt{3}}(a^{(0)}_{2\lambda'0} + \sqrt{2}a^{(0)}_{0\lambda'0}), & \lambda = 1, \\ -\frac{\mathcal{M}_{FP}}{\sqrt{2}}(\sqrt{2}a^{(0)}_{2\lambda'0} - a^{(0)}_{0\lambda'0}), & \lambda = 2, \end{cases} \quad (68)$$

and eight transitions between the states with $M = \pm 1$,

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 1 \xrightarrow{\pi} n'^3P_{\lambda'} \pm 1) = \begin{cases} 0, & \lambda = 1, \\ \frac{\mathcal{M}_{FP}}{\sqrt{2}}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1}), & \lambda = 2, \\ \frac{\sqrt{3}\mathcal{M}_{FP}}{2}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1}), & \lambda = 3. \end{cases} \quad (69)$$

Among σ transitions are eight transitions whose matrix elements are strictly proportional to F_0,

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 1 \xrightarrow{\sigma} n'^3P_1 0) = \begin{cases} \mp\sqrt{\frac{5}{8}}\mathcal{M}_{FP}, & \lambda = 1, \\ 0, & \lambda = 2, \\ \pm\frac{\sqrt{3}}{4}\mathcal{M}_{FP}, & \lambda = 3, \end{cases} \quad (70)$$

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 1 \xrightarrow{\sigma} n'^3P_2 \pm 2) = \begin{cases} 0, & \lambda = 1, 2, \\ -\sqrt{\frac{3}{8}}\mathcal{M}_{FP}, & \lambda = 3, \end{cases} \quad (71)$$

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 3 \xrightarrow{\sigma} n'^3P_2 \pm 2) = \begin{cases} 0, & \lambda = 1, \\ -\sqrt{\frac{5}{4}}\mathcal{M}_{FP}, & \lambda = 2, \end{cases} \quad (72)$$

and 32 matrix elements whose matrix elements depend on the superposition coefficients of the n'^3P_JM states with $M = 0$,

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 1 \xrightarrow{\sigma} n'^3P_{\lambda'}0) = \begin{cases} -\sqrt{\frac{5}{24}}\mathcal{M}_{FP}(a^{(0)}_{2\lambda'0} + \sqrt{2}a^{(0)}_{0\lambda'0}), & \lambda = 1, \\ -\frac{1}{\sqrt{3}}\mathcal{M}_{FP}(\sqrt{2}a^{(0)}_{2\lambda'0} - a^{(0)}_{0\lambda'0}), & \lambda = 2, \\ -\frac{1}{4}\mathcal{M}_{FP}(a^{(0)}_{2\lambda'0} + \sqrt{2}a^{(0)}_{0\lambda'0}), & \lambda = 3, \end{cases} \quad (73)$$

and with $M = \pm 1$,

$$\mathcal{M}^{(1)}(n^3F_3 0 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \mp\frac{\mathcal{M}_{FP}}{2}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1}), \quad (74)$$

$$\mathcal{M}^{(1)}(n^3F_\lambda 0 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \begin{cases} -\frac{\mathcal{M}_{FP}}{2}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1}), & \lambda = 1, \\ \frac{\sqrt{3}\mathcal{M}_{FP}}{4}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1}), & \lambda = 2, \end{cases} \quad (75)$$

$$\mathcal{M}^{(1)}(n^3F_\lambda \pm 2 \xrightarrow{\sigma} n'^3P_{\lambda'} \pm 1) = \begin{cases} 0, & \lambda = 1, \\ -\frac{1}{2}\sqrt{\frac{5}{2}}\mathcal{M}_{FP}(a^{(0)}_{2\lambda'\pm1} \pm a^{(0)}_{1\lambda'\pm1}), & \lambda = 2, \\ -\frac{1}{\sqrt{2}}\mathcal{M}_{FP}(a^{(0)}_{2\lambda'\pm1} \mp a^{(0)}_{1\lambda'\pm1}), & \lambda = 3. \end{cases} \quad (76)$$

4.3. Relation between the Amplitudes of Electric-Field-Induced and Electromultipole Transitions

The probability of dipole-forbidden transitions contains, in addition to the field-induced part determined by the matrix element $\mathcal{M}^{(1)}_{n'n}$, a component determined by the electroquadrupole interaction between an atom and the field of a photon. For $\sigma^\pm$ radiation along the external field, both these corrections are different from zero and the corresponding operator of quadrupole interaction is expressed as [8]

$$\hat{V}^{q\pm} = \frac{i\alpha\omega_{nn'}F}{4\sqrt{3}}r^2C_{2\pm1}(\theta, \varphi),$$

where α is the fine-structure constant. The ratio of the first-order dipole matrix element (65) to the quadrupole

matrix element proves to be independent of the momenta of interacting sublevels:

$$\frac{\mathcal{M}^{(1)}_{n'n}}{\mathcal{M}^{q(0)}_{n'n}} = i\frac{2F_0\sigma^+_{L-1}}{\alpha\omega_{n'n}\langle\phi_{n'L-2}|r^2|\phi_{nL}\rangle} \approx i\frac{2F_0}{\alpha\omega_{n'n}(E_{nL-1}-E_{nL})}. \quad (77)$$

This expression is obtained with the use of a simple estimate for the second-order radial matrix element that takes into account the main contribution, to the Green function $g^{(n)}_{L-1}$, of the wavefunction of the state with the orbital momentum $L-1$ from the n shell,

$$\sigma^+_{L-1} \approx \frac{\langle n'L-2|r^2|nL\rangle}{E_{nL-1}-E_{nL}}.$$

For the lines with $L' = L$, the ratio (77) has a more complicated form and contains combinations of the superposition coefficients of interacting sublevels. Numerical estimates for the absolute value of ratio (77) (for n and n' ranging from 2 to 6) in the anticrossing field of the lower level give values ranging from 10^3 to 10^6. Hence, one may neglect multipole corrections to the probability of an electric-field-induced transition when considering the anticrossing effects on the fine-structure sublevels.

5. CONCLUSIONS

The formulas and numerical results obtained above for the superposition coefficients that define wavefunction (1) allow one to extract new information about the spectral properties of an atom in a field. In particular, the variation in the intensities of the Stark components of multiplet lines makes it possible to accurately determine the ratio between the field and spin–orbit interactions of orthohelium in a dc electric field. The probabilities of the radiative transitions versus F_0 obtained above may be useful not only in describing experimentally observed electro-optical effects but also in controlling the optical spectra by a dc electric field.

The properties of fine-structure components in an electric field have been studied earlier for low-lying states of alkaline atoms and alkalinelike ions by diagonalizing the matrix of the spin–orbit and Stark Hamiltonian on a small number of effective sublevels [23]. In particular, the results of the numerical calculations obtained in [23] exhibit an effect similar to that described above, namely, the equalization of the intensities of the Stark components of doublet emission lines (see Fig. 2) near the anticrossing field. The vanishing of nearly half of the components of radiative multiplets as the field strength increases results from the "destruction" of spin–orbit interaction by an electric field and corresponds to the transition with enhancement of the field to the Stark effect on a spinless atom.

The field dependence of the intensities of dipole-forbidden transitions is of no less interest. The decomposition of the amplitude of a field-induced process into irreducible parts (56) makes it possible to reduce the analysis to the calculation of the second-order radial matrix elements that determine invariant atomic quantities—the polarizabilities of transitions (59). The numerical values of the polarizability tensor components allow us to evaluate the probabilities of electric-field-induced transitions not only qualitatively but also quantitatively. In particular, in the anticrossing field of the lower level, this probability may amount to 2–3% of the probability of the dipole-allowed process both for π and σ transitions.

The method for calculating the wavefunctions of multiplet states presented here may be useful for solving spectroscopy problems of quantum systems with multiplet structure in an external electric field.

ACKNOWLEDGMENTS

We are grateful to N.L. Manakov for useful discussions.

This work was supported by the US Civil Research and Development Foundation (US CRDF) and by the Ministry of Education of the Russian Federation (project nos. VZ-010-0 and Y1-P-10-02).

REFERENCES

1. O. Reusch, C. Dieste, S. Garnica, and G. von Oppen, J. Phys. B: At. Mol. Opt. Phys. **34**, 2145 (2001).
2. R. Drozdowski, M. Busch, and G. von Oppen, J. Phys. B: At. Mol. Opt. Phys. **35**, 2494 (2002).
3. R. Schumann, M. Dammasch, U. Eichmann, *et al.*, J. Phys. B: At. Mol. Opt. Phys. **30**, 2581 (1997).
4. L. V. Gorchakov, V. P. Demkin, I. I. Murav'ev, and A. M. Yancharina, *Emission of Inert Gas Atoms in Electrical Field* (Tomsk. Gos. Univ., Tomsk, 1984) [in Russian].
5. M. Bellermann, T. Bergeman, A. Haffmans, *et al.*, Phys. Rev. A **46**, 5836 (1992).
6. A. A. Kamenski and V. D. Ovsiannikov, J. Phys. B **33**, 491, 5543 (2000).
7. A. A. Kamenskiĭ and V. D. Ovsyannikov, Zh. Éksp. Teor. Fiz. **120**, 52 (2001) [JETP **93**, 43 (2001)].
8. A. A. Kamenski and V. D. Ovsiannikov, Can. J. Phys. **81**, 755 (2003).
9. A. Derevyanko, W. R. Johnson, V. D. Ovsyannikov, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 494 (1999) [JETP **88**, 272 (1999)].
10. A. Derevyanko, W. R. Johnson, V. D. Ovsiannikov, *et al.*, Phys. Rev. A **60**, 986 (1999).
11. I. L. Bolgova, V. D. Ovsyannikov, V. G. Pal'chikov, *et al.*, Zh. Éksp. Teor. Fiz. **123**, 1145 (2003) [JETP **96**, 1006 (2003)].

12. I. I. Sobel'man, *Atomic Spectra and Radiative Transitions* (Nauka, Moscow, 1977; Springer, Berlin, 1979), Chap. 9.
13. N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
14. V. D. Ovsiannikov and S. V. Goossev, Phys. Scr. **57**, 506 (1998).
15. V. A. Davydkin and V. D. Ovsiannikov, J. Phys. B **17**, L207 (1984); **19**, 2071 (1986).
16. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
17. NIST Atomic Spectra Database Levels Data, http://physics.nist.gov.
18. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, Berlin, 1957; Fizmatgiz, Moscow, 1960).
19. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Relativistic Quantum Theory* (Nauka, Moscow, 1968; Pergamon, Oxford, 1971), Part 1.
20. L. P. Rapoport, B. A. Zon, and N. L. Manakov, *Theory of Multiphoton Processes in Atoms* (Atomizdat, Moscow, 1978) [in Russian].
21. V. A. Davydkin, V. D. Ovsiannikov, and B. A. Zon, Laser Phys. **3**, 449 (1993).
22. P. P. Herrmann, J. Hoffnagle, N. Schlumpf, *et al.*, J. Phys. B **19**, 1271, 5543 (1986).
23. E. Stambulchik and Y. Maron, Phys. Rev. A **56**, 2713 (1997).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 505–515.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 570–582.

ATOMS, MOLECULES, OPTICS

Selecting Molecules Embedded in Nanodroplets (Clusters) of Superfluid Helium

V. N. Lokhman and G. N. Makarov*

Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: gmakarov@isan.troitsk.ru*

Received September 22, 2004

Abstract—A method of selecting molecules embedded in nanodroplets (clusters) of superfluid helium is proposed, which is based on the selective vibrational excitation of embedded molecules by intense IR laser radiation. This action leads to a significant decrease in size of the excited clusters, after which these clusters are separated with respect to size via scattering of the cluster beam on a crossing atomic beam. The method is described in detail and the possibility of selecting SF_6 molecules in liquid helium nanodroplets using the excitation by CO_2 laser radiation and the angular separation via scattering on a xenon atomic beam is demonstrated. The results show that, by using this technique, it is possible to separate molecules with respect to isotope (element) composition. Advantages and drawbacks of the method are analyzed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In recent years, much attention has been devoted to experimental and theoretical investigations of nanodimensional superfluid helium droplets (clusters) obtained upon gas efflux from cooled nozzle sources (see, e.g., review papers [1–11]). The size of a cluster, as determined by the number N of helium atoms in the droplet, varies from several dozens of atoms to $N \approx 10^7$. Special techniques have been developed for obtaining such superfluid helium nanodroplets and introducing single foreign molecules into these clusters, which makes it possible to study the spectra of molecules and molecular clusters at a very low temperature ($T \approx 0.37$ K) in liquid superfluid helium representing a soft quantum matrix of the new type [1–9]. The results of such spectroscopic measurements also provide important information about the properties of superfluid helium nanodroplets [5, 6, 9].

In addition, the experiments with molecules embedded in liquid helium droplets open wide possibilities for the investigation of many physicochemical processes on an atomic-molecular level at ultimately low temperatures. For example, by introducing molecules into helium nanodroplets, it is possible to obtain long chains of polar molecules oriented in a definite way inside the droplets [12], which is of considerable interest for the synthesis of biologically active molecules. Unique conditions provided for the growth of molecular clusters inside superfluid helium nanodroplets make it possible to obtain high-energy isomers, for example, cyclic water hexamers representing the smallest pieces of ice [13]. It is also possible to study chemical reactions proceeding inside helium nanodroplets at very low temperatures [14] and the orientation effects accompanying the interaction of molecules with nanodroplets [15].

Recently, we have suggested and analyzed the process of selecting molecules embedded in superfluid helium nanodroplets with respect to isotope (elemental) composition [16]. In the present paper, we report on the results of a more extensive and deeper investigation of the possibility of using beams of superfluid helium nanodroplets (clusters) for such separation of embedded molecules. The proposed method is described in detail and the possibilities of this technique are demonstrated and analyzed in application to helium droplets containing SF_6 molecules.

2. PRINCIPLES OF THE METHOD

The results of previous experimental [1, 2, 17, 18] and theoretical [19–22] investigations showed that helium-4 (^{4}He) nanodroplets (clusters) containing $N \geq 100$ particles occur in a superfluid state [23]. Molecules embedded in such droplets can freely rotate and exhibit very narrow IR absorption band owing to a rather low temperature of the medium ($T \approx 0.37$ K [1–3, 6–9]). For example, the total width of the absorption band due to ν_3 vibrations of $^{32}SF_6$ molecules in a superfluid helium nanodroplet is as small as 0.25 cm^{-1} [24–26]. Therefore, the IR absorption spectra can be almost completely resolved not only for the molecules such as SF_6 and SiF_4 with relatively large (~5–10 cm^{-1}) isotope shifts, but also for the molecules containing heavy elements whose isotope shifts in the IR spectra are on the order of (or smaller than) 1 cm^{-1} (OsO_4, WF_6, UF_6, etc.). Thus, a sharp decrease in the absorption lines of molecules embedded in superfluid helium nanodroplets provides the possibility of highly selective excitation of these molecules using IR laser radiation: essentially,

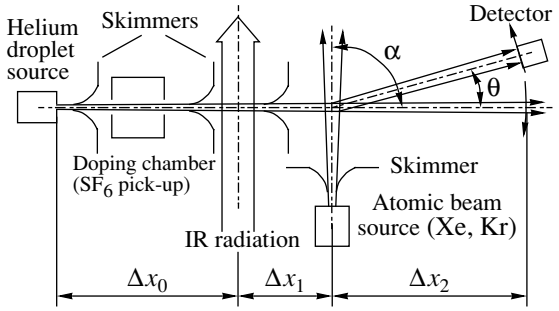

Fig. 1. Schematic diagram of the experimental setup (see the text for explanations).

only those helium clusters in a beam which contain molecules of a selected isotope will be excited.

The absorption of radiation by molecules embedded in a liquid helium nanodroplet leads to heating of this droplet, which results in evaporation of a certain amount of He atoms [1–9, 26]. For example, the absorption of one quantum of a CO_2 laser radiation ($\lambda \approx 10$ μm, $\hbar\omega \approx 0.12$ eV) leads to the evaporation of about 200 atoms [1, 5, 9, 26]. As a result of this evaporation, the droplet cools down and its temperature again stabilizes at $T \approx 0.37$ K [1, 7, 9. 27]. Both heating and evaporation of He atoms upon the absorption of a laser radiation quantum proceed quite rapidly (within a period of time $t \leq 10^{-8}$–10^{-9} s [26–29]). These phenomena provide a basis for the method of selecting molecules embedded in superfluid helium nanodroplets with respect to isotope and elemental composition.

The proposed method is essentially as follows (Fig. 1). A beam of superfluid helium nanodroplets with embedded molecules is irradiated at a certain distance Δx_0 from the nozzle output by intense laser radiation whose frequency is in resonance with the vibrations of molecules of a selected isotope. As was noted above, only those helium clusters in a beam which contain resonant molecules of the selected isotope will be excited, while the other clusters (containing molecules not absorbing this radiation) will remain unexcited. The absorption of radiation energy leads to heating of the droplets that results in the evaporation of helium atoms. Evaporated atoms isotropically fly away from clusters. The binding energy per atom (expressed in kelvins) as a function of the number N of atoms in a liquid helium cluster can be determined using the relation [19, 30]

$$E/N = -7.21 + 17.71N^{-1/3} - 5.95N^{-2/3}. \quad (1)$$

In large clusters ($N \geq 10^6 - 10^7$) the value of E/N is approximately the same (about 7.2 K [19]) as in bulk liquid helium, and this value decreases with a decrease in the cluster size. For example, in clusters with $N \leq 10^3$, the binding energy per atom is $E/N \leq 5.5$ K, while in clusters with $N \leq 200$, this value does not exceed 4 K [19]. It should be taken into account that the binding energy of He atoms in doped clusters is higher than in impurity-free clusters [19]. The interaction between a foreign article and helium atoms is stronger than the interaction between He atoms in the same cluster. The difference in binding energies significantly depends on the impurity type and is more pronounced in small-size clusters ($N \leq 100$) [19]. For example, the binding energy per atom in a helium cluster containing one SF_6 molecule is $E/N \approx 10$ K for $N \approx 100$ [19], but this energy increases to $E/N \approx 20$ K in clusters with $N \approx 40$.

Under conditions when an excited cluster have absorbed 5–10 quanta of laser radiation with a wavelength of about 10 μm, the cluster loses 1000–2000 atoms. If the clusters doped with foreign molecules in a beam initially contain $N \sim (2–3) \times 10^3$ atoms, the absorption of IR photons will lead to a significant decrease in the cluster size. Clusters containing less than 10^3 atoms may exhibit complete fragmentation upon excitation with the formation of free (completely striped) SF_6 molecules. Thus, the IR excitation under such conditions leads to a significant change in the cluster size distribution as compared to that in the initial beam, whereby selectively excited clusters will be much smaller than unexcited ones.

The next stage consists in separation of the clusters with respect to their dimensions. This is achieved via scattering of the cluster beam on a crossing molecular (or atomic) beam [31–33], which crosses the laser-excited cluster beam at a distance of $\Delta x_0 + \Delta x_1$ from the nozzle (Fig. 1). In particular, it is possible to use a crossing beam of xenon or krypton atoms. The interaction with this beam leads to separation of the cluster beam with respect to the scattering angle [31] because clusters of different sizes are deflected by unequal angles. The scattering of helium clusters on the deflecting beam involves trapping of the particles of this beam by the clusters (droplets) [31–33], whereby the moment of a trapped particle is transferred to the corresponding droplet. As a result, the droplet is heated and a certain number of He atoms is evaporated (as in the case of laser-induced heating), which is determined by the trapped particle energy. In the case of a monoatomic beam, the cluster loses about 100–200 helium atoms, while for polyatomic molecules, the number of evaporated He atoms exceeds 600 [32]. In the course of scattering on the crossing beam, liquid helium droplets of smaller size are deflected by greater angles. Therefore, selective excitation of the molecules inside liquid helium clusters by high-power IR laser radiation, followed by differential deflection of various clusters as a result of their scattering by various angles provides a means of selecting molecules embedded in liquid helium nanodroplets. Below we present the corresponding theoretical calculations, consider a particular case of embedded SF_6 molecules, and analyze the performance of the proposed method for selecting molecules with respect to the isotope composition.

3. RESULTS AND DISCUSSION

3.1. Calculation of the Cluster Size Distribution

The cluster size distribution in a beam is described by a log-normal distribution model [31, 32, 34, 35],

$$f(N) = \frac{1}{N\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln N-\mu)^2}{2\sigma^2}\right), \quad (2)$$

which is Gaussian in a coordinate system with the logarithmic abscissa scale. The mean value $\bar{N}$ and the standard deviation S are defined as

$$\bar{N} = \exp(\mu+\sigma^2/2), \quad S = \bar{N}\sqrt{\exp(\sigma^2)-1}, \quad (3)$$

where

$$\mu = \ln\left(\frac{\bar{N}^2}{\sqrt{S^2+\bar{N}^2}}\right), \quad \sigma^2 = \ln\left(\frac{S^2}{\bar{N}^2}+1\right).$$

The maximum of this distribution (N_{max}) and the $f(N_{max})$ value are given by the formulas [31, 35]

$$N_{max} = \exp(\mu-\sigma^2),$$
$$f(N_{max}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{\sigma^2}{2}-\mu\right). \quad (4)$$

The halfwidth of the asymmetric distribution

$$\Delta N_{1/2} = \exp(\mu-\sigma^2+\sigma\sqrt{\ln 4}) - \exp(\mu-\sigma^2-\sigma\sqrt{\ln 4}) \quad (5)$$

is approximately equal to the average cluster size [31].

Let us consider a log-normal distribution $f(N)$ with $N_{max} = 1800$, $\bar{N} = 2846$, and $S = 1701$. For the sake of simplicity, let the distribution be normalized to unity as

$$\int f(N)dN = 1.$$

This distribution is quite close to that obtained experimentally in [31]. A rather narrow distribution is used for obtaining a sufficiently high selectivity (see Section 3.3 below). First, we will analyze transformation of the cluster size distribution in the course of sequential (i) pick-up of molecules in the doping chamber, (ii) selective laser excitation, and (iii) trapping atoms of the deflecting beam. Then we will obtain expressions for the selectivity and efficiency of the proposed separation process.

The probabilities of trapping molecules and atoms are proportional to the cluster cross section σ_N or $N^{2/3}$ (since $\sigma_N \sim N^{2/3}$). First, consider transformation of the initial log-normal distribution as a result of doping. In the case of clusters doped by single molecules, the resulting distribution function is

$$fs(N) = f(N)\left(\frac{N}{\bar{N}}\right)^{2/3}\frac{1}{\int f(N)\left(\frac{N}{\bar{N}}\right)^{2/3}dN}, \quad (6)$$

where the last term is introduced in order to restore the normalization. The subsequent evaporation of m atoms from each cluster caused by doping with single molecules in the simplest case shifts the initial distribution so that

$$fs1(N) = fs(N+m).$$

Laser excitation leading to the evaporation of L molecules from each cluster changes the distribution to

$$fs1L(N) = fs1(N+L), \quad (7)$$

where

$$\int fs1LdN \neq 1.$$

For $L = 1200$, the value of the integral

$$\int fs1LdN = I1L$$

is approximately 0.95.

In the course of scattering on the crossing atomic beam, the distribution is modified to

$$fs1Ls(N) = fs1L(N)\left(\frac{N}{\bar{N}}\right)^{2/3}\frac{\int fs1LdN}{\int fs1L\left(\frac{N}{\bar{N}}\right)^{2/3}dN}, \quad (8)$$

where the last term is introduced (as in relation (6)) in order to restore the normalization. Analogous transformation takes place for unexcited clusters as well. A change in the cluster size distribution related to the evaporation of He atoms caused by the trapping of atoms from the deflecting beam is not taken into account because this factor does not influence the angle of cluster deflection.

Figure 2 illustrates the transformation of the cluster size distribution function in the course of processes described above, showing the initial log-normal distribution (curve *1*) and the distributions modified as a result of the pick-up of zero-energy dopant molecules ($m = 0$, curve *2*) and the trapping of both dopant molecules and atoms from a deflecting atomic beam (curve *3*). Curve *4* shows the result of the pick-up of dopant molecules and the subsequent laser excitation, while curve *5* reflects the influence of doping, the laser excitation, and the trapping of atoms of the deflecting

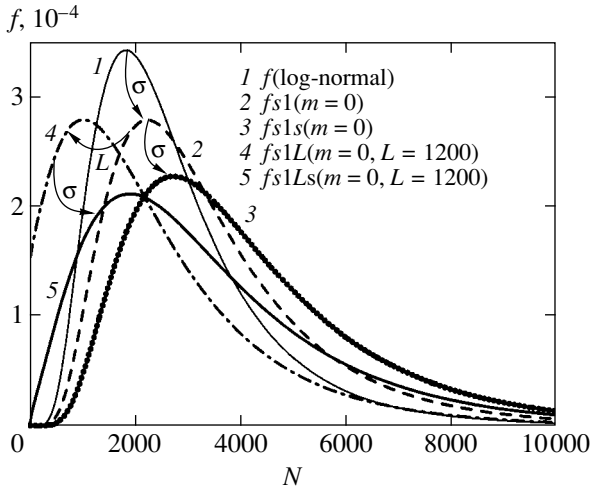

Fig. 2. The initial log-normal helium cluster size distribution (curve *1*) and its transformation in various stages of the process for $m = 0$ (curves *2*–*5*). See the text for explanations.

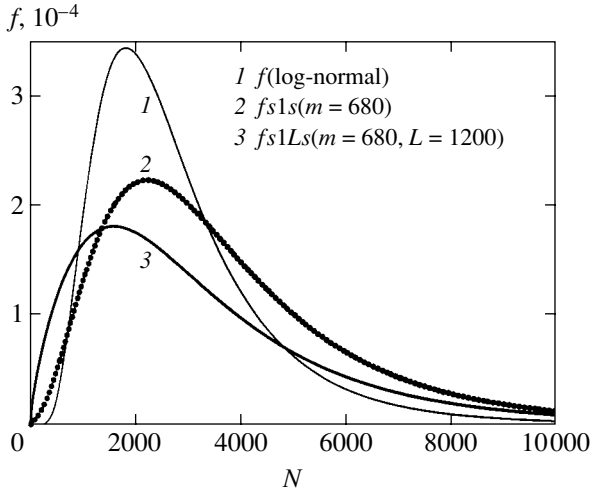

Fig. 3. The initial log-normal helium cluster size distribution (curve *1*) and its transformation (*2*) upon pick-up of an energetic doping molecule ($m = 680$) and trapping of an atom of the deflecting beam and (*3*) upon pick-up of an energetic doping molecule ($m = 680$), laser-induced excitation, and trapping of an atom of the deflecting beam.

beam. Thus, comparison of curves *3* and *5* shows the pure effect of laser excitation. As can be seen from Fig. 2, the difference between distributions *3* (without laser excitation) and *5* (after the excitation) in the ideal case ($m = 0$) is very large, especially in the region of small clusters (low N values).

Figure 3 shows the initial log-normal distribution (curve *1*) and the distributions *2*, and *3* modified by the pick-up of SF_6 molecules with $m = 680$ (these curves which are counterparts of the curves *3* and *5*, respectively, in Fig. 2). As can be seen, the laser-induced effect in this case is less pronounced, although the difference between curves *2* and *3* is still significant.

When the cluster beam (with a cluster momentum of $Nm_{\mathrm{He}}v_1$) is scattered on a monochromatic atomic beam (with a momentum of each atom m_2v_2), the angle of deflection upon scattering is (see Section 3.2)

$$\theta \approx \frac{m_2v_2}{Nm_{\mathrm{He}}v_1}, \quad N = \frac{m_2v_2}{\theta m_{\mathrm{He}}v_1} = \frac{k_0}{\theta}, \qquad k_0 = \frac{m_2v_2}{m_{\mathrm{He}}v_1}. \tag{9}$$

The transformation of the distribution of clusters with respect to sizes into their distribution with respect to deflection angles can be written as

$$f(N)dN \longrightarrow f\left(\frac{k_0}{\theta}\right)\frac{k_0}{\theta^2}d\theta = G(\theta)d\theta, \tag{10}$$

where

$$\int f(N)dN = \int G(\theta)d\theta. \tag{11}$$

In order to obtain the relations determining the selectivity and efficiency of the process, we have to take into account the following transformations,

$$fs1Ls(N) \longrightarrow G1L(\theta), \quad fs1s(N) \longrightarrow G1(\theta).$$

In the case of clusters doped with single molecules, the angle-dependent (differential) selectivity is defined by the ratio

$$S1(\theta) = \frac{G1L(\theta)}{G1(\theta)}. \tag{12}$$

Introducing integral functions of the type

$$IG(\theta)d\theta \equiv \int_{\theta}^{1} G(\zeta)d\zeta \tag{13}$$

and assuming that all doped clusters deflected by angles exceeding θ are collected, we introduce the integral selectivity as

$$IS1(\theta) = \frac{IG1L(\theta)}{IG1(\theta)}. \tag{14}$$

The efficiency of the separation process can be defined as

$$E1(\theta) = \frac{IG1L(\theta)}{IG1L(0)}. \tag{15}$$

Analogous calculations of the selectivity and efficiency have been also performed for the case of clusters trap-

ping two dopant molecules (some results of these calculations are presented below in Section 3.3).

Figure 4 shows the angular distributions of doped clusters (for $m = 0$) before (curve *1*) and after (curve *2*) laser excitation, as well as the differential (curve *3*) and integral (curve *4*) selectivities. As can be seen from these results, a rather high selectivity can be reached in the limiting (ideal) case ($m = 0$).

3.2. Application to SF_6 Molecules

For example, let us consider the case of helium clusters doped with SF_6 molecules. The content of sulfur isotopes in the natural gas is as follows (%): ^{32}S, 95.02; ^{33}S, 0.75; ^{34}S, 4.2; and ^{36}S, 0.017 [36]. The isotope shift in the ν_3 IR absorption band is about 8.5 cm^{-1} per atomic mass unit (amu) [37]. The absorption spectrum of ν_3 vibrations in $^{32}SF_6$ is well known [24–26]. The maximum of the absorption band is at 946.55 cm^{-1} and the total width is about 0.25 cm^{-1} [24–26]. This band coincides quite well with a 10.6 μm generation band of a CO_2 laser. In particular, the 10*R*(50) line of the $^{13}CO_2$ laser (946.54 cm^{-1}) and the 10*R*(35) line of the $^{16}O^{13}C^{18}O$ laser (946.64 cm^{-1}) are in good resonance with the absorption band of $^{32}SF_6$ embedded in superfluid helium droplets. Some emission lines of the CO_2 laser fall in resonance with ν_3 vibrations of the other SF_6 isotopomers occurring inside these droplets. At the same time, optimum conditions for the selective excitation of molecules embedded in superfluid helium clusters are offered by tunable IR lasers, in particular, by high-pressure CO_2 lasers with continuous frequency tuning as well as by lasers on color centers.

It should be noted that the effective excitation of clusters requires rather high laser radiation intensity, such that the excitation rate would be comparable with the energy relaxation rate. This condition can be expressed as

$$\sigma I \tau_{V-T} \approx 1, \tag{16}$$

where σ is the SF_6 absorption cross section, I is the radiation intensity [photons/(cm^2 s)], and τ_{V-T} is the characteristic time of the excitation energy relaxation. In order to provide that the clusters would be capable of absorbing on the average, for example, about five quanta, the excitation pulse duration τ_p must satisfy the condition

$$\tau_p \geq 10\tau_{V-T}.$$

Assuming that the absorption cross section of SF_6 molecules inside helium clusters is equal to the gaskinetic cross section ($\sigma \approx 2.4 \times 10^{-15}$ cm^2 [38]), and that the relaxation time is $\tau_{V-T} \approx 0.56$ ns [26], we obtain from relation (16) an estimate of the laser radiation intensity

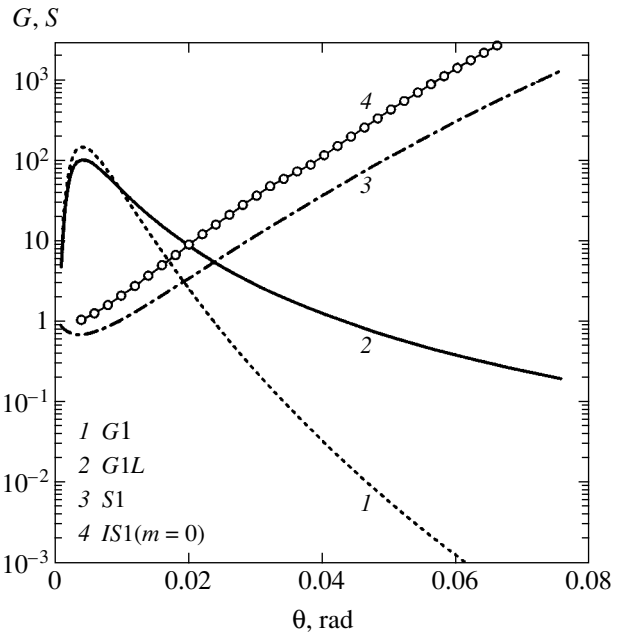

Fig. 4. The angular distributions of (*1*) unexcited and (*2*) laser-excited doped helium clusters scattered on an atomic beam. Curves *3* and *4* show the angular dependence of the differential and integral selectivity.

$I \geq 7 \times 10^{23}$ photons/(cm^2 s). Condition (16) is readily satisfied if the clusters are excited using a CO_2 laser with a pulse energy of $E \geq 10^{-2}$ J and a pulse duration on the order of 100 ns.

Now let us consider in more detail the process of cluster deviation via scattering on a crossing molecular (or atomic) beam. As was noted above, a molecule colliding with a superfluid helium droplet is trapped and its momentum is transferred to this droplet [31–33]. For a helium cluster of mass $m_1 = Nm_{He}$ and velocity v_1, colliding with a particle of mass m_2 and velocity v_2 in a beam crossing the cluster beam at an angle of α, the deflection angle θ is given by the relation

$$\tan\theta = \frac{\sin\alpha}{m_1 v_1/m_2 v_2 + \cos\alpha}. \tag{17}$$

Let the deflecting beam consist of xenon atoms with $m_2 = 131$ amu and $v_2 \approx 300$ m/s. In the case of $m_1 \gg m_2$ (e.g., $m_1 = 8000$ amu) and $v_1 \approx 1.6 v_2$ ($v_1 \approx 480$ m/s [31]), the angle θ is relatively small and can be calculated with sufficiently high precision as

$$\theta = \frac{m_2 v_2}{N m_{He} v_1}\sin\alpha = \frac{1}{N}\frac{m_2}{m_{He}}\frac{v_2}{v_1}\sin\alpha. \tag{18}$$

For the cluster beam and the deflecting beam crossing at an angle of $\alpha = 90°$, helium droplets with $N \approx 2000$ will be deflected by an angle of $\theta \approx 0.56°$ (0.01rad). In these estimates we ignored the mass of the SF_6 mole-

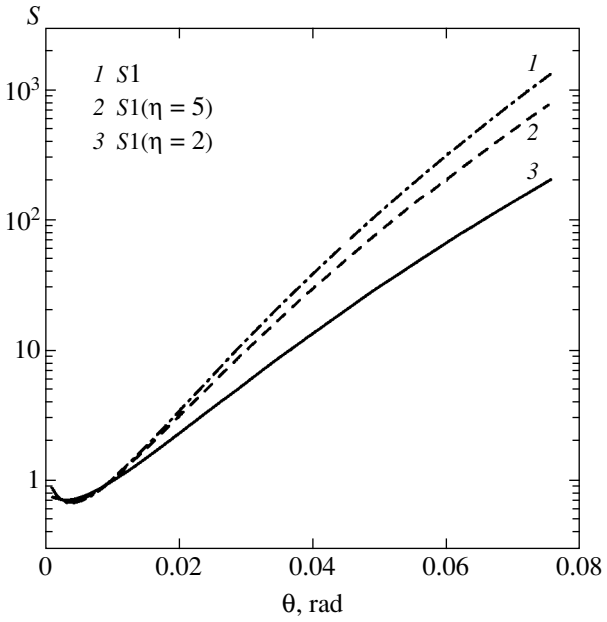

Fig. 5. The angular dependence of the differential selectivity S for a helium cluster beam scattered (*1*) on a monochromatic atomic beam ($\eta = \infty$) and (*2*, *3*) on the beams with $\eta = 5$ and 2, respectively (see text).

cule embedded in the droplet as compared to the droplet mass.

The selectivity of separation, for example, of $^{32}SF_6$ and $^{34}SF_6$ molecules, is defined as

$$S = \frac{[^{32}SF_6]^*/[^{34}SF_6]^*}{[^{32}SF_6]^0/[^{34}SF_6]^0}, \qquad (19)$$

where figures in square brackets denote the concentrations of molecules in the initial gas (index 0) and in the droplets deflected by angle θ (asterisk). The selectivity will exceed unity in the interval of angles $\theta \geq 0.01$ rad, and it is below unity for smaller deflection angles (curve *3* in Fig. 4).

It should be noted that free SF_6 molecules, appearing in the cluster beam as a result of excitation and subsequent evaporation of small doped clusters, will be reflected by xenon atoms (in cases of direct impact) by an angle of $\theta \approx 30°$. However, the probability of scattering on xenon atoms for such SF_6 molecules is much smaller than that for the clusters because of a significant difference in the corresponding cross sections: $\sigma_{SF_6} \approx 2.4 \times 10^{-15}$ cm^2 [38] versus $\sigma_N = \sigma_0 N^{2/3}$ [19] (here $\sigma_0 \approx 2.2 \times 10^{-15}$ cm^2 is the effective cross section of a He atom in clusters and σ_N is the cross section of a cluster containing N such atoms [19]). For example, the cross section of a helium cluster with $N = 1000$ is $\sigma_N \approx 2.08 \times 10^{-13}$ cm^2, while that for $N = 100$ is $\sigma_N \approx 6 \times 10^{-14}$ cm^2 [19]. Therefore, most of SF_6 molecules will be retained in the direct cluster beam, rather than deflected by the xenon beam. Large-size clusters ($N \geq 10^4$), which are deflected according to relations (17) and (18) by very small angles ($\theta \leq 0.1°$), will be also practically retained within the direct beam.

For these reasons, experiments on the separation of embedded molecules via scattering are most conveniently carried out using the beams of medium-size clusters with $\bar{N} \approx (1–3) \times 10^3$, in which case the excited clusters are deflected by relatively large angles ($\theta \geq 2°–3°$). In addition, helium atoms are more readily evaporated from such clusters because of sufficiently low binding energy. In addition, a beam of relatively small clusters has a narrower distribution width [31], which allows a higher selectivity of separation to be obtained upon scattering on a crossing beam. It should be noted that, on the other hand, the mass of liquid helium droplets is still rather large and the deflection angles upon scattering are yet quite small (1°–2°). For this reason, the experiments should be performed with well-collimated cluster beams [31]. In addition, it is necessary to provide for sufficiently large flight pathlengths ($\Delta x_2 \geq 50$ cm) after interaction with the deflecting beam, so as to reach a high angular resolution (see Fig. 1). Finally, it should be noted that, in the case of using pulsed CO_2 lasers, the experiments are expediently performed with modulated cluster beams so as to ensure approximately equal on–off (duty) ratios for the laser and cluster beams and reach a higher contrast in the selection of embedded molecules.

3.3. Factors Influencing the Selectivity and Efficiency of Separation

Let us consider the main factors influencing the selectivity and efficiency of the process under consideration. One of the factors decreasing the selectivity is the finite width of the distribution of the velocity of particles (e.g., xenon atoms) in the defecting beam. We have studied the dependence of the selectivity S on the degree of monochromaticity of this atomic beam, which was characterized by the ratio of the average atomic velocity to the scatter of these velocities:

$$\eta = \bar{v}/\Delta v.$$

The results of this analysis are presented in Fig. 5. As can be seen, the selectivity for $\eta \geq 5$ differs but little from S for the ideal monochromatic beam (cf. curves *2* and *1*). It is only rather large scatter of the atomic velocities ($\eta \leq 2$) that significantly (by a factor of approximately three) reduces the selectivity (curve *3*) as compared to that for ideal monochromatic beam. Since the values of $\eta \geq 5$ are usually readily achieved in experiment, the distribution of atomic velocities in the deflecting beam can be considered as weakly influencing the process selectivity. It should be also noted that the primary helium cluster beams are also characterized

by very narrow velocity distributions with $\eta \geq 50$ (see, e.g., [31–33]). Therefore, the influence of η on the selectivity of separation can be ignored.

The main factor decreasing the selectivity is nonselective thermal evaporation of atoms caused by the pick-up of energetic dopant molecules. This effect is illustrated in Fig. 6, where the integral selectivity and efficiency are plotted versus deflection angle for various numbers of evaporated He atoms (m = 0, 200, and 680). As can be seen, S sharply decreases with increasing m. These results indicate that, in the case of large embedded molecules (such as SF_6, OsO_4, and UF_6), it is expedient to cool the gas in the doping chamber. For example, the trapping of SF_6 molecules at 150 K leads to the evaporation of only 160–170 atoms from a helium droplet. According to Fig. 6, the integral selectivity for $\theta \geq 0.04$ rad in this case is almost ten times greater as compared to the S value upon SF_6 trapping at room temperature.

Another factor that may lead to a certain decrease in the selectivity is a finite size Δx_{int} of the region of interaction between a doped cluster beam and the deflecting atomic beam (typically, $\Delta x_{int} \approx 1$ cm). However, for sufficiently large flight pathlengths ($\Delta x_2 \geq 50$ cm), this decrease in S is rather insignificant (on the order of $\Delta x_{int}/\Delta x_2$) and can be further reduced by increasing Δx_2.

Now let us proceed to the analysis of factors influencing the efficiency of separation. A decrease in the efficiency (as well as of the selectivity) is primarily related to the fact that the probability of trapping dopant molecules by helium nanodroplets obeys the Poisson distribution [32],

$$P(N) = N_0 \frac{(n\sigma L)^N}{N!} \exp(-n\sigma L), \qquad (20)$$

where N_0 is the amplitude factor, n is the number density of molecules in the doping chamber, σ is the trapping cross section, and L is the length of the interaction region. For these reasons, a certain fraction of clusters of the primary beam will trap single dopant molecules. Depending on the experimental conditions (gas pressure and the length of the interaction region in the doping chamber), the cluster beam may also contain a relatively large fraction of droplets trapping two or more SF_6 molecules, which form SF_6 clusters inside these droplets. It should be noted that helium droplets trapping more than one SF_6 molecule will have smaller dimensions compared to the size of droplets containing a single dopant molecule, because a greater number of He atoms is evaporated in the former case. Upon scattering of the cluster beam on a crossing atomic beam, these smaller clusters will be deflected by greater angles, which may significantly decrease the selectivity. In order to avoid this, experiments should be performed under conditions that minimize the fraction of helium nanodroplets trapping more than one dopant molecule. However, this will also significantly reduce the total number of doped helium clusters in the beam, thus naturally decreasing the efficiency of the process. For example, our calculations showed (Fig. 6) that, if the fraction of clusters trapping two SF_6 molecules is about half the number of clusters doped with a single molecule, both selectivity and efficiency (for $\theta \geq 0.02$ rad) for the former fraction are also about twice as small as those for the singly doped droplets.

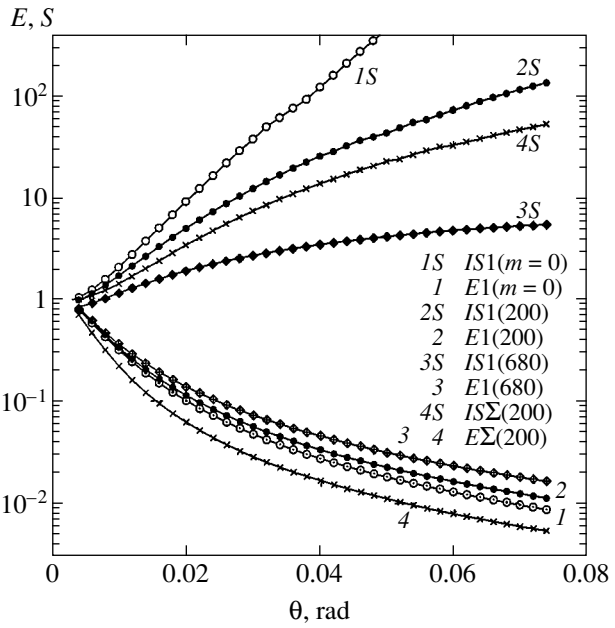

Fig. 6. The angular dependences of the efficiency E (curves *1–3*) and integral selectivity S (curves *1S–3S*) for doped helium clusters losing different numbers of atoms evaporated upon doping: m = 0 (*1*), 200 (*2*), and 680 (*3*). Curves *4* and *4S* show the efficiency and selectivity, respectively, for a beam containing clusters doped with one and two molecules in the 2:1 ratio (for m = 200).

Finally, let us consider the possible influence of double collisions of the doped helium clusters with atoms of the deflecting beam on the efficiency and selectivity of the separation process. Such double collisions, which become possible under the conditions of a sufficiently high density of doped clusters and/or large length of the interaction region (Δx_{int}), obviously affect both the efficiency and selectivity of the process. Indeed, the effect of double collisions is essentially the same as if the beam would be scattered twice from the deflecting atomic beam. In the corresponding calculation, we have to take into account both the evaporation of about 200 helium atoms upon trapping of the first xenon atom ($m \approx 200$ for atoms trapped from a xenon beam [32]) and the subsequent transformation of the distribution function (multiplied by $N^{2/3}$) upon trapping of the second atom.

We omit rather lengthy formulas describing the sequential transformation of the initial distribution and present here only the most pessimistic result obtained

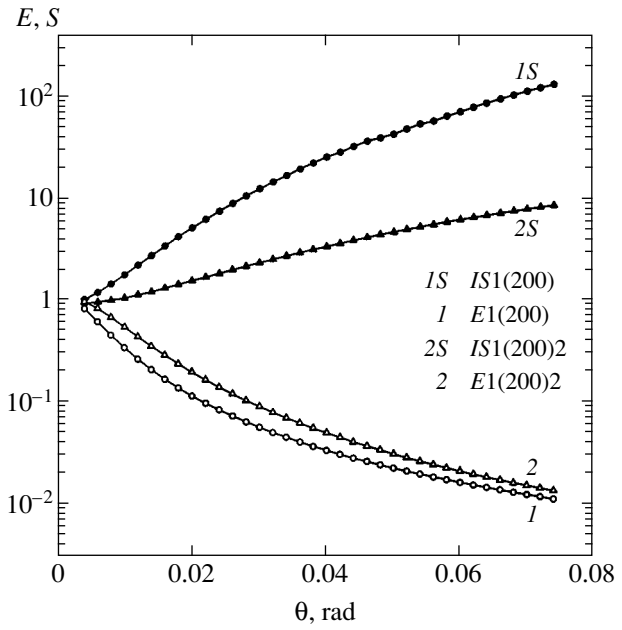

Fig. 7. The angular dependences of the integral selectivity S (curves *1S* and *2S*) and efficiency E (curves *1* and *2*) for doped helium clusters scattered on a monochromatic atomic beam in the case of (*1S*, *1*) single collisions and (*2S*, *2*) double collisions. Helium clusters lost m = 200 atoms upon doping and m_1 = 200 atoms upon trapping atoms in the deflecting beam.

upon calculation taking into account the aforementioned factors and the conversion of the size distribution into the angular distribution. Note that, since the total deflection angle upon double collisions is approximately twice as large as that upon single collisions, the quantity k_0 in relations (9) and (10) is replaced by $2k_0$. The values of selectivity and efficiency obtained with allowance for all the above factors are presented in Fig. 7 (curers *2S* and *2*, respectively) in comparison with the corresponding results for single collisions (curves *1S* and *1*). As can be seen, the double collisions significantly reduce the selectivity and slightly increase the efficiency. For example, at an equal efficiency of $E = 0.1$, the selectivity in the case of double collisions decreases by a factor of about 2.5 as compared to that in the regime of single collisions. Therefore, in order to increase the selectivity, the experiments have to be carried out under the conditions minimizing the probability of double collisions (e.g., for a sufficiently low density of atoms in the deflecting beam).

3.4. Estimates of the Yield of Enriched Products

In view of the potential applicability of the proposed method for the separation of isotopes, it is of interest to estimate the yield of enriched products. Of course, we are speaking of very rough estimates. The values presented below cannot be used as the initial data for calculating the efficiency of a real separation process. Such calculations can be based only on reliable experimental results. Nevertheless, we believe that even rough estimates of the product yield are expedient for assessing the method.

The evaluations below proceed from the intensities of real helium cluster beams and the capacity of diffusion pumps used in real systems [5, 9]. It should be noted that, owing to low binding energies of atoms in small helium clusters (see Section 2), the formation of liquid helium droplets usually requires high densities and low temperatures of gas in the expansion region [5], both these factors leading to a sufficiently high gas flow rate. For this reason, the setups producing helium cluster beams employ effective pumps (capable of ensuring high pumping speeds) and/or nozzles with very small apertures. The gas flow rate is proportional to the product $p_0 d^2 T_0^{-1/2}$, where d is the nozzle output diameter and p_0 and T_0 are the gas pressure and temperature in front of the nozzle. In the typical setups producing liquid helium nanodroplets, the nozzle diameter is within d = 5–20 μm, the gas pressure ranges within p_0 = 1–20 bar, and the temperature is T_0 = 10–30 K [5, 9]. In the ideal gas approximation, a nozzle with d = 10 μm operating at $T_0 \approx 20$ K and $p_0 \approx 15$ bar provides for a gas flow equivalent to that formed by a nozzle with d = 50 μm operating at room temperature and a pressure of 2 bar [5], which amounts to approximately 4 (bar cm^3)/s (≈0.16 mmol/s). In order to pump this flow, a diffusion pump must ensure a pumping speed of about 10000 l/s for He (≈4000 l/s for nitrogen) at a limiting working pressure of about 3×10^{-4} Torr. Under such conditions, it is possible to obtain liquid helium nanodroplets containing several thousand He atoms [5].

For the evaluation, we use the above estimate of the gas flow rate—namely, 4 (bar cm^3)/s (which corresponds to a particle flux of about 10^{20} s^{-1})—and consider a beam of clusters with an average size of $\bar{N} = 10^3$. If all clusters in the beam were doped with single SF_6 molecules, the total SF_6 flux would be equal to 10^{17} s^{-1}. However, let us take into account factors decreasing this value, for example, in a system with a two-dimensional (slit) nozzle with a slit width of 50 cm (such nozzles were used in investigations of the excitation and isotope-selective multiphoton IR dissociation (ISMD) of UF_6 [39] and SF_6 [40]. For the proposed separation method, it also desirable to use slit nozzles (see Section 3.3). In the case of a narrow directed beam, it is possible to select about 10% of clusters from the initial flux, and no more that 20% of clusters can be doped with single molecules.

Let us consider the case of a continuous cluster beam excited by IR radiation of a CO_2 laser operating at a pulse repetition rate of $f \approx 500$ Hz (lasers operating in this regime are used in practice for the laser separation of carbon isotopes [41, 42]). The fraction of laser-

excited clusters in the beam (i.e., the on–off ratio of the excitation process) is

$$c_f = \frac{\Delta x_{\rm irr} f}{v_1},$$

where $\Delta x_{\rm irr}$ is the length of the laser-irradiated region and v_1 is the velocity of the cluster beam. The latter value is typically about $v_1 \approx 500$ m/s [31, 32]. If the laser pulse energy is about 2 J and the laser beam is perpendicular to the cluster beam, it is easy to provide for an irradiated region with $\Delta x_{\rm irr} \approx 20$ cm (or a laser beam cross section of 20×1 cm^2 in this region). The resulting energy flux is about 0.1 J/cm^2, which is one order of magnitude greater than the value required for the effective excitation of clusters (see relation (16)). Thus, the on–off ratio of the excitation process is $c_f \approx 0.2$.

Finally, we have to take into account that no more than 20% of doped clusters can be deflected by single collisions in the course of scattering on the atomic beam. Therefore, we eventually deal with approximately a 10^{-3} fraction of the initial beam, or with a flux of SF_6 molecules about 10^{14} s^{-1}. Taking into account that, for the integral selectivity $IS \approx 10$ the process selectivity is $E \approx 0.1$ (see Fig. 6, curves *2S* and *2* for m = 200), the yield of SF_6-enriched product will amount to about 10^{13} s^{-1} or 3.6×10^{16} h^{-1}, which is equivalent to approximately 0.01 mg/h. Obviously, this is a very low yield, even though we have considered a small setup with a single nozzle and a single pump.

It should be noted that small yields of enriched product are characteristic, in particular, of the process of isotope separation by ISMD in gasdynamically cooled jets and flows. For example, the laser separation of uranium isotopes using ISMD of UF_6 molecules employs highly diluted mixtures with a relative density of UF_6 molecules in the carrier gas below 0.01 [39, 43]. The effective excitation and dissociation of molecules requires rather high laser energy density (at least, exceeding 1 J/cm^2 for the dissociation of preliminarily excited molecules), which leads to a decrease in the on–off ratio of the excitation process down to $c_f \le 0.001$. High selectivity ($s \approx 3$–5) can probably be reached only at the expense of a low yield (≤0.01). However, in the case of isotope separation by ISMD in gasdynamically cooled flows, it is possible to increase the gas pressure in the doping chamber by two orders of magnitude (up to 10^{-2} Torr). This implies that the gas flux also increases by two orders of magnitude to reach about 10^{22} s^{-1}. Thus, our estimates obtained taking into account all the factors mentioned above show that the yield of enriched products can amount to about 10^{15} s^{-1} or 3.6×10^{18} h^{-1}, which is about two orders of magnitude higher compared to the case of the cluster beam. On the other hand, it should be noted that highly selective dissociation in the case of molecules characterized by small isotope shifts can be achieved only in high-frequency IR fields, which significantly complicates the process of isotope separation using the ISMD method. In addition, it is necessary to provide high energy densities and ensure a much greater (at least by two orders of magnitude) contribution of laser photons to the dissociation of molecules. This is related to rather high dissociation energy and a low selectivity of this process, whereby a considerable fraction of laser photons is spent for the excitation of molecules containing isotopes other than the target one. In the case of cluster beams, the selective excitation and dissociation of clusters only requires a single-mode radiation of a moderate-power laser. For this reason, one laser can irradiate relatively greater volumes as compared to those in the case of the ISMD process. The selectivity of excitation in helium cluster beams can be also higher.

In connection with the above evaluations, it should be noted that the development of commercial systems for the laser separation of uranium isotopes using the ISMD of UF_6 molecules is considered unprofitable (for the modern level of laser technologies) [44]. Further investigations (using the existing experimental facilities) are necessary for the entire set of problems pertaining to the development of laser systems and the optimization of gasdynamic flows. The method under consideration can probably find use for the laser separation of exotic molecules and/or small amounts of some isotopes. However, it is not excluded that further progress in technology will make possible wide use of such methods in the future.

4. CONCLUSIONS

The results of our investigation show that, using selective vibrational excitation of superfluid helium clusters containing embedded molecules by intense IR laser radiation, followed by separation of the clusters with respect to size via scattering of the cluster beam on a crossing atomic beam, it is possible to select the embedded molecules with respect to isotope (elemental) composition. In order to increase the process efficiency, it is recommended to obtain the initial helium cluster beams using slit nozzles and/or slit diaphragms, which allow the beams with large transverse dimensions and large number of liquid helium nanodroplets of optimum size to be obtained.

A significant advantage of using superfluid helium droplets (clusters) for selecting molecules are (in addition to a low temperature of clusters and the related very small width of the optical absorption bands of embedded molecules) a low binding energy of He atoms in clusters (≤7.2 K [19]) and the free rotation of molecules inside the clusters. Because of the low binding energy, the absorption of a single IR photon leads to a significant decrease in the cluster size. A reduction in the cluster size leads to further decrease in the binding energy and, in turn, decrease requirements on the pumping energy.

In contrast to the method based on the ISMD of molecules (see, e.g., [45–47] and references therein), which is applicable only to polyatomic molecules, the method under consideration can also be used for selecting small (e.g., diatomic) molecules. Owing to the aforementioned free rotation, the IR absorption spectra of small (e.g., di- and triatomic) molecules embedded in superfluid helium droplets contain narrow vibrational–rotational lines [1, 5, 9], which makes possible highly selective excitation of the molecules. In the course of doping of helium clusters with such small molecules, their size distribution is not very strongly modified (as compared to the case of large molecules) and the laser excitation of such doped clusters can be performed with a higher selectivity (Fig. 6).

Another advantage of the separation method considered in this study is the possibility of exciting molecules embedded in liquid helium droplets by means of microwave radiation. The results of experiments show (see, e.g., [5. 9] and references therein) that use of the microwave radiation for the excitation of molecules inside the droplets and the evaporation of He atoms from the excited molecules lead to the same effects as the IR laser radiation. The main disadvantages of the proposed method are a rather complicated procedure and a relatively low yield. It should be noted that detailed description of the methods of formation of liquid helium nanodroplets and the methods of introducing single molecules (and clusters) into these droplets can be found in [1, 2, 5, 9].

ACKNOWLEDGMENTS

The authors are grateful to the referee for valuable remarks. This study was supported in part by the Russian Foundation for Basic Research, project no. 03-02-17067.

REFERENCES

1. J. P. Toennies, A. F. Vilesov, and K. B. Whaley, Phys. Today **54**, 31 (2001).
2. E. Lugovoj, J. P. Toennies, S. Grebenev, *et al.*, in *Spectroscopy of Single Molecules and Clusters inside Superfluid Helium Droplets, in Atomic and Molecular Beams, The State of the Art 2000*, Ed. by R. Compargue (Springer, Berlin, 2000), p. 755.
3. S. Grebenev, M. Hartmann, A. Lindinger, *et al.*, Physica B (Amsterdam) **280**, 65 (2000).
4. S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, Science **289**, 1532 (2000).
5. C. Callegary, K. L. Lehmann, R. Schmied, and G. Scoles, J. Chem. Phys. **115**, 10090 (2001).
6. F. Stienkemeier and A. F. Vilesov, J. Chem. Phys. **115**, 10119 (2001).
7. S. Grebenev, J. P. Toennies, and A. F. Vilesov, Science **279**, 2083 (1998).
8. J. P. Toennies and A. F. Vilesov, Annu. Rev. Phys. Chem. **49**, 1 (1998).
9. G. N. Makarov, Usp. Fiz. Nauk **174**, 225 (2004) [Phys. Usp. **47**, 217 (2004)].
10. Y. Kwon, P. Huang, M. V. Patel, *et al.*, J. Chem. Phys. **113**, 6469 (2000).
11. J. A. Northby, J. Chem. Phys. **115**, 10065 (2001).
12. K. Nauta and R. E. Miller, Science **283**, 1895 (1999).
13. K. Nauta and R. E. Miller, Science **287**, 293 (2000).
14. E. Lugovoj, J. P. Toennies, and A. F. Vilesov, J. Chem. Phys. **112**, 8217 (2000).
15. N. Portner, A. F. Vilesov, and M. Havenith, Chem. Phys. Lett. **368**, 458 (2003).
16. V. N. Lokhman and G. N. Makarov, Chem. Phys. Lett. **398**, 453 (2004).
17. M. Pi, R. Maiol, and M. Barranco, Phys. Rev. Lett. **82**, 3093 (1999).
18. N. Portner, J. P. Toennies, and A. F. Vilesov, J. Chem. Phys. **117**, 6054 (2002).
19. S. A. Chin and E. Krotschek, Phys. Rev. B **52**, 10405 (1995).
20. P. Zindzingre, M. L. Klein, and D. M. Ceperley, Phys. Rev. Lett. **63**, 1601 (1989).
21. M. Casas, F. Dalfovo, A. Lastri, *et al.*, Z. Phys. D **35**, 67 (1995).
22. M. V. Rama Krishna and K. B. Whaley, J. Chem. Phys. **93**, 746 (1990).
23. P. L. Kapitsa, Zh. Éksp. Teor. Fiz. **11**, 581 (1941).
24. M. Hartmann, R. E. Miller, J. P. Toennies, and A. F. Vilesov, Phys. Rev. Lett. **75**, 1566 (1995).
25. J. Harms, M. Hartmann, J. P. Toennies, and A. F. Vilesov, J. Mol. Spectrosc. **185**, 204 (1996).
26. M. Hartmann, N. Portner, B. Sartakov, *et al.*, J. Chem. Phys. **110**, 5109 (1999).
27. D. Brink and S. Stringari, Z. Phys. D **15**, 257 (1990).
28. K. Nauta and R. E. Miller, J. Chem. Phys. **115**, 8384 (2001).
29. F. Madeja, P. Markwick, M. Havenith, *et al.*, J. Chem. Phys. **116**, 2870 (2002).
30. F. Dalfovo and S. Stringari, J. Chem. Phys. **115**, 10078 (2001).
31. M. Lewerenz, B. Schilling, and J. P. Toennies, Chem. Phys. Lett. **206**, 381 (1993).
32. M. Lewerenz, B. Schilling, and J. P. Toennies, J. Chem. Phys. **102**, 8191 (1995).
33. J. Gspann, Z. Phys. B: Condens. Matter **98**, 405 (1995).
34. J. Harms, J. P. Toennies, and F. Dalfovo, Phys. Rev. B **58**, 3341 (1998).
35. H. Pauly, *Atom, Molecule, and Cluster Beams II: Cluster Beams, Fast and Slow Beams, Accessory Equipment and Applications* (Springer, New York, 2000).
36. *CRC Handbook of Chemistry and Physics*, Ed. by D. R. Lide, 74th ed. (CRC Press, Boca Raton, 1993–1994).

37. R. S. McDowell, J. B. Krohn, H. Flicker, and M. C. Vasquez, Spectrochim. Acta A **42**, 351 (1986).

38. I. Burak, J. I. Steinfeld, and D. G. Sutton, J. Quant. Spectrosc. Radiat. Transf. **9**, 959 (1969).

39. Y. Okada, Y. Kuga, K. Sunouchi, *et al.*, Appl. Phys. B **59**, 475 (1994).

40. G. N. Makarov, E. Ronander, S. P. van Heerden, *et al.*, Appl. Phys. B **65**, 583 (1997).

41. V. Yu. Baranov and A. P. Dyad'kin, in *Isotopes: Properties, Production, Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), p. 343 [in Russian].

42. A. P. Dyad'kin, V. B. Zarin, V. B. Laptev, *et al.*, in *Proceedings of the VIII All-Russian (International) Conference on Physicochemical Processes under Atomic and Molecular Selection, Zvenigorod, 2003*, Ed. by Yu. A. Kolesnikov (TsNIIatominform, Moscow, 2003), p. 121.

43. K. Takeuchi, H. Tashiro, S. Kato, *et al.*, J. Nucl. Sci. Technol. **26**, 301 (1989).

44. V. Yu. Baranov, E. I. Kozlova, Yu. A. Kolesnikov, and A. A. Kotov, in *Isotopes: Properties, Production, Application*, Ed. by V. Yu. Baranov (IzdAT, Moscow, 2000), p. 357 [in Russian].

45. G. N. Makarov and A. N. Petin, Zh. Éksp. Teor. Fiz. **119**, 5 (2001) [JETP **92**, 1 (2001)].

46. G. N. Makarov and A. N. Petin, Chem. Phys. **266**, 125 (2001).

47. G. N. Makarov, Usp. Fiz. Nauk **173**, 913 (2003) [Phys. Usp. **46**, 889 (2003)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 516–520.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 583–588.

ATOMS, MOLECULES, OPTICS

Double Ionization of Helium by Relativistic Highly Charged Ion Impact¶

A. B. Voitkiv*, B. Najjari, and J. Ullrich
Max-Planck Institut für Kernphysik D-69117, Heidelberg, Germany
**e-mail: Alexander.Voitkiv@mpi-hd.mpg.de*
Received October 1, 2004

Abstract—We consider an interesting realization of the fundamental four-body problem: double ionization of helium in superintense electromagnetic fields generated by highly charged ions in relativistic collisions. We show how the simultaneous interaction of such fields with all three target constituents (which is not described by first-order theory) strongly influences the collision dynamics even at very high collision energies and how a "genuine" photoemission-like pattern may emerge in collisions at extreme relativistic energies. A very good agreement with available experimental data is obtained. *© 2005 Pleiades Publishing, Inc.*

The question of the dynamics of quantum mechanical few-particle systems on various time scales is among the most interesting topics in modern atomic, molecular, and optical physics [1]. One of fundamental examples of the quantum few-body problem is given by ionization of helium in collisions with fast highly charged ions. During the last decade, there has been remarkable progress in this field [1, 2]. Most of the studies of helium ionization, however, have been performed for single ionization and for nonrelativistic collision velocities.

Whereas single ionization of helium is normally treated as a three-body problem (projectile, "active" electron, and recoil ion), double ionization represents a particularly strong challenge for theory because it is a pure four-body problem. Indeed, a satisfactory (but still incomplete) understanding of helium double ionization by charged projectiles has been reached only for collisions with sufficiently fast electrons where the first Born approximation (FBA) in the projectile–target interaction is valid. Helium double ionization by highly charged ions is more difficult to describe, and it has attracted much less attention so far. In particular, helium double ionization by relativistic ions with such a high charge Z_p that $Z_p/v_p \sim 1$ even for collision velocities v_p approaching the speed of light c ($c \approx 137$ a.u.) has remained terra incognita to a large extent.

The first measurements of differential cross sections for double ionization of helium in relativistic collisions (1 GeV/u U^{92+}, $v_p = 120$ a.u., $\gamma = (1 - v_p^2/c^2)^{-1/2} \approx 2$, and $Z_p/v_p \approx 0.77$) were performed in [3]. Detailed experimental studies of helium ionization by highly charged ions in collisions at $\gamma \approx 1.5$–2 are scheduled for 2005 (GSI, Germany) and collision energies up to those corresponding to $\gamma \approx 30$ will become routinely accessible for atomic physics experiments in the near future [4].

Relativistic collisions with ions like U^{92+} may expose helium atoms to extreme conditions. Indeed, rough estimates show that electromagnetic pulses with effective power densities as high as 10^{19} to 10^{23} W/cm^2 can be generated by relativistic highly charged ions in collisions at $\gamma \sim 10$–30 for impact parameters between 2 and 10 a.u. such that the whole target atom is exposed to a nearly homogeneous field. In addition, such pulses are ultrashort and, despite the enormous intensities, may "gently" irradiate the target, making its "snapshots" on the subatomic time scale.

Only a few attempts have been made to evaluate differential cross sections for double ionization of helium in relativistic collisions with highly charged ions. Estimates in [3] and [5] were based on the Weizsäcker–Williams method of equivalent photons. However, for collisions with light targets, strictly speaking, this method may be applied only at extreme relativistic energies [6]. In addition, the results in [5] were obtained only for a fixed collision impact parameter and cannot therefore be related to experiment. In [7], helium ionization was treated using the classical-trajectory Monte Carlo approach. However, cross sections reported in [7] were too small because this approach fails to properly describe collisions with relatively small momentum transfers, which become of great importance at very high impact energies.

In this paper, we consider helium double ionization in relativistic collisions with very highly charged ions by developing an approach that, for the first time, enables a detailed description of this extraordinary case of the four-body quantum dynamical problem.

We start with the following remarks. First, even in collisions with relativistic projectiles, the overwhelming majority of electrons emitted from light targets have

¶ This article was submitted by the authors in English.

nonrelativistic energies.[1] Therefore, we consider helium ionization in the target frame and use a nonrelativistic description for the electron motion. Second, because the momentum exchange does not actually exceed several atomic units in collisions of interest for the present study, the recoil velocity of the target nucleus and the deflection angle of the projectile are always very small. This allows us to begin the consideration with the semiclassical picture in which (i) only the electrons are treated quantum mechanically; (ii) the target nucleus is assumed to be at rest and is taken as the origin of the target frame; and (iii) in this frame, the projectile moves along a straight-line classical trajectory $\mathbf{R}(t) = \mathbf{b} + \mathbf{v}_p t$, where $\mathbf{b}$ is the impact parameter. The corresponding Schrödinger equation is

$$i\frac{\partial \Psi}{\partial t} = \left\{\sum_{j=1}^{2}\left[\frac{1}{2}\left(\hat{\mathbf{p}}_j + \frac{\mathbf{A}_j}{c}\right)^2 - \varphi_j\right] + V_{at} + Z_{t\varphi}\right\}\Psi. \quad (1)$$

Here, $\hat{\mathbf{p}}_j$ is the momentum operator for the jth atomic electron, φ_j and $\mathbf{A}_j$ are the scalar and vector potentials of the projectile field at the position of the jth atomic electron, and φ is the corresponding scalar potential at the origin. Furthermore, $Z_t = 2$ is the charge of the target nucleus and $V_{at} = -Z_t/r_1 - Z_t/r_2 + 1/r_{12}$ is the interaction between the target particles, where $\mathbf{r}_j$ is the coordinate of the jth electron with respect to the target nucleus and $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$. The spin-flip transitions are suppressed in our case by a factor $\sim v_p/c^2$ compared to the non-spin-flip ones, and the spin terms are therefore ignored in Eq. (1).

In the Lorentz gauge, the projectile potentials are given by [8]

$$\varphi_j = \frac{\gamma Z_p}{s_j}, \quad \varphi = \frac{\gamma Z_p}{s}, \quad \mathbf{A}_j = \frac{\mathbf{v}_p}{c}\varphi_j, \quad (2)$$

where $\mathbf{s}_j$ and $\mathbf{s}$ are the coordinates of the jth target electron and the target nucleus with respect to the projectile ion given in the projectile rest frame.

Taking into account that in both the initial and final channels the projectile velocity is much higher than typical electron velocities (1–3 a.u.), we use the symmetric eikonal approximation (SEA). In the SEA, the state Ψ is replaced by Ψ_i and Ψ_f in the initial and final channels respectively, where

$$\begin{aligned}\Psi_i(t) &= \psi_i(\mathbf{r}_1, \mathbf{r}_2)\exp(-i\varepsilon_i t)(vs + \mathbf{v}\cdot\mathbf{s})^{i\eta_p} \\ &\times (vs_1 + \mathbf{v}\cdot\mathbf{s}_1)^{-i\nu_p}(vs_2 + \mathbf{v}\cdot\mathbf{s}_2)^{-i\nu_p}, \\ \Psi_f(t) &= \psi_f(\mathbf{r}_1, \mathbf{r}_2)\exp(-i\varepsilon_f t)(vs - \mathbf{v}\cdot\mathbf{s})^{-i\eta_p} \\ &\times (vs_1 - \mathbf{v}\cdot\mathbf{s}_1)^{i\nu_p}(vs_2 - \mathbf{v}\cdot\mathbf{s}_2)^{i\nu_p},\end{aligned} \quad (3)$$

[1] Actually, in the target frame, the energies of most emitted electrons do not exceed few atomic units (see Fig. 2).

ψ_i and ψ_f are the initial and final states of the three-body target subsystem with energies ε_i and ε_f, $\nu_p = Z_p/v_p$, and $\eta_p = Z_p Z_t/v_p$. We note that in Eq. (3), the Coulomb boundary conditions (due to the projectile field) are satisfied for all the three target particles.

Within the SEA, the prior form of the semiclassical transition amplitude is

$$a_{fi}(\mathbf{b}) = -i\int_{-\infty}^{\infty}(dt\langle\Psi_f(t)|\hat{W}(t)|\Psi_i(t)\rangle), \quad (4)$$

where the distortion interaction $\hat{W}(t)$ is given by

$$\begin{aligned}\hat{W}\Psi_i = (vs_1 - \mathbf{v}\cdot\mathbf{s}_1)^{-i\nu_p}(vs_2 - \mathbf{v}\cdot\mathbf{s}_2)^{-i\nu_p}(vs - v\cdot\mathbf{s})^{i\eta_p} \\ \times \nu_p\exp(-i\varepsilon_i t)\sum_{j=1}^{2}(\mathbf{C}_j\cdot\hat{\mathbf{p}}_j + \nu_p D_j)\psi_i,\end{aligned} \quad (5)$$

with

$$\begin{aligned}\mathbf{C}_j &= -s_j^{-1}(s_{j,x}(s_j + s_{j,z})^{-1}; s_{j,y}(s_j + s_{j,z})^{-1}; \gamma^{-1}), \\ D_j &= (s_j(s_j + s_{j,z}))^{-1} - 0.5\,v_p^2(cs_j)^{-2},\end{aligned} \quad (6)$$

where $s_{j,z} = \mathbf{s}_j\cdot\mathbf{v}_p/v_p$ and $(s_{j,x}; s_{j,y}) = s_j - s_{j,z}\mathbf{v}_p/v_p$.

The full quantum dynamics of the collision cannot be treated with semiclassical amplitude (4). However, for collisions with very small projectile scattering angles and negligible velocities of the target nucleus, the quantum transition amplitude S_{fi} can be obtained from semiclassical amplitude (4) as

$$S_{fi}(\mathbf{Q}) = \frac{1}{2\pi}\int d^2b\exp(i\mathbf{Q}\cdot\mathbf{b})a_{fi}(\mathbf{b}), \quad (7)$$

where $\mathbf{Q}$ is the two-dimensional transverse part ($\mathbf{Q}\times\mathbf{v}_p = 0$) of the momentum transfer $\mathbf{q}$ to the target. In contrast to the impact parameter $\mathbf{b}$, the momentum transfer is accessible to direct measurement. We have $\mathbf{q} = (\mathbf{Q}; q_{\min})$, where $q_{\min} = \omega_{fi}/v_p$ with $\omega_{fi} = \varepsilon_f - \varepsilon_i$.

Amplitude (7) is the first term of the symmetric eikonal distorted wave series. The analysis shows that for the most important part of the emission, the expansion parameter of this series is essentially given by $\varsigma = Z_p/v_p^2$. In relativistic collisions, ς does not exceed 0.01 even for the highest possible projectile charge states $Z_p \sim v_p$. Therefore, the first term of this series alone may already be sufficient for a successful treatment of the collision dynamics. This is to be contrasted with the standard Born series, which is generated from Eq. (1) in the usual way and has the expansion parameter $\nu_p = Z_p/v_p$. In collisions with the heaviest bare nuclei, the parameter ν_p is never much less than unity. Therefore,

not only might the first Born approximation be insufficient[2] but also the whole Born series is likely to become meaningless.

The success of distorted wave models for nonrelativistic ion–atom collisions was to a very large extent caused by the facts that (i) the interaction between the projectile and the target nucleus (the *n–n* interaction) does not affect the electron emission spectra integrated over the projectile deflection angle and (ii) for collisions with hydrogen-like targets, the transition amplitude of type (7) can be evaluated analytically provided the *n–n* interaction is ignored. Taking into account the interaction between the projectile and the second "active" target electron tremendously complicates calculations, and the situation is certainly not simplified if the *n–n* interaction must also be included, for instance, in the case where the full collision dynamics has to be considered.

At $\nu_p \sim 1$, a direct numerical integration of the multiple integral in Eq. (7) faces difficulties because in both the initial arid final channels, the motion of the projectile is not bounded in space. Therefore, the integral over $d^3R = d^3b v_p dt$ in Eq. (7) is not absolutely convergent and should be taken analytically. The result is

$$S_{fi}(\mathbf{Q}) = \frac{i\nu_p}{2\pi v_p \gamma} \times \int d^2\boldsymbol{\zeta} d^2\boldsymbol{\xi} \langle \psi_f | \mathbf{G}_1 \cdot \hat{\mathbf{p}}_1 + \mathbf{G}_2 \cdot \hat{\mathbf{p}}_2 + F_1 + F_2 | \psi_i \rangle. \quad (8)$$

Here, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ are two-dimensional vectors perpendicular to $\mathbf{v}_p$,

$$\mathbf{G}_1 = \mathbf{G}(\nu_p, \eta_p; \gamma; \boldsymbol{\zeta}, \boldsymbol{\xi}, \mathbf{q}; \mathbf{r}_1, \mathbf{r}_2),$$

$$F_1 = F(\nu_p, \eta_p; \gamma; \boldsymbol{\zeta}, \boldsymbol{\xi}, \mathbf{q}; \mathbf{r}_1, \mathbf{r}_2),$$

$\mathbf{G}_2 = \mathbf{G}_1(\mathbf{r}_1 \longleftrightarrow \mathbf{r}_2)$, and $F_2 = F_1(\mathbf{r}_1 \longleftrightarrow \mathbf{r}_2)$, where $\mathbf{G}$ and F are expressions containing exponential, gamma, and hypergeometric functions. The explicit forms of $\mathbf{G}$ and F are very cumbersome and will be given elsewhere.

We note that the right-hand side of Eq. (5) was written with ψ_i assumed to be an exact state of the free target. If this is not the case, an additional term appears in the right-hand side of Eq. (5). However, if $\varepsilon_f \neq \varepsilon_i$, this term makes zero contribution to the transition amplitude. Therefore, there are no formal restrictions imposed on ψ_i and ψ_f by the use of Eq. (8). Because the three-body problem has no exact solution, the actual choice of ψ_i and ψ_f is dictated by two main points: these states should be "sufficiently good" and, simultaneously, allow performing at least the ten-fold integration in Eq. (8) necessary to obtain the fully differential cross section

$$\frac{d\sigma}{d^2Q d^3k_1 d^3k_2} = |S_{fi}|^2, \quad (9)$$

where $\mathbf{k}_1$ and $\mathbf{k}_2$ are the electron momenta in the final state.

As was already remarked, the SEA is superior to the FBA at $\nu_p = Z_p/v_p \sim 1$. One would expect that as $\nu_p \longrightarrow 0$, the results of both approximations converge if the exact target states ψ_i and ψ_f can be used. However, even with such states, the ultrarelativistic limits of these two approaches are still different: the symmetric eikonal approximation yields the correct asymptotic behavior for cross sections as $\gamma \longrightarrow \infty$, but the first Born approximation does not. This point is very important and deserves a separate and detailed discussion. Here, we only note that at $\nu_p \ll 1$, the first Born approximation with exact target states would strongly fail only at $\gamma \sim c/\nu^2_p$ and higher.

The results of both the SEA and FBA using approximate target states ψ_i and ψ_f do not coincide even as $\nu_p \longrightarrow 0$. Therefore, a consistent way to "highlight" higher order effects in the projectile–target interaction is as follows. For a given $\nu_p \sim c$, calculations in the SEA are performed for the actual projectile (Z_p) and for the proton impact. The first-order result for the actual projectile is then obtained from that for the proton using the first-order scaling, i.e., via multiplication with Z_p^2. We call this first-order approach the SEA-1.

In calculations of the fully differential cross section given by Eq. (9), we approximate the initial state by the four-parameter Hylleraas wavefunction

$$\psi_i = N_i[\exp(-\alpha r_1 - \beta r_2) + (r_1 \longleftrightarrow r_2)] \times [1 - \delta \exp(-\lambda r_{12})], \quad (10)$$

where $N_i = 1.638$ is the normalization factor, $\alpha = 1.4096$, $\beta = 2.2058$, $\delta = 0.6054$, and $\lambda = 0.2420$. Wavefunction (10) yields $\varepsilon_i = -2.902$ a.u., which is close to the exact value –2.904 a.u. The final state is taken as

$$\psi_f = \psi_{3C} - \langle \psi_i | \psi_{3C} \rangle \psi_i,$$
$$\psi_{3C} = \frac{1}{\sqrt{2}}[\psi_{\mathbf{k}_1}(\mathbf{r}_1)\psi_{\mathbf{k}_2}(\mathbf{r}_2)\chi_{\mathbf{k}_{12}}(\mathbf{r}_{12}) + (\mathbf{r}_1 \longleftrightarrow \mathbf{r}_2)], \quad (11)$$

where $\mathbf{k}_{12} = (\mathbf{k}_1 - \mathbf{k}_2)/2$ and ψ_{3C} is the so-called 3C state, a (symmetrized) product of three Coulomb waves describing all pairwise interactions between the constituents of the target. The above approximations are chosen because they yield good results for helium double

[2] In collisions at very high γ, where very small momentum transfers contribute the most to double ionization, even for $\nu_p \sim 1$, a properly formulated first-order approach may be applied to the total cross section for the double ionization.

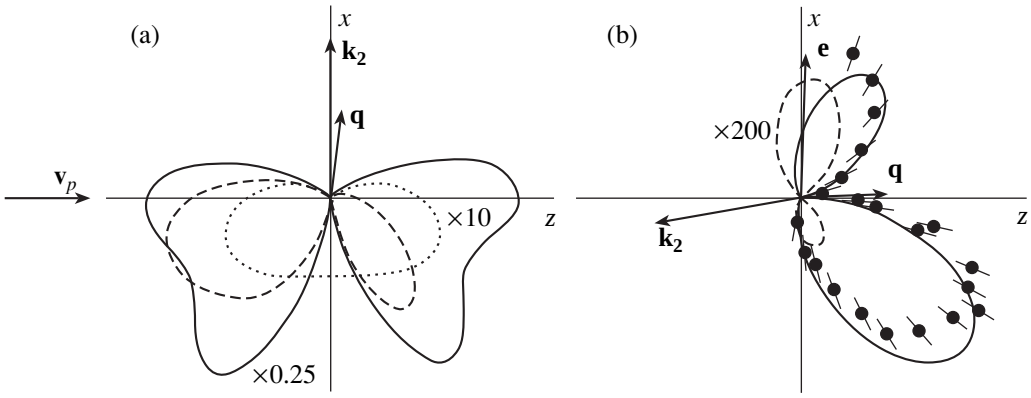

Fig. 1. The fully differential cross section (in arb. units) as a function of the polar emission angle ϑ_1 of the "first" electron, given in the plane defined by $\mathbf{v}_p = (0, 0, v_p)$ and $\mathbf{q} = (Q, 0, q_{\min})$. Emission energies $E_1 = E_2 = 10$ eV, azimuthal emission angles $\varphi_1 = \varphi_2 = 0°$. (a) $v_p = 120$ a.u., $Q = 0.25$ a.u., $\vartheta_2 = 90°$. The solid curve corresponds to the SEA, the dashed curve to the SEA without the *n*–*n* interaction, the doted curve to the SEA-1. (b) $v_p \approx 137$ a.u. ($\gamma = 26$), $Q = 10^{-3}$ a.u., $\vartheta_2 = 192°$. The solid curve correspond to the SEA, the dashed curve to the nonrelativistic SEA ($c = \infty$), the symbols are the experimental data on double photoionization [9] (the incident real photon is polarized along the *x* axis) normalized to the SEA results.

ionization due to the photoeffect and by fast electrons in collisions with relatively small momentum transfers. Such collisions become especially important at relativistic impact energies. In addition, with the states in Eqs. (10) and (11), the sixfold integrals over the electron coordinates in Eq. (8) can be reduced to two-fold integrals.

The results for the fully differential cross section in collisions with U^{92+} are shown in Fig. 1. Two important points should be mentioned.

First, within any first-order approach, the projectile may exchange only a single virtual photon with the target and can therefore directly interact with just one electron. Double ionization may then only occur due to electron–electron correlations and/or rearrangement in the target final state. However, the highly charged projectile, due to its strong field, can directly and very effectively interact with all the three target particles simultaneously. Therefore, such (higher-order) effects in the projectile–target interaction, which are properly described within the SEA, may profoundly influence the collision dynamics (Fig. 1a). Not only the direct interaction of the projectile with both electrons but also the *n*–*n* interaction (which itself does not lead to ionization) may very strongly affect the fully differential emission pattern.

Second, in collisions at very high γ and very low Q, the higher-order effects become of minor importance even at $\nu_p \sim 1$. A very interesting peculiarity of such collisions is that the physics of the impact ionization may become very similar to that of the photoeffect. A certain similarity between impact ionization and photoionization has been the subject of a long-term discussion in studies of the double ionization by fast nonrelativistic electrons. Such discussions, however, are of superficial character and can even be misleading because the fundamental similarity between these processes is only possible if $\gamma \gg 1$. Indeed, the emission pattern in Fig. 1b is almost indistinguishable from that due to the photoeffect because it is produced by the absorption of a virtual photon whose properties are very close to those of a real photon [6]. As a result, not the virtual photon momentum $\mathbf{q}$ but its polarization $\mathbf{e} \sim \mathbf{q}/\omega_{fi} - \mathbf{v}_p/c^2$ [6], almost perpendicular to $\mathbf{q}$, determines the shape of the emission cross section in Fig. 1b.

In [3], the emission spectrum differential in the energy of one of the ejected electrons has been reported. To produce such a spectrum from the cross section in (9), one has to perform seven additional integrations. This task is not feasible if ψ_i and ψ_f are given by Eqs. (10) and (11), but can be carried out if the terms depending on $\mathbf{r}_{12}$ are neglected in these equation, which allows evaluating the integrals over the electron coordinates in Eq. (8) analytically. Of course, neglecting the electron correlation would be a very improper approximation in the study of the fully differential cross section given by Eq. (9). Nevertheless, it is known that for collisions with highly charged ions, this approximation can still be used to estimate the total cross section and the energy emission spectrum integrated over the momentum transfer and all emission angles. The basic reasons for this are twofold. First, the double ionization in our case is dominated by the so-called TS-2 process, in which the electrons undergo transitions due to the "direct" interaction between the projectile and each of the two electrons. Second, while the electron–electron interaction in the continuum can strongly affect angular distributions, it cannot change the total energy of the electrons.

The results of such calculations ($\alpha = 1.885$, $\beta = 2.1832$, and $\delta = 0 \Rightarrow \varepsilon_i = -2.876$ a.u. and $\chi_{\mathbf{k}_{1,2}} = 1$) are shown in Fig. 2. For completeness, the energy spectrum

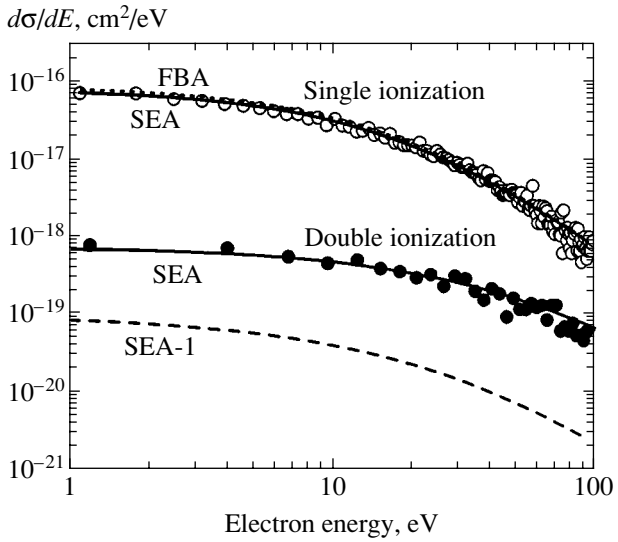

Fig. 2. Energy spectra of electrons emitted in 1-GeV/u U^{92+} + He($1s^2$) collisions. Symbols are the experimental data from [3]. See the text for more explanations.

of electrons emitted in singly ionizing collisions is also displayed.[3] For both single and double ionization, a very good agreement between the SEA results and experimental data is observed.[4] The overall effect of the higher-order terms in the projectile–target interaction is clearly seen in Fig. 2: it only slightly decreases the single ionization cross section but is very strong for the double ionization. Compared to the first-order result, the energy spectrum for double ionization decreases substantially slower as the emission energy increases and is larger on an absolute scale by a factor of 10–30 due to the large contribution from collisions in which both target electrons are removed simultaneously by their "independent" interactions with the projectile.

In conclusion, using a novel approach that treats, within the SEA, the interaction of the projectile with all the three target constituents on an equal footing, we have considered the double ionization of helium in relativistic collisions with highly charged ions. By exploring the basic dynamics of these collisions for the first time, we have demonstrated how the direct interaction of the projectile with all the three target particles can strongly affect the fully differential cross section. We have further shown that the fundamental similarity between the impact double ionization and double photoionization of helium naturally emerges in extreme relativistic collisions with very small transverse momentum transfers.

[3] Results for this spectrum were obtained by combining the SEA and FBA with the Hartree–Fock description of the "active" electron. We note that within the effective three-body collision model, where the (active) electron moves in the same Hartree–Fock potential in both initial and final states, the results of SEA and FBA nicely converge at $\nu_p \ll 1$.

[4] We note that for ionization by 1-GeV protons, the SEA, with ψ_i and ψ_f used to produce the spectrum in Fig. 2, yields the double-to-single ionization ratio $\sigma^{2+}/\sigma^{+} \approx 2 \times 10^{-3}$, which is quite close to the established high-velocity limit for this ratio approximately equal to 2.5×10^{-3} (see, e.g., [10]).

ACKNOWLEDGMENTS

The work was supported, in part, by the INTAS-GSI (grant 03-54-3604).

REFERENCES

1. *Many-Particle Quantum Dynamics in Atomic and Molecular Fragmentation*, Ed. by J. Ullrich and V. P. Shevelko (Springer, New York, 2003).
2. J. Ullrich, R. Moshammer, A. Dorn, *et al.*, Rep. Prog. Phys. **66**, 1463 (2003).
3. R. Moshammer, W. Schmitt, J. Ullrich, *et al.*, Phys. Rev. Lett. **79**, 3621 (1997).
4. *An International Accelerator Facility for Beams of Ions and Antiprotons. Conceptual Design Report* (GSI, Germany, 2003).
5. S. Keller, H. J. Lüdde, and R. M. Dreizler, Phys. Rev. A **55**, 4215 (1997).
6. A. B. Voitkiv and J. Ullrich, J. Phys. B **34**, 4513 (2001).
7. C. J. Wood, R. E. Olson, W. Schmitt, *et al.*, Phys. Rev. A **56**, 3746 (1997).
8. J. Eichler and W. Meyerhof, *Relativistic Atomic Collisions* (Academic, New York, 1995).
9. O. Schwarzkopf, B. Krässig, and V. Schmidt, J. Phys. Colloq. **C6**, 169 (1993).
10. J. Ullrich, R. Moshammer, H. Berg, *et al.*, Phys. Rev. Lett. **71**, 1697 (1993).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 521–527.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 589–596.

ATOMS, MOLECULES, OPTICS

Preparation of Biphotons in Arbitrary Polarization States

L. A. Krivitskiĭ, S. P. Kulik*, G. A. Maslennikov, and M. V. Chekhova

Moscow State University, Moscow, 119992 Russia
**e-mail: postmast@qopt.phys.msu.su*

Received October 6, 2004

Abstract—An experiment on preparation of entangled photon pairs (biphotons) in an arbitrary polarization state is described. The biphotons are qutrits (three-state quantum systems). They can be used in ternary quantum cryptography protocols. A theoretically derived orthogonality criterion for the prepared biphotons is validated experimentally. The criterion can be used to identify orthogonal biphoton states. © *2005 Pleiades Publishing, Inc.*

1. BIPHOTONS AS QUTRITS

Most quantum cryptography protocols are based on binary encoding (with qubits) [1, 2]. Qubits can be prepared as polarization states of a single-photon wave packet, states of a spin 1/2 particle, states of a single-photon wave packet in a two-arm interferometer, and in various other ways.

In recent studies, it was proposed to use ternary logic (qutrits) in quantum cryptography instead of qubits [3–5]. These studies are generally motivated by the higher efficiency [3] and higher security [6] of quantum channels of higher dimension. A qutrit is a three-state quantum system, as a three-level atom or a spin 1 particle. However, photons are known as the best means of data transmission. There exist several methods for making photonic qutrits. In particular, ternary encoding can use photon states obtained in a three-arm interferometer [7], single-photon wave packets with helical wavefronts [8], and four-photon states created by parametric down-conversion [9].

The ternary quantum cryptography scheme proposed in [4] made use of polarization-entangled states of photon pairs (biphotons) obtained as a result of spontaneous parametric down-conversion (SPDC). It is the simplest method for preparing an arbitrary polarization state of a qutrit [10, 11], i.e., a state of the form

$$|\Psi\rangle = c_1|2, 0\rangle + c_2|1, 1\rangle + c_3|0, 2\rangle \qquad (1)$$

with arbitrary amplitudes c_1, c_2, and c_3. The ket notation $|m, n\rangle$ in (1) means m vertically polarized photons and n horizontally polarized ones. Theoretically, the resulting biphotons belong to the same spatiotemporal mode. Even though the biphotonic field created in an experiment always spans a frequency–angle spectrum of finite width, representation (1) is valid if the optical detector employed in the scheme does not resolve the frequency–angular spectrum. By virtue of the normalization

$$|c_1|^2 + |c_2|^2 + |c_3|^2 = 1$$

and the unimportance of the overall phase of state (1), a biphoton is defined by four real numbers, e.g., the two amplitudes

$$d_1 \equiv |c_1|, \quad d_2 \equiv |c_2|$$

and the two phases

$$\varphi_{13} \equiv \arg(c_1 c_3^*), \quad \varphi_{23} \equiv \arg(c_2 c_3^*).$$

State (1) is also conveniently represented as a pair of photons in arbitrary pure polarization states [12]:

$$|\Psi\rangle = \frac{a^\dagger(\vartheta, \varphi)a^\dagger(\vartheta', \varphi')|vac\rangle}{|a^\dagger(\vartheta, \varphi)a^\dagger(\vartheta', \varphi')|vac\rangle|}, \qquad (2)$$

where $a^\dagger(\vartheta, \varphi)$ and $a^\dagger(\vartheta', \varphi')$ are the operators of creation of photons in arbitrary polarization modes characterized by azimuthal (ϑ, ϑ') and polar (φ, φ') angles on the Poincaré sphere.

Representation (2) can be used to depict a photon pair on the Poincaré sphere. Furthermore, it can be shown [12] that important polarization characteristics of state (1), such as the Stokes vector and degree of polarization, are readily calculated by using representation (2). For example, the degree of polarization of state (1) is uniquely determined by the angular distance σ between the points representing the states φ, ϑ and φ', ϑ' on the Poincaré sphere:

$$P = \frac{2\cos(\sigma/2)}{1 + \cos^2(\sigma/2)}. \qquad (3)$$

It should be emphasized here that this quantity is interpreted as the conventional (classical) polarization

degree defined in terms of the Stokes parameters S_0, S_1, S_2, and S_3 [13]:

$$P \equiv \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0}. \quad (4)$$

Degree of light polarization (4) can be measured in experiment as the maximum visibility of the modulation observed in polarization-dependent intensity [14]. Since this quantity is completely determined by second-order moments of the field, it is not an optimal characteristic of biphotonic light (whose most interesting properties manifest themselves in the behavior of fourth-order moments). An alternative definition of degree of polarization, the photon–photon polarization degree formulated in terms of fourth-order moments, has been proposed to describe the polarization state of a biphoton [14–16]. However, since the "photon–photon" polarization degree of a pure state having the form of (1) is always unity [14], it provides no information about the relative location of the two points representing a biphoton on the Poincaré sphere. Only the quantity defined by (4) provides information of this kind.

We should also mention here the biphotons created as mixed states instead of pure state (1). In this case, the components of state (1) are multiplied by phase factors exhibiting classical fluctuations; i.e., the biphoton is a statistical mixture of several basis states. A biphoton state of this kind can be prepared by SPDC implemented by means of incoherent pumping in two or three crystals. However, it cannot be represented as (2), because (2) is a pure state.

2. EXPERIMENT ON PREPARATION OF A BIPHOTON WITH ARBITRARY DEGREE OF POLARIZATION

An arbitrary state having the form of (1) can be prepared by using an interferometric scheme in which SPDC is implemented by means of common pumping [17]. However, this scheme is impracticable because of its instability. In this study, arbitrary biphotons (qutrits) were prepared by using a different scheme, which does not require the use of an interferometer.

First, we prepared and detected biphotons having an arbitrary polarization degree (4), i.e., biphotons based on photons in arbitrary polarization states (varying from similar to orthogonal ones).

The experimental setup is schematized in Fig. 1. Collinear, frequency-degenerate Type I SPDC was implemented in two beta-barium borate (BBO) crystals pumped by a 325 nm He-Cd laser. The optical axes of the crystals were oriented so that vertically and horizontally polarized photon pairs (states $|0, 2\rangle$ and $|2, 0\rangle$) were generated in the first and second crystals, respectively. The ratio of the corresponding absolute amplitudes were varied by rotating a half-wave plate placed in the pump beam. In addition, the phase difference ε between the states $|2, 0\rangle$ and $|0, 2\rangle$ was varied by tilting two quartz plates with vertical optical axes. The light generated by the two crystals was the coherent superposition

$$|\Psi_1\rangle = \sin(2\chi)|2, 0\rangle + e^{i\varepsilon}\cos(2\chi)|0, 2\rangle, \quad (5)$$

where χ is the half-wave plate rotation angle relative to the vertical axis. Thus, two of the four real parameters that determine state (1) (χ and ε) could be varied so that

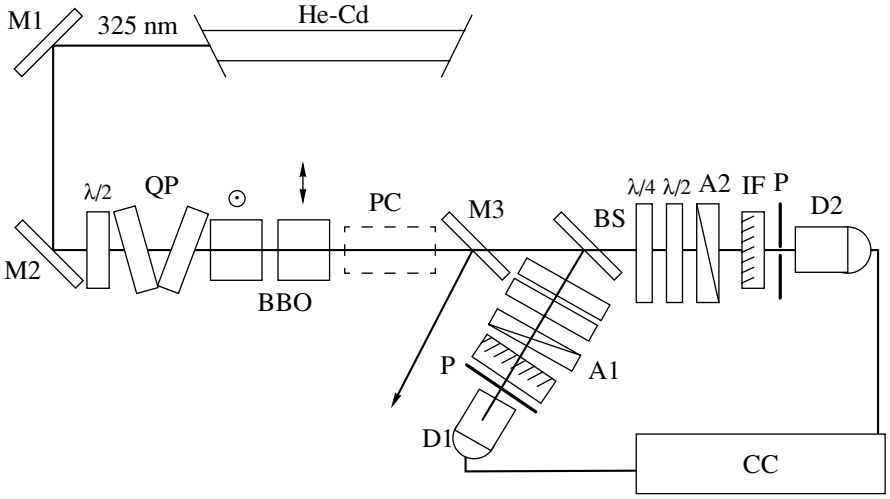

Fig. 1. Experimental setup: He-Cd = helium-cadmium laser; M1, M2, M3 = mirrors reflecting the pump beam; QP = quartz plates; BBO = beta-barium borate crystals; P = pinhole; IF = interference filter; BS = nonpolarizing beamsplitter; A1, A2 = polarizers; D1, D2 = detectors (avalanche photodiodes); CC = coincidence circuit; PC = Pockels cell.

the polarization degree of the generated state would take any value between 0 and 1.

Initially, the phase difference ε was set equal to π. By varying χ from 0 to 45°, the generated state was transformed from $|2, 0\rangle$ into $|0, 2\rangle$, so that the two points corresponding to a biphoton state on the Poincaré sphere traversed its equator (see Fig. 2), remaining symmetric relative to the axis *HV*. This state can be represented as (2) with

$$\vartheta = \vartheta' = 2\arctan[\sqrt{\cot(2\chi)}], \quad \varphi = 0, \quad \varphi' = \pi. \quad (6)$$

In particular, we used $\chi = 22.5°$ to obtain the state $|+45°, -45°\rangle$, i.e., a pair of photons polarized linearly at angles ±45° relative to the vertical axis. The generated light was totally unpolarized (with $P = 0$). Overall, the degree of polarization varied from 1 to 0; the angular distance between the points representing the biphoton on the Poincaré sphere varied from zero to π.

Note that the degree of polarization P of state (5) depends only on the rotation angle χ of the half-wave plate placed in the pump beam: $P = |\cos(4\chi)|$. Thus, when the plate was fixed in a certain position, the two points on the Poincaré sphere corresponding to biphoton (5) were located symmetrically with respect to the axis *HV* and separated by a constant angular distance. As the phase difference was varied, the points moved simultaneously about the axis *HV*.

The pump beam that passed through the two crystals was eliminated by a mirror, and the biphoton light was selected both spatially and spectrally (by using a pinhole and a 10 nm bandwidth interference filter with transmittance peak at 650 nm, respectively) and directed into a Hanbury-Brown–Twiss interferometer in order to detect fourth-order moments of the field. The interferometer consisted of a 50% nonpolarizing beamsplitter (a plane-parallel plate set at a small angle of 15°, relative to the beam so that both reflected and transmitted light polarizations were similar to that of the incident light), two photodetectors (EG&G avalanche photodiodes), and a coincidence circuit with a resolution of 1.5 ns. Thin-film polarizers were inserted into the interferometer arms and used as linear polarization filters.

3. DEMONSTRATION OF ORTHOGONALITY OF BIPHOTONS (QUTRITS)

The prepared biphoton states were used to validate the operational orthogonality criterion formulated in [17]. It was shown in [17] that the orthogonality of biphotons $|\Psi_1\rangle$ and $|\Psi_2\rangle$ is equivalent to zero counting rate in the Hanbury-Brown–Twiss interferometer output when its input is the biphoton state $|\Psi_1\rangle$ and the polarization filters inserted into its arms select the photon polarization states that constitute $|\Psi_2\rangle$. The experimental setup schematized in Fig. 1 always prepared the state $|\Psi_2\rangle$ as a pair of linearly polarized photons, while

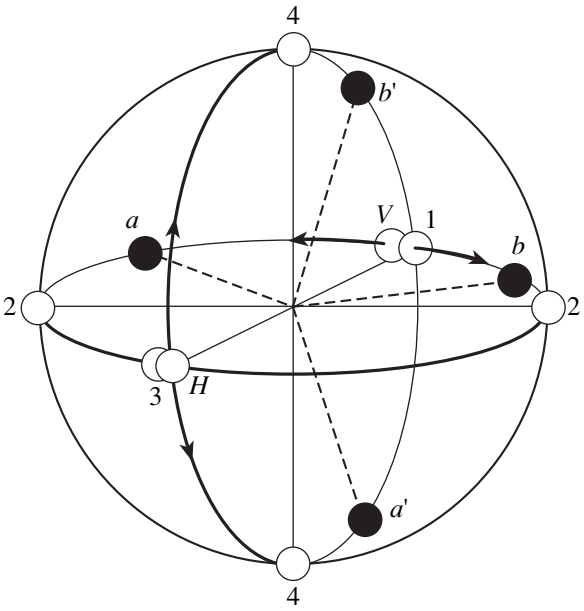

Fig. 2. Prepared states on the Poincaré sphere: (1) state $|2, 0\rangle$ generated at $\chi = 0$; (2) state $|+45°, -45°\rangle$, $\chi = 22.5°$; (3) state $|0, 2\rangle$, $\chi = 45°$; (4) state $|R, L\rangle$ (pair of left- and right-polarized photons), $\chi = 67.5°$. The states $|a, b\rangle$ and $|a', b'\rangle$ prepared with $\chi = 15°$ and 75°, respectively, were studied in the present experiment.

different states $|\Psi_1\rangle$ were used as input (both pairs of linearly polarized photons and pairs of elliptically polarized photons) and the biphoton polarization degree varied from 1 to 0 (see above).

The state $|\Psi_1\rangle$ with a polarization degree of 0.5 was selected as input. In this case, the biphoton $|\Psi_1\rangle \equiv |a, b\rangle$ is represented by a pair of points located on the equator of the Poincaré sphere at angles of ±74.5° relative to the *HV* axis. According to (5), this state corresponds to $\chi = 15°$. Whereas there exist an infinite number of biphoton states $|\Psi_2\rangle$ orthogonal to the input state, the orthogonality criterion uniquely determines the polarization state of one of the photons that make up a biphoton if the state of the other is preset. In our experiment, it was convenient to set one polarizer in the Hanbury-Brown–Twiss interferometer at an angle of 45° relative to the vertical. A calculation showed that the other polarizer must then be set at 60° to the vertical.

Figure 3a shows the coincidence rate measured versus χ for $\varepsilon = \pi$ when the polarizers were held at 45 and 60°. According to our calculations, the minimum coincidence rate corresponds to $\chi = 15°$. When the half-wave plate is fixed at $\chi = 15°$ and polarizer $A1$ is rotated while polarizer $A2$ is held at 45°, then the minimum coincidence rate corresponds to an angle of 60° (Fig. 3b).

As the half-wave plate was rotated, the biphoton state transformed into $|0, 2\rangle$ (at $\chi = 45°$). With further increase in the rotation angle, the two points representing the biphoton components moved in opposite direc-

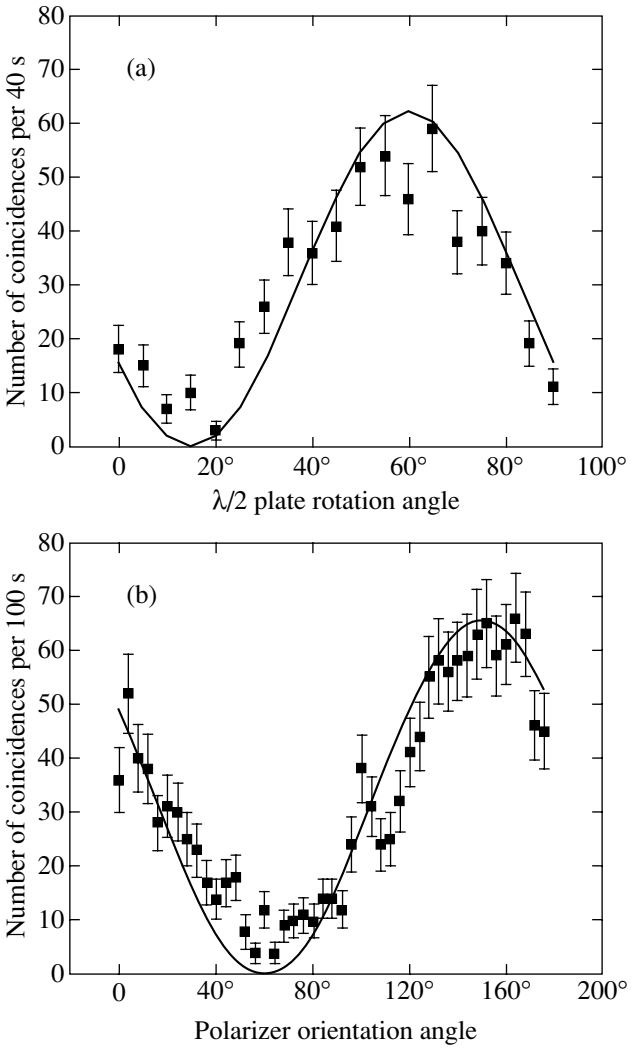

Fig. 3. Orthogonality of biphotons in the case when both input photons are linearly polarized (state $|a, b\rangle$ in Fig. 2), $\varepsilon = \pi$, and polarizers held at 45 and 60°: (a) coincidence rate versus half-wave plate rotation angle χ, the orthogonality condition is satisfied at $\chi = 15°$; (b) coincidence rate versus the polarizer rotation angle relative to the vertical, the orthogonality condition is satisfied at an angle of 45°, the other polarizer is held at 60° relative to the vertical.

tions along a meridian of the Poincaré sphere (see Fig. 2). At $\chi = 67.5°$, the output was the state $|R, L\rangle$ consisting of circularly polarized photons. At $\chi = 75°$, the degree of polarization of the output was again 0.5 (as in $|a, b\rangle$), but the corresponding biphoton ($|a', b'\rangle$ in Fig. 2) consisted of elliptically polarized photons. Then, it can be verified by calculation that the orthogonality condition is satisfied when the polarizers are oriented at 45 and –60°. Accordingly, the coincidence rate measured for these polarizer positions versus the rotation angle of the half-wave plate placed in the pump beam reaches a minimum at approximately $\chi = 75°$ (Fig. 4).

Figure 5 shows the coincidence rate measured as a function of ε for $\chi = 15°$ (i.e., input state $|a, b\rangle$ in Fig. 2) while the polarizers were held at 45 and 60°. The minimum coincidence rate was observed in the vicinity of $\varepsilon = \pi$.

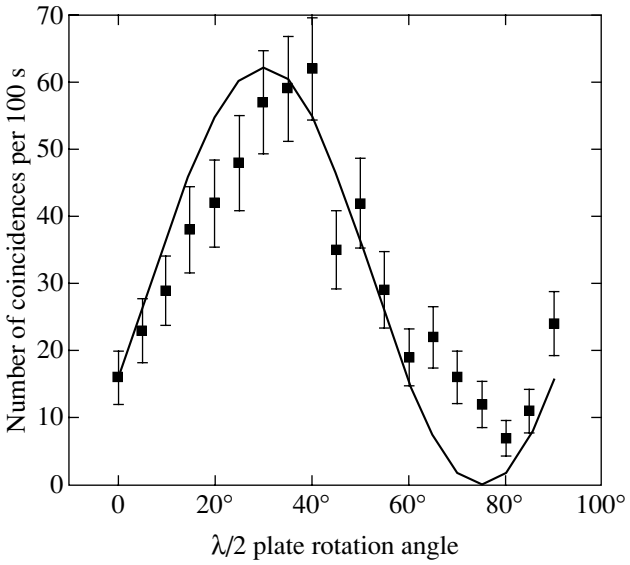

Fig. 4. Orthogonality of biphotons in the case when both input photons are elliptically polarized (state $|a', b'\rangle$ in Fig. 2), $\varepsilon = \pi$, and polarizers held at 45 and –60°: coincidence rate versus half-wave plate rotation angle χ. The orthogonality condition is satisfied at $\chi = 75°$.

The curves plotted in Figs. 3–5 were calculated for particular biphoton states and polarizer positions (e.g., see [18]). The only fitted parameter was the vertical scale.

4. TRANSITION TO AN ARBITRARY BIPHOTON STATE

To change from a biphoton with an arbitrary degree of polarization (represented by two points separated by an arbitrary angular distance on the Poincaré sphere) to a biphoton in an arbitrary polarization state (represented by two arbitrary points on the Poincaré sphere), one must be able to perform any required transformation of state (5). This can be done by varying ε, χ, and two additional parameters characterizing state (5). These parameters can be the retardation δ and orientation α of a retarding plate placed in the prepared biphoton beam.

The feasibility of transition from state (5) to state (1) with a retarding plate of arbitrary thickness and orientation added to the setup shown in Fig. 1 can be illustrated by performing a simple geometric construction on the Poincaré sphere.

An arbitrary biphoton state $|a, b\rangle$ (see Fig. 6) can be obtained by using a retarding plate with certain δ and α to transform the state $|a', b'\rangle$ represented by two points on the Poincaré sphere located symmetrically relative to the axis *HV*. The required plate parameters are determined by the condition that the transformation on the Poincaré sphere maps the symmetry axis of the pair

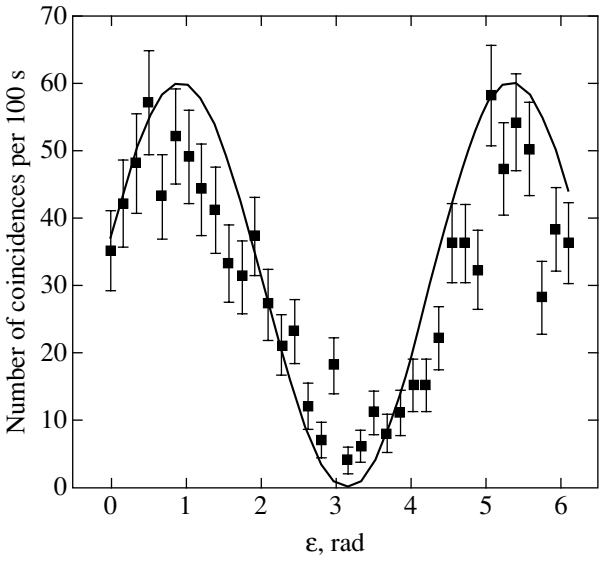

Fig. 5. Coincidence rate versus phase ε for half-wave plate set χ = 15° and polarizers set at 45 and 30°. The orthogonality condition is satisfied at ε = π.

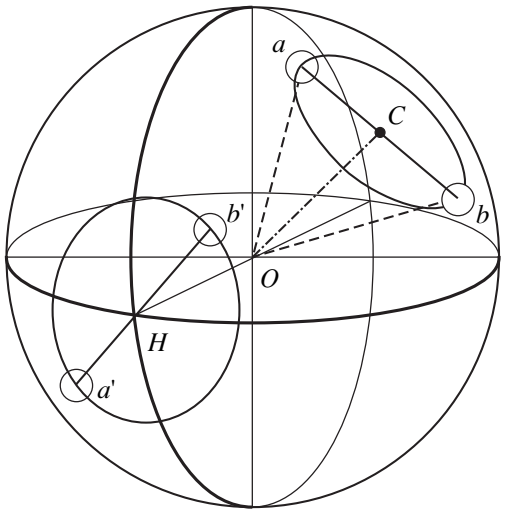

Fig. 6. Preparation of an arbitrary polarized biphoton $|a, b\rangle$ by using the setup schematized in Fig. 1 and a Pockels cell acting as a retarding plate with variable retardation and orientation. The Pockels cell transforms the state $|a', b'\rangle$ created by using the setup schematized in Fig. 1 into the desired state $|a, b\rangle$.

$|a, b\rangle$ (axis *OC* in Fig. 6) to the *HV* axis. Note that the state $|a', b'\rangle$ can be prepared by using the setup shown in Fig. 1 so that the angular distance between points *a* and *b* on the sphere and the angle of their rotation relative to its equator are determined by the orientation χ of half-wave plate placed in the pump beam and the phase difference ε, respectively.

In other words, the retarding plate placed after the two crystals will map state (5) to state (1) by rotating the points representing the biphoton in Fig. 1 as a whole. The resulting state (1) is characterized by the four parameters ε, χ, α, and δ. One practical difficulty in this method is that the parameter δ can be varied only by changing plates with different retardations.

To vary δ by an arbitrary amount, one should use a Pockels cell instead of a set of wave plates. Varying the voltage applied to the cell, one can use it as a retarding plate with variable retardation. The parameter α can be varied gradually by revolving the cell about the biphoton beam.

A state described by (1) was prepared by placing a Pockels cell after the two crystals in the optical arrangement schematized in Fig. 1. The element used in the Pockels cell was a 3 cm long crystal of lithium niobate cut along the optical *z* axis. When a dc voltage was applied along the *x* axis, the crystal became weakly birefringent, and the plane of its optical axes made an angle of 45° with the *xz* plane. As the dc voltage applied to the crystal was varied from zero to 2.8 kV, the phase difference between the ordinary wave (polarized in the plane of the optical axes, i.e., at an angle of 45° relative to the *xz* plane) and the extraordinary wave (polarized at an angle of −45° relative to the *xz* plane) increased from zero to 2π. Note that the cell acted as a zeroth-order retarding plate with δ varying between 0 and π. Thus, the electrically induced transformation of polarization in the cell was similar at any frequency within the SPDC bandwidth (about 40 nm) and, therefore, within the bandwidth of the interference filter. Note also that the orientation α of the optical axis of a Pockels cell treated as a retarding plate is determined by the plane of the induced optical axes, i.e., makes an angle of 45° with the direction of the electric field applied to the cell.

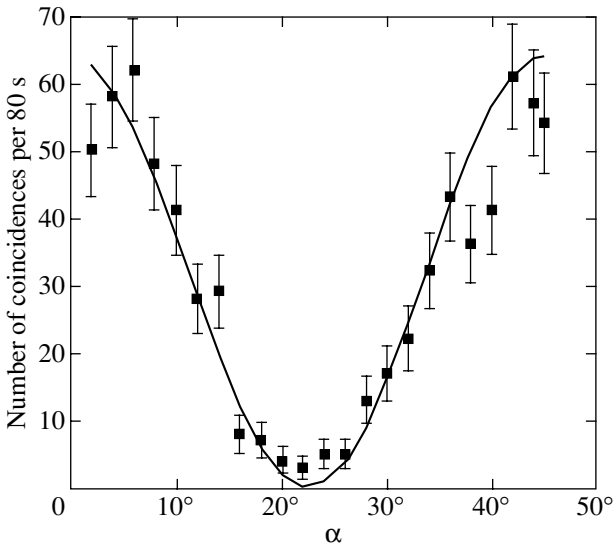

Fig. 7. Coincidence rate versus Pockels cell orientation angle in the case of half-wave voltage applied to the Pockels cell (retardation δ = π/2), half-wave plate orientation χ = 22.5°, phase ε = π, and polarizers set at 45 and −45°.

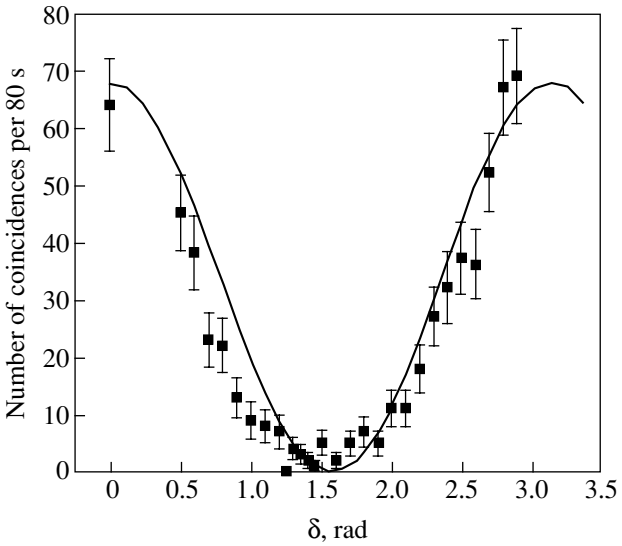

Fig. 8. Coincidence rate versus retardation δ for Pockels cell set at α = 22.5°, half-wave plate orientation χ = 22.5°, phase ε = π, and polarizers set at 45 and –45°.

To begin our measurements, we switched off the voltage applied to the Pockels cell, placed the half-wave plate at the angle χ = 22.5° in the pump beam, and set the phase difference ε equal to π (see (5)). This combination of parameters corresponded to the highest coincidence rate when the positions of the polarization filters were set at 45 and –45° so that the input state was $|+45°, -45°\rangle$. Next, we applied the half-wave voltage to the Pockels cell. Since the optical axis of the cell was initially aligned with the vertical direction, the biphoton remained in the same state $|+45°, -45°\rangle$ when the half-wave voltage was applied: the cell executed a rotation by π relative to the axis *HV* on the Poincaré sphere. The state of the output biphoton varied as the Pockels cell was revolved, and the coincidence rate varied accordingly. Figure 7 shows the coincidence rate versus the angle α between the optical axis of the Pockels cell and the vertical axis. It is clear that coincidences virtually vanished at α = 22.5° because the Pockels cell transformed the selected state $|+45°, -45°\rangle$ into the state $|H, V\rangle$, which is orthogonal to $|+45°, -45°\rangle$.

Finally, we set α = 22.5° and varied the voltage applied to the cell from zero to 2.8 kV. Figure 8 shows the corresponding coincidence rate as a function of δ. The minimum coincidence rate was reached at δ = π/2, as in the preceding case, because the states $|H, V\rangle$ and $|+45°, -45°\rangle$ are mutually orthogonal.

5. CONCLUSIONS

We have experimentally demonstrated the preparation of arbitrarily polarized two-photon states (qutrits). The states were generated by using a scheme that did not include any interferometer. This highly stable preparation scheme can be employed in practical quantum cryptography. The arbitrarily polarized biphotons produced were utilized to validate a method for identifying orthogonal biphoton states based on zero counting rate in the Hanbury-Brown–Twiss scheme. The feasibility of experimental identification of orthogonal biphotons (qutrits) means that they can be used in quantum data transmission protocols, including quantum cryptography protocols. The 12 biphoton states required to implement the ternary analog of the BB84 protocol were calculated and represented on the Poincaré sphere in [19].

We should also note that the present demonstration of orthogonality of biphoton states generalizes the well-known experiment on anticorrelation dip reported in [20]. Previously, this effect was observed only for similarly polarized [20] or orthogonally polarized [21] photon pairs. The present study is the first demonstration of this effect for arbitrarily polarized photon pairs.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 02-02-16664 and 03-02-16444; by INTAS, project no. 2122-01; and under the State Program for Support of Leading Science Schools, grant 166.2003.02. One of us (L.A.K.) gratefully acknowledges the support provided by INTAS, YS fellowship grant no. 03-55-1971.

REFERENCES

1. *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*, Ed. by D. Bouwmeester, A. K. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
2. C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
3. K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger, Phys. Rev. Lett. **76**, 4656 (1996).
4. A. V. Burlakov, M. V. Chekhova, O. A. Karabutova, *et al.*, Phys. Rev. A **60**, R4209 (1999).
5. H. Bechmann-Pascuinucci and A. Peres, Phys. Rev. Lett. **85**, 3313 (2000).
6. D. Bruss and C. Machiavello, Phys. Rev. Lett. **88**, 127901 (2002).
7. R. Thew, A. Acin, H. Zbinden, and N. Gisin, quant-ph/0307122.
8. A. Vaziri, G. Weihs, and A. Zeilinger, Phys. Rev. Lett. **89**, 240401 (2002).
9. J. C. Howell, A. Lamas-Linares, and D. Bouwmeester, Phys. Rev. Lett. **88**, 030401 (2002).
10. Yu. Bogdanov, M. Chekhova, S. Kulik, *et al.*, Phys. Rev. A **70**, 042303 (2004).
11. A. V. Burlakov and D. N. Klyshko, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 795 (1999) [JETP Lett. **69**, 839 (1999)].

12. A. V. Burlakov and M. V. Chekhova, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 505 (2002) [JETP Lett. **75**, 432 (2002)].

13. W. A. Shurcliff, *Polarized Light: Production and Use* (Harvard Univ. Press, Cambridge, Mass., 1962; Mir, Moscow, 1965).

14. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **111**, 1955 (1997) [JETP **84**, 1065 (1997)].

15. A. S. Chirkin, A. A. Orlov, and D. Yu. Parashchuk, Kvantovaya Élektron. (Moscow) **20**, 999 (1993).

16. G. Bjork, J. Soderholm, A. Trifonov, *et al.*, in *Proceedings of SPIE on Quantum and Atomic Optics, High-Precision Measurements in Optics, and Optical Information Processing, Transmission, and Storage, ICONO 2001*, Ed. by S. N. Bagayev, S. S. Chesnokov, A. S. Chirkin, and V. N. Zadkov (2002), Proc. SPIE, Vol. 4750.

17. A. A. Zhukov, G. A. Maslennikov, and M. V. Chekhova, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 696 (2002) [JETP Lett. **76**, 596 (2002)].

18. L. A. Krivitskiĭ, S. P. Kulik, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **124**, 943 (2003) [JETP **97**, 846 (2003)].

19. G. A. Maslennikov, M. V. Chekhova, S. P. Kulik, and A. A. Zhukov, J. Opt. B: Quantum Semiclassic. Opt. **5**, 530 (2003).

20. C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).

21. Y. H. Shih and A. V. Sergienko, Phys. Lett. A **186**, 29 (1994).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 528–538.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 597–609.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

The Accretion of Dark Energy onto a Black Hole

E. O. Babichev, V. I. Dokuchaev, and Yu. N. Eroshenko

Institute for Nuclear Research, Russian Academy of Sciences,
pr. Shestidesyatiletiya Oktyabrya 7a, Moscow, 117312 Russia
e-mail: babichev@inr.npd.ac.ru; dokuchaev@inr.npd.ac.ru; erosh@inr.npd.ac.ru

Received June 15, 2004

Abstract—The stationary, spherically symmetric accretion of dark energy onto a Schwarzschild black hole is considered in terms of relativistic hydrodynamics. The approximation of an ideal fluid is used to model the dark energy. General expressions are derived for the accretion rate of an ideal fluid with an arbitrary equation of state $p = p(\rho)$ onto a black hole. The black hole mass was found to decrease for the accretion of phantom energy. The accretion process is studied in detail for two dark energy models that admit an analytical solution: a model with a linear equation of state, $p = \alpha(\rho - \rho_0)$, and a Chaplygin gas. For one of the special cases of a linear equation of state, an analytical expression is derived for the accretion rate of dark energy onto a moving and rotating black hole. The masses of all black holes are shown to approach zero in cosmological models with phantom energy in which the Big Rip scenario is realized. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In recent years, strong observational evidence that the Universe is currently expanding with acceleration has been obtained. In the Einstein theory of gravitation, this positive acceleration is explained by the dominance of dark energy with a negative pressure in the Universe [1–4]. Several theoretical models of dark energy have been suggested: the vacuum energy (the cosmological constant Λ) or such dynamical components as quintessence [5–10] and *k* essence [11–13]. Models with dynamical dark energy seem more realistic, since tracker [14, 15], or attractor, solutions are realized in them. Thus, the problem of fine tuning the parameters of the Universe is solved [11–13].

A peculiar property of cosmological models with dark energy is the possibility of a Big Rip [16, 17]: an infinite increase in the scale factor of the Universe in a finite time. The Big Rip scenario is realized in the case of dark energy, the so-called phantom energy (for which $\rho + p < 0$). In the Big Rip scenario, the cosmological phantom energy density tends to infinity, and all of the bound objects are torn apart up to subnuclear scales. It should be noted, however, that the condition $\rho + p < 0$ alone is not enough for the Big Rip scenario to be realized [18]. Alam *et al.* [19] analyzed data on distant supernovas in a model-independent way and showed that the presence of phantom energy with $-1/2 < p/\rho < -1$ in the Universe at present is highly likely. The quantum properties of the phantom energy in curved spacetime were considered in [20]. The entropy of the Universe filled with phantom energy was discussed in [21]. Models with phantom energy are also used to construct mole burrows [22, 23]. The accretion of a scalar field onto a black hole from special potentials $V(\phi)$ was considered in [24–29]. We use a different approach to describe the accretion of dark energy onto a black hole; more specifically, we model the dark energy by an ideal fluid with a negative pressure.

In our recent paper [30] (see also [31]), we showed that the masses of all black holes in the Universe with phantom energy gradually decrease, and the black holes disappear completely by the Big Rip. In this paper, we consider in detail the stationary spherical accretion of dynamical dark energy onto a black hole. The dark energy is modeled by an ideal fluid with a negative pressure. The history of research on the accretion of an ideal fluid onto a compact object begins with Bondi's classic paper [32]. A relativistic generalization was made by Michel [33] (see also [34–41] for further generalizations and supplements to Michel's solution). Carr and Hawking [42] considered the accretion of dust and radiation onto a black hole by solving the complete system of Einstein equations and taking into account the back reaction of the surrounding matter (see also [43] for a description of the progress made in this area and for a discussion of fundamental questions). Below, we obtain a solution for the stationary accretion of a test relativistic ideal fluid with an arbitrary equation of state $p(\rho)$ onto a Schwarzschild black hole. Using this solution, we show that the black hole mass decreases during the accretion of phantom energy. The masses of black holes can decrease during accretion in the case of phantom energy due to the violation of the energy dominance condition ($\rho + p \geq 0$) that underlies the theorem on the nondecreasing area of the event horizon of a classical black hole [44].

This paper is structured as follows. In Section 2, we derive general equations for the spherical accretion of an ideal fluid and describe basic parameters of the steady energy flux onto a black hole. We consider an

arbitrary equation of state, $w = p/\rho$, where the pressure p can be positive (for ordinary matter) and negative (for dark energy, including phantom energy $w < -1$). Note that the parameter w of the equation of state need not be constant in our approach. Accretion causes the black hole mass to change: the mass increases for $\rho + p > 0$ and decreases for $\rho + p < 0$. The energy flux turns out to be completely determined by the black hole mass M, the dark energy density at infinity ρ_∞, and the equation of state $p = p(\rho)$ only if $0 < \partial p/\partial\rho < 1$. In this case, there is a critical point that fixes the flux just as for an ordinary fluid. When the condition $0 < \partial p/\partial\rho < 1$ is violated, the dark energy flux onto a black hole can formally be arbitrary. For $0 < \partial p/\partial\rho < 1$, we describe the method of calculating the fluid parameters at the critical point and the energy flux onto a black hole for given M, ρ_∞, and $p = p(\rho)$. In Section 3, we consider specific models of the equation of state for dark energy. In the first model, we use a simple equation of state with a linear density dependence of the pressure. We consider the special cases of accretion of several types of ideal fluid: thermal radiation, matter with an ultrahard equation of state, dark energy with $\partial p/\partial\rho \lesseqgtr 0$, and linear phantom energy. The accretion rate of dark energy onto a moving black hole was calculated for the special case of $\partial p/\partial\rho = 1$. As the second model, we investigate the accretion of a Chaplygin gas onto a black hole. The evolution of the black hole mass in the Universe with the Big Rip is considered in Section 4. The possibility that the presence of phantom energy will lead the Universe to the Big Rip in the future has been discussed in recent years. The problem of the fate of black holes in this Universe is solved in a rather unexpected way: black holes are not torn apart, but disappear by the Big Rip due to the accretion of phantom energy, irrespective of their initial masses. In Section 5, we discuss the correspondence between the accretion of dark energy modeled by an ideal fluid onto a black hole and the accretion of a scalar field. The results obtained are briefly discussed in Section 6.

2. GENERAL EQUATIONS FOR SPHERICAL ACCRETION

Let us consider the stationary, spherically symmetric accretion of an ideal fluid that models the dark energy in the special case of a negative pressure onto a black hole. The dark energy density is assumed to be low enough for the metric to be a Schwarzschild one with a high accuracy:

$$ds^2 = \left(1 - \frac{2M}{r}\right)dt^2 - \left(1 - \frac{2M}{r}\right)^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2), \quad (1)$$

where M is the mass of the black hole, r is the radial coordinate, and θ and ϕ are the angular spherical coordinates. We model the dark energy by an ideal fluid with the energy–momentum tensor

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu}, \quad (2)$$

where ρ is the density, p is the dark energy pressure, and $u^\mu = dx^\mu/ds$ is the radial 4-velocity component. The pressure is assumed to be an arbitrary function of the density, $p = p(\rho)$. Integrating the zeroth (time) component of the conservation law $T^{\mu\nu}_{;\nu} = 0$ yields the first integral of motion for stationary, spherically symmetric accretion (Bernoulli's relativistic equation or the energy equation):

$$(\rho + p)\left(1 - \frac{2}{x} + u^2\right)^{1/2} x^2 u = C_1, \quad (3)$$

where $x = r/M$, $u = dr/ds$, and C_1 is the constant determined below. To find the second integral of motion, we use the equation for the component of the energy–momentum tensor conservation law along the 4-velocity u^μ:

$$u_\mu T^{\mu\nu}_{;\nu} = 0. \quad (4)$$

In our case, this equation is [45]

$$u^\mu \rho_{,\mu} + (\rho + p)u^\mu_{;\mu} = 0. \quad (5)$$

For the given equation of state

$$p = p(\rho), \quad (6)$$

the auxiliary function $n = n(\rho)$ can be defined by the relation

$$\frac{d\rho}{\rho + p} = \frac{dn}{n}. \quad (7)$$

The function n is identical to the particle concentration for an atomic gas, but it can also be used to describe a continuous medium that does not consist of any particles. In this case, the "concentration" n is a formal auxiliary function. For an arbitrary equation of state $p = p(\rho)$, we obtain a solution for n from Eq. (7):

$$\frac{n(\rho)}{n_\infty} = \exp\left(\int_{\rho_\infty}^{\rho} \frac{d\rho'}{\rho' + p(\rho')}\right). \quad (8)$$

Using (8), we find the sought second integral of motion from Eq. (5):

$$\frac{n(\rho)}{n_\infty} u x^2 = -A, \quad (9)$$

where n_∞ (the dark energy "concentration" at infinity) was introduced for convenience. In the case of a fluid flow directed toward the black hole, $u = dr/ds < 0$, and,

therefore, the numerical constant $A > 0$. From (3) and (9), we can easily obtain

$$\frac{\rho + p}{n}\left(1 - \frac{2}{x} + u^2\right)^{1/2} = C_2, \quad (10)$$

where

$$C_2 = \frac{\rho_\infty + p(\rho_\infty)}{n(\rho_\infty)}. \quad (11)$$

Let us now calculate the radial 4-velocity component and the fluid density on the event horizon of the black hole, $r = 2M$. Setting $x = 2$, we obtain from Eqs. (9)–(11)

$$\frac{A}{4}\frac{\rho_H + p(\rho_H)}{\rho_\infty + p(\rho_\infty)} = \frac{n^2(\rho_H)}{n^2(\rho_\infty)}, \quad (12)$$

where ρ_H is the density on the $x = 2$ horizon. Thus, having specified the density at infinity ρ_∞, the equation of state $p = p(\rho)$, and the flux A and using definition (7) of the concentration, we can calculate the fluid density ρ_H on the event horizon of the black hole from (12). Given the density on the horizon ρ_H, we can easily determine the radial fluid 4-velocity component on the horizon from (9):

$$u_H = -\frac{A}{4}\frac{n(\rho_\infty)}{n(\rho_H)}. \quad (13)$$

Below, we will see that the constant A, which defines the energy flux onto the black hole, can be calculated for hydrodynamically stable ideal fluids with $\partial p/\partial\rho > 0$. This can be done by determining the fluid parameters at the critical point. Following Michel [33], we find the relationship between the parameters at the critical point:

$$u_*^2 = \frac{1}{2x_*}, \quad V_*^2 = \frac{u_*^2}{1 - 3u_*^2}, \quad (14)$$

where

$$V^2 = \frac{n}{\rho + p}\frac{d(\rho + p)}{dn} - 1. \quad (15)$$

Together with (7), this yields

$$V^2 = c_s^2(\rho), \quad (16)$$

where $c_s^2 = \partial p/\partial\rho$ is the square of the effective speed of sound in the medium. We derive the following relation for the critical point from Eqs. (14), (16), (11), and (10):

$$\frac{\rho_* + p(\rho_*)}{n(\rho_*)} = [1 + 3c_s^2(\rho_*)]^{1/2}\frac{\rho_\infty + p(\rho_\infty)}{n(\rho_\infty)}, \quad (17)$$

which fixes the fluid density at the critical point ρ_* for an arbitrary equation of state $p = p(\rho)$. Specifying ρ_* and using (8), we can determine $n(\rho_*)$. Accordingly, the quantities x_* and u_* can be calculated from (14) and (16). As a result, the numerical constant A can be calculated by substituting the derived quantities into (9). For $c_s^2 < 0$ or $c_s^2 > 1$, no critical point exists beyond the event horizon of the black hole ($x_* > 1$), implying that the dark energy flux onto the black hole depends on the initial conditions for an unstable ideal fluid ($c_s^2 < 0$) or a "superluminal" fluid ($c_s^2 > 1$). This result has a simple physical interpretation: the accreted fluid has a critical point if its speed increases from subsonic to supersonic values as it approaches the black hole. In contrast, for $c_s^2 < 0$ or $c_s^2 > 1$, the critical point either does not exist or is formally within the event horizon of the black hole. It should also be noted that fluids with $c_s^2 < 0$ are hydrodynamically unstable (see [46, 47] for a discussion).

Equation (10), together with (6), (8), and (9), defines the accretion rate onto a black hole. These equations are valid for an ideal fluid with an arbitrary equation of state $p = p(\rho)$, in particular, for a gas of massless particles (thermalized radiation) and a gas of massive particles. For a gas of massive particles, the system of equations (9) and (10) reduces to a similar system of equations found by Michel [33]. It should be noted, however, that Eqs. (6), (8), (9), and (10) are also valid for dark energy, including phantom energy with $\rho + p < 0$. In these cases, the concentration $n(\rho)$ is positive for any ρ, while the constant C_2 in Eq. (10) is negative.

The rate of change in the black hole mass (the energy flux onto the black hole) through accretion is

$$\dot{M} = -4\pi r^2 T_0^r.$$

Using (9) and (10), this expression can be rewritten as [30]

$$\dot{M} = 4\pi A M^2[\rho_\infty + p(\rho_\infty)]. \quad (18)$$

It follows from Eq. (18) than the mass of the black hole increases as it accretes the gas of particles when $p > 0$, but decreases as it accretes the phantom energy when $\rho + p < 0$. In particular, this implies that the black hole masses in the Universe filled with phantom energy must decrease. This result is general in nature. It does not depend on the specific form of the equation of state $p = p(\rho)$; only the satisfaction of the condition $p + \rho < 0$ is important. The physical cause of the decrease in the black hole mass is as follows: the phantom energy falls to the black hole, but the energy flux associated with this fall is directed away from the black hole.

If we ignore the cosmological evolution of the density ρ_∞, then we find the law of change in the black hole mass from (18) to be

$$M = M_i\left(1 - \frac{t}{\tau}\right)^{-1}, \quad (19)$$

where M_i is the initial mass of the black hole, and τ is the evolution time scale:

$$\tau = \frac{1}{4\pi A M_i[\rho_\infty + p(\rho_\infty)]}. \quad (20)$$

3. ANALYTICAL ACCRETION MODELS

3.1. Model of a Linear Equation of State

Let us consider the model of dark energy with a linear density dependence of pressure [30]:

$$p = \alpha(\rho - \rho_0), \quad (21)$$

where α and ρ_0 are constants. Among the other cases, this model describes an ultrarelativistic gas ($p = \rho/3$), a gas with an ultrahard equation of state ($p = \rho$), and the simplest model of dark energy ($\rho_0 = 0$ and $\alpha < 0$). The quantity α is related to the parameter $w = p/\rho$ of the equation of state by $w = \alpha(\rho - \rho_0)/\rho$.

An equation of state with $w = \text{const} < 0$ throughout the cosmological evolution is commonly used to analyze cosmological models. The matter with such an equation of state is hydrodynamically unstable and can exist only for a short period. Our equation of state (21) for $\alpha > 0$ does not have this shortcoming. For $\alpha > 0$, it also allows the case of hydrodynamically stable phantom energy to be described, which is not possible when using an equation of state with $w = \text{const} < -1$. In the real Universe, the equation of state changes with time (i.e., w depends on t). Therefore, Eq. (21) has the meaning of an approximation to the true equation of state only in a limited ρ range. From the physical point of view, the condition $\rho > 0$ must be satisfied for any equation of state in a comoving frame of reference. In particular, the state of matter with $\rho = 0$, but $p \neq 0$, is physically unacceptable. The corresponding constraints for the equation of state (21) are specified by conditions (29) and (30) given below.

For $\alpha < 0$, there is no critical point for the accreted fluid flow. For $\alpha > 0$, using (14) and (16), we obtain the parameters of the critical point

$$x_* = \frac{1 + 3\alpha}{2\alpha}, \quad u_*^2 = \frac{\alpha}{1 + 3\alpha}. \quad (22)$$

Note that the parameters of the critical point (22) in the linear model (21) are determined only by $\partial p/\partial\rho = \alpha$ and do not depend on ρ_0, which fixes the physical nature of the fluid under consideration: a relativistic gas, dark energy, or phantom energy. Note also that no critical point exists beyond the event horizon of the black hole for $\alpha > 1$ (this corresponds to a nonphysical situation with a superluminal speed of sound).

Let us calculate the constant A, which defines the energy flux onto the black hole. We find from (8) that

$$\frac{n}{n_\infty} = \left|\frac{\rho_{\text{eff}}}{\rho_{\text{eff},\infty}}\right|^{1/(1+\alpha)}, \quad (23)$$

where we introduced the effective density

$$\rho_{\text{eff}} \equiv \rho + p = -\rho_0\alpha + (1 + \alpha)\rho.$$

Using (17), we obtain

$$\left(\frac{\rho_{\text{eff}_*}}{\rho_{\text{eff},\infty}}\right)^{\alpha/(1+\alpha)} = (1 + 3\alpha)^{1/2}, \quad (24)$$

where ρ_{eff_*} and $\rho_{\text{eff},\infty}$ are the effective densities at the critical point and at infinity, respectively. Substituting (24) into (23) and using (9), we obtain for the linear model

$$A = \frac{(1 + 3\alpha)^{(1+3\alpha)/2\alpha}}{4\alpha^{3/2}}. \quad (25)$$

It is easy to see that $A \geq 4$ for $0 < \alpha < 1$. $A = 4$ for $\alpha = 1$ (this corresponds to $c_s = 1$); i.e., the constant A is on the order of 1 for relativistic speeds of sound. Using (25), we obtain from (20)

$$\tau = \left[\pi M_i(\rho_\infty + p_\infty)\frac{(1 + 3\alpha)^{(1+3\alpha)/2\alpha}}{\alpha^{3/2}}\right]^{-1}. \quad (26)$$

To determine the fluid density on the event horizon of the black hole, we substitute (23) into (12) to yield

$$\rho_H = \frac{\alpha\rho_0}{1 + \alpha} + \left(\rho_\infty - \frac{\alpha\rho_0}{1 + \alpha}\right)\left(\frac{A}{4}\right)^{(1+\alpha)/(1-\alpha)}, \quad (27)$$

where A is given by (25). For $0 < \alpha < 1$, the effective density on the horizon $\rho_{\text{eff},H}$ cannot be lower than $\rho_{\text{eff},\infty}$. The radial 4-velocity component on the horizon can be found from (13) and (27):

$$u_H = -\left(\frac{A}{4}\right)^{-\alpha/(1-\alpha)}. \quad (28)$$

The value of u_H changes from 1 to 1/2 for $0 < \alpha < 1$.

The linear model (21) describes the phantom energy when

$$\frac{\rho_\infty}{\rho_0} < \frac{\alpha}{1 + \alpha}.$$

In this case, $\rho + p < 0$. However, the requirement that the density ρ be nonnegative should be taken into account. This parameter can formally be negative in the range $0 < \alpha \leq 1$. Such a nonphysical situation imposes a constraint on the linear model (21) under consideration. For a physically proper description of the accretion process, we must require that the density ρ be nonnegative. We obtain the following constraint on the validity range of the linear model from (27) for hydrodynamically stable phantom energy:

$$\frac{\alpha}{1 + \alpha}\left[1 - \left(\frac{A}{4}\right)^{-(1+\alpha)/(1-\alpha)}\right] < \frac{\rho_\infty}{\rho_0} < \frac{\alpha}{1 + \alpha}. \quad (29)$$

As follows from (29), at a given α, we can always choose the parameters ρ_0 and ρ_∞ in such a way that $\rho > 0$ for any $r > 2M$.

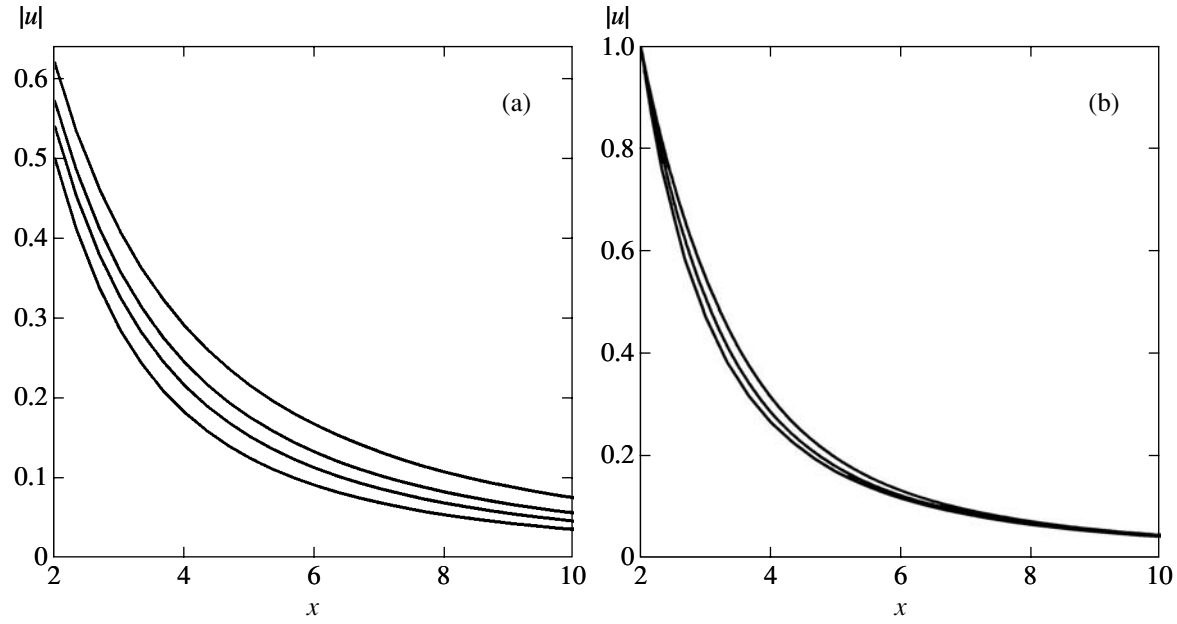

Fig. 1. Accreted fluid velocity u in the linear model (21) versus radial coordinate x.

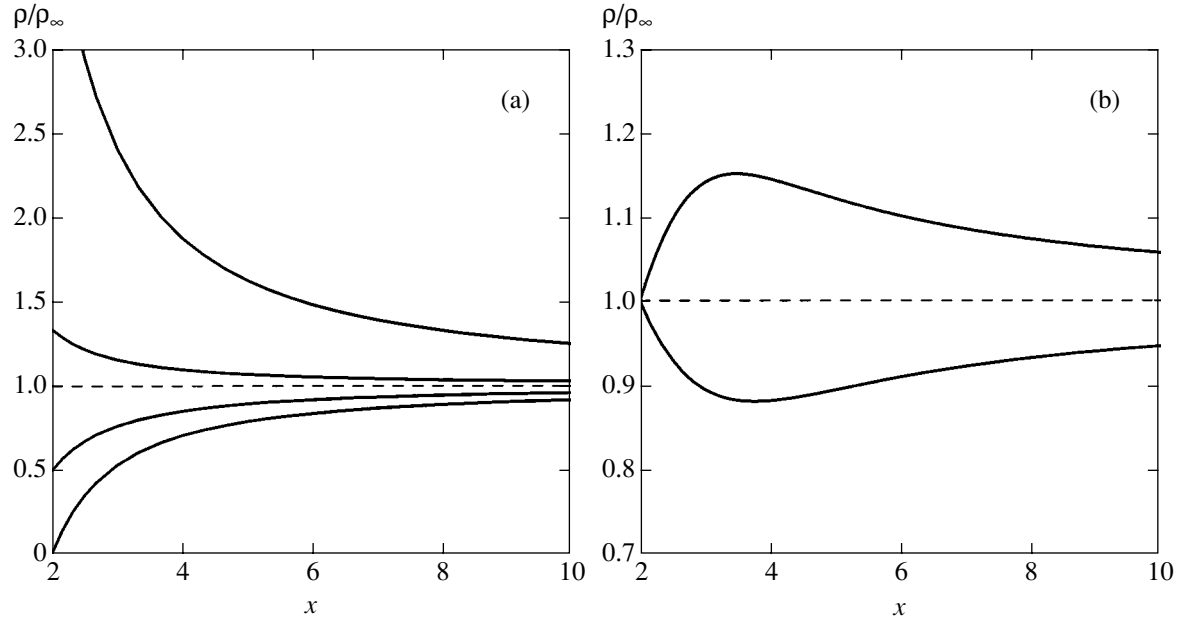

Fig. 2. Accreted fluid density normalized to the density at infinity, ρ/ρ_∞, versus radial coordinate x for the linear model (21) (solid curves). The dashed lines indicate the density of the Λ term, ρ/ρ_∞.

On the other hand, model (21) describes the quintessence (not the phantom energy) for the entire r range only if $p < 0$. Consequently, a physically proper description of the quintessence can be obtained from (27) if

$$\frac{\alpha}{1+\alpha} < \frac{\rho_\infty}{\rho_0} < \frac{\alpha}{1+\alpha}\left[\frac{1}{\alpha} + \left(\frac{A}{4}\right)^{-(1+\alpha)/(1-\alpha)}\right]. \qquad (30)$$

For some of the specific choices of α (more specifically, for $\alpha = 1/3$, 1/2, 2/3, and 1), $\rho(x)$ and $u(x)$ can be calculated analytically (see the Appendix for details on these calculations). In Figs. 1 and 2, the radial 4-velocity component, u, and the density normalized to the density at infinity, ρ/ρ_∞, are plotted against the coordinate $x = r/2M$. Figure 1a shows the plots of $u(x)$ for hydrodynamically stable fluids with $c_s^2 > 0$ at $\alpha = 1/3$, 1/2, 2/3,

and 1 (the curves are arranged from top to bottom, respectively). Figure 1b shows the plots of $u(x)$ for $c_s^2 < 0$ at $\alpha = -1.1$, -2, and $-1/2$ (the curves are also arranged from top to bottom). For this case, we chose the boundary condition $u_H = 1$ on the horizon. Figure 2a shows the plots of ρ/ρ_∞ for a hydrodynamically stable fluid with $\alpha = 1$ for various cases: $\rho_0 = 0$ (the model of neutron star matter); $\rho_0/\rho_\infty = 16/9$ (the linear model of nonphantom dark energy); $\rho_0/\rho_\infty = 7/3$ (the linear model of phantom energy); and $\rho_0/\rho_\infty = 7/3$ (the linear model of phantom energy with $\rho_H = 0$) (the curves are arranged from top to bottom, respectively). Figure 2b shows the plots of ρ_0/ρ_∞ for $\alpha < 0$ for various cases: $\alpha = -2$, $\rho_0 = 0$, and $A = 4$ (the linear model of phantom energy, the upper curve); and $\alpha = -1/2$, $\rho_0 = 0$, and $A = 4$ (the linear model of nonphantom energy, the lower curve). For this case, we chose the velocity $|u_H| = 1$ on the horizon.

3.2. Accretion onto a Moving and Rotating Black Hole

Let us consider the accretion onto a moving and rotating black hole in the special case of a linear equation of state with $\alpha = 1$. The condition $\alpha = 1$ allows an exact analytical expression to be derived for the accretion rate of dark energy onto a black hole.

For $\alpha = 1$, we easily find from (23) that

$$\frac{n}{n_\infty} = \left|\frac{\rho_{\text{eff}}}{\rho_{\text{eff},\infty}}\right|^{1/2}. \tag{31}$$

We obtain the following continuity equation for the particle concentration from (5):

$$(nu^\mu)_{;\mu} = 0.$$

We can introduce the scalar field ϕ in terms of which the fluid velocity can be expressed as follows (there is no torsion in the fluid):

$$\frac{\rho + p}{n}u_\mu = \phi_{,\mu}. \tag{32}$$

We derive an equation for the auxiliary function ϕ by using Eqs. (31) and (32),

$$\phi^\mu_{;\mu} = 0. \tag{33}$$

Exactly the same equation arises in the problem of the accretion of a fluid with the equation of state $p = \rho$ [41]. Thus, we reduced the problem of a black hole moving in dark energy with the equation of state $p = \rho - \rho_0$ to the problem of a fluid with an extremely hard equation of state, $p = \rho$. Using the method suggested in [41], we obtain the mass evolution law for a moving and rotating black hole immersed in dark energy with the equation of state $p = \rho - \rho_0$:

$$\dot{M} = 4\pi(r_+^2 + a^2)[\rho_\infty + p(p_\infty)]u^0_{BH}, \tag{34}$$

where

$$r_+ = M + (M^2 - a^2)^{1/2}$$

is the radius of the event horizon for a rotating black hole, $a = J/M$ is the specific angular momentum of the black hole (rotation parameter), and u^0_∞ is the zeroth 4-velocity component of the black hole relative to the fluid. Expression (34) for $u^0_{BH} = 0$ reduces to (18) for a Schwarzschild ($a = 0$) black hole at rest.

3.3. Chaplygin Gas

Let us consider a Chaplygin gas with the following equation of state as another example of the solvable model:

$$p = -\frac{\alpha}{\rho}, \tag{35}$$

where $\alpha > 0$. The range of parameters $\rho^2 < \alpha$ represents the phantom energy with a superluminal speed of sound, implying that the phantom energy flux onto the black hole is not fixed by the condition of its passage through the critical point. The case of $\rho^2 > \alpha$ corresponds to dark energy with $\rho + p > 0$ and $0 < c_s^2 < 1$. We can easily find from Eq. (8) that

$$\frac{n}{n_\infty} = \left|\frac{\rho^2 - \alpha}{\rho_\infty^2 - \alpha}\right|^{1/2}. \tag{36}$$

The density at the critical point can be calculated from (17) and (36):

$$\rho_*^2 = 4\rho_\infty^2 - 3\alpha. \tag{37}$$

The velocity and the radial coordinate at the critical point are given by

$$x_* = \frac{2\rho_\infty^2}{\alpha}, \quad u_*^2 = \frac{\alpha}{4\rho_\infty^2}. \tag{38}$$

We then find the constant A from Eq. (9):

$$A = 4\left(\frac{\rho_\infty^2}{\alpha}\right)^{3/2}. \tag{39}$$

For $0 < c_s^2 < 1$, the constant A cannot be smaller than 4, as in the case of the linear model. The evolution time scale of the black hole mass without any cosmological change in the dark energy density is given by

$$\tau = 8\pi M_i\frac{\rho_\infty^2}{\alpha}(\rho_\infty + p_\infty). \tag{40}$$

Note that Eqs. (36)–(39) are applicable only for dark energy with $\rho + p > 0$ and are invalid for phantom

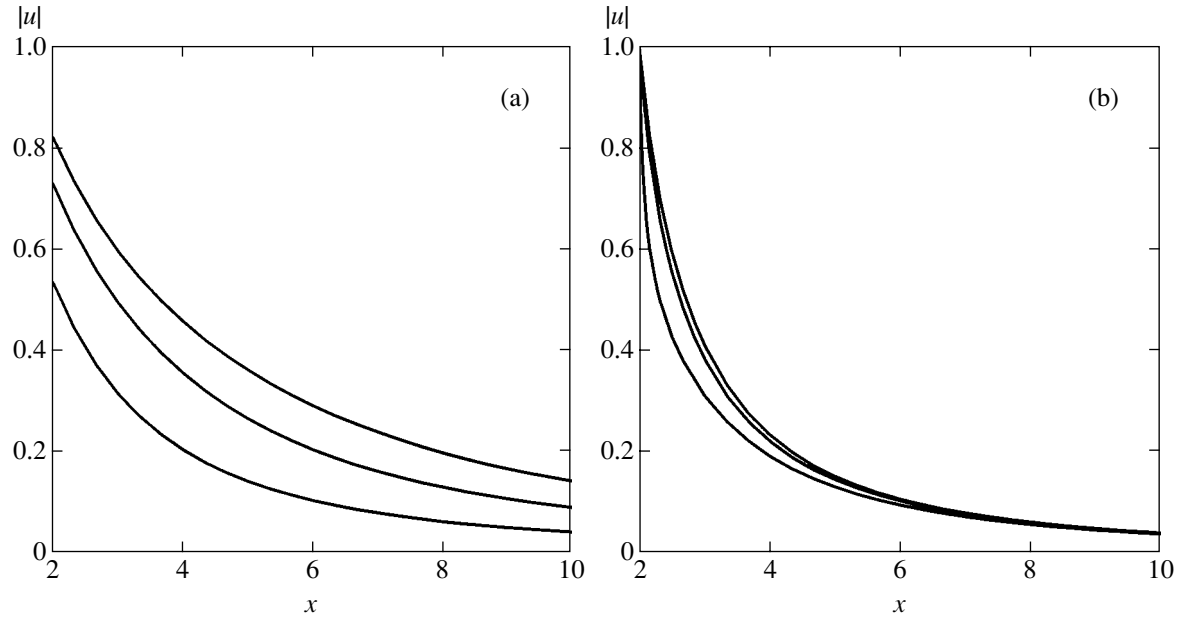

Fig. 3. Velocity u versus coordinate x for Chaplygin gas [35]

energy. On the black hole horizon,

$$\rho_H = \frac{A}{4}\rho_\infty,$$
$$u_H = -\frac{A}{4}\left[\frac{\rho_\infty^2 - \alpha}{(A/4)^2\rho_\infty^2 - \alpha}\right]^{1/2}. \quad (41)$$

For $0 < c_s^2 < 1$, the density on the horizon ρ_H cannot be lower than ρ_∞, and u_H changes from 1 to 1/2. The Chaplygin gas density distribution can be determined from the general equations (9) and (10):

$$\rho = \rho_\infty\left(\frac{\rho_\infty^2}{\alpha}\right)^{3/2}\left(\frac{2}{x} - \frac{\alpha}{\rho_\infty^2}\right)^{-1/2}$$
$$\times\left[\frac{16}{x^4}\left(1 - \frac{\alpha}{\rho_\infty^2}\right) - \left(\frac{\alpha}{\rho_\infty^2}\right)^4\left(1 - \frac{2}{x}\right)\right]^{1/2}. \quad (42)$$

The velocity distribution $u(r)$ can be calculated by using Eqs. (9), (36), and (42). In Figs. 3 and 4, the velocity u and the density normalized to the density at infinity, ρ/ρ_∞, are plotted against the coordinate $x = r/2M$. Figure 3a shows the plots of $u(r)$ for nonphantom dark energy at $\rho_\infty^2/\alpha = 3$, 2, and 1.1 (the curves are arranged from top to bottom, respectively). Figure 3b shows the plots of $u(r)$ for phantom energy at $\rho_\infty^2/\alpha = 0.3$, 0.5, and 0.9 (the curves are also arranged from top to bottom, respectively). In this case, the boundary condition $u_H = 1$ is set on the horizon. Figure 4a shows the plots of the normalized density, ρ/ρ_∞, for nonphantom dark energy at $\rho_\infty^2/\alpha = 3$, 2, and 1.1 (the curves are arranged from top to bottom, respectively). Figure 4b shows the plots of the normalized density, ρ/ρ_∞, for phantom energy at $\rho_\infty^2/\alpha = 0.9$, 0.5, and 0.3 (the curves are also arranged from top to bottom, respectively). For this case, we chose the boundary condition $u_H = 1$ on the horizon.

4. THE FATE OF BLACK HOLES DURING THE BIG RIP

Let us now consider the evolution of black holes in the cosmological Big Rip scenario, where the scale factor $a(t)$ increases to infinity in a finite time [16, 17]. For simplicity, we take into account only the dark energy and disregard the other forms of energy. In the linear model (21), the Big Rip takes place at $\rho + p < 0$ and $\alpha < -1$. The following relation can be derived from Friedmann's equations in the case of a linear equation of state:

$$|\rho + p| \propto a^{-3(1+a)}.$$

Setting, for simplicity, $\rho_0 = 0$, we find the evolution law of the phantom energy density in this Universe:

$$\rho_\infty = \rho_{\infty,i}\left(1 - \frac{t}{\tau}\right)^{-2}, \quad (43)$$

where

$$\tau^{-1} = -\frac{3(1+\alpha)}{2}\left(\frac{8\pi}{3}\rho_{\infty,i}\right)^{1/2}. \quad (44)$$

Here, $\rho_{\infty,i}$ is the initial cosmological phantom energy, and the initial time was chosen in such a way that the

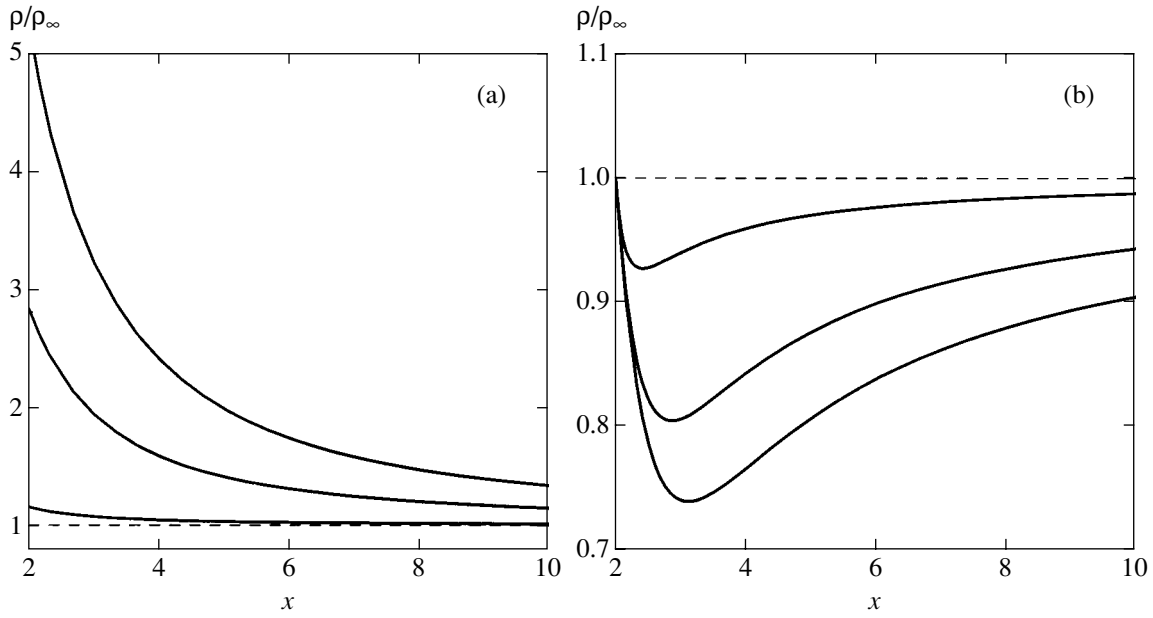

Fig. 4. Density normalized to the density at infinity, ρ/ρ_∞, versus coordinate x for model (35) (solid curves). The dashed line indicates the normalized density of the Λ term, ρ/ρ_∞.

Big Rip occurs at time τ. We easily see from Eqs. (20) and (43) that the Big Rip takes place at $\alpha \equiv \partial p/\partial\rho < -1$. In general, the condition $\rho + p < 0$ alone is not enough for the cosmological evolution to be ended with the Big Rip [18].

Using Eq. (43), we find the evolution of the black hole mass in the cosmological Big Rip scenario from Eq. (18):

$$M = M_i\left(1 + \frac{M_i}{\dot{M}_0\tau}\frac{t}{\tau - t}\right)^{-1}, \qquad (45)$$

where

$$\dot{M}_0 = (3/2)A^{-1}|1 + \alpha|, \qquad (46)$$

and M_i is the initial mass of the black hole. At $\alpha = -2$ and a typical value of $A = 4$ (which correspond to $u_H = -1$), $\dot{M}_0 = 3/8$. In the limit $t \longrightarrow \tau$ (i.e., near the Big Rip), the t dependence of the black hole mass becomes linear, $M \approx \dot{M}_0(\tau - t)$. When t approaches τ, the rate of decrease in the black hole mass ceases to depend on the initial black hole mass and the phantom energy density:

$$\dot{M} \approx -\dot{M}_0.$$

In other words, the masses of all black holes near the Big Rip are approximately equal and approach zero. This implies that the accretion of phantom energy dominates over the Hawking evaporation until the black hole mass decreases to the Planck mass. Formally, however, all black holes in the Universe completely evaporate during the Hawking radiation in the Planck time before the Big Rip occurs.

5. THE ACCRETION OF A SCALAR FIELD

In this section, we compare our calculations of the accretion of an ideal fluid with similar calculations of the accretion of a scalar (nonphantom) field onto a black hole [24–29]. The dark energy is commonly modeled by a scalar field with a potential $V(\phi)$. The approximation of an ideal fluid is rougher, since the scalar field ϕ and $\partial_m\phi$ cannot be unambiguously reproduced for given ρ and p, which characterize an ideal fluid. Despite this difference between the scalar field and the ideal fluid, we will show that our results are in close agreement with the corresponding calculations of the accretion of a scalar field onto a black hole.

The Lagrangian of the scalar field is $L = K - V$, where K is the kinetic term and V is the potential. For the standard choice of the kinetic term

$$K = \phi_{;\mu}\phi^{;\mu}/2,$$

the corresponding energy flux onto the black hole is

$$T_{0r} = \phi_{,t}\phi_{,r}.$$

Jacobson [24] found a solution for the scalar field in the Schwarzschild metric for a zero potential, $V = 0$:

$$\phi = \dot{\phi}_\infty[t + 2M\ln(1 - 2M/r)],$$

where ϕ_∞ is the scalar field at infinity. Frolov and Kofman [26] showed that this solution is also valid for

many scalar fields with a nonzero potential $V(\phi)$ under certain conditions. For this solution,

$$T_0^r = -(2M)^2\dot{\phi}_\infty^2/r^2$$

and, accordingly,

$$\dot{M} = 4\pi(2M)^2\dot{\phi}_\infty^2.$$

The energy–momentum tensor constructed using Jacobson's solution is identical to the energy–momentum tensor for an ideal fluid with an extremely hard equation of state, $p = \rho$, after the substitution

$$p_\infty \longrightarrow \dot{\phi}_\infty^2/2, \quad \rho_\infty \longrightarrow \dot{\phi}_\infty^2/2.$$

This is not surprising, since the theory of a scalar field with a zero potential, $V(\phi) = 0$, is identical to the model of an ideal fluid [48]. In view of this correspondence, we easily see agreement between our result (18) for $\dot{M}$ in the case of $p = \rho$ and the corresponding results from [24, 26].

The Lagrangian of the scalar field that describes the phantom energy must have a negative kinetic term [16, 17], for example,

$$K = -\phi_{;\mu}\phi^{;\mu}/2$$

(see [49] for more general cases). In this case, the phantom energy flux onto the black hole has the opposite sign,

$$T_{0r} = -\phi_{,t}\phi_{,r},$$

where ϕ is the solution of the same Klein–Gordon equation as that for the standard scalar field, but with the substitution $V \longrightarrow -V$. For a zero potential, this solution is identical to Jacobson's solution [24] obtained for a scalar field with a positive kinetic term.

However, the Lagrangian with a negative kinetic term and $V(\phi) = 0$ does not describe the phantom energy. At the same time, the solution for a scalar field with $V(\phi) = 0$ is identical to the solution for a positive constant potential, $V_0 = \text{const}$, which can be chosen in such a way that

$$\rho = -\dot{\phi}^2/2 + V_0 > 0.$$

In this case, the scalar field describes the accreted phantom energy with $\rho > 0$ and $p < -\rho$, which leads to a decrease in the black hole mass at the rate

$$\dot{M} = -4\pi(2M)^2\dot{\phi}_\infty^2.$$

A simple example of phantom cosmology (but without the Big Rip) is realized by a scalar field with the potential

$$V = m^2\phi^2/2,$$

where $m \sim 10^{-33}$ eV [50]. After a short transition period, this cosmological model approaches an asymptotic state with

$$H \approx m\phi/3^{1/2}, \quad \dot{\phi} \approx 2m/3^{1/2}.$$

In the Klein–Gordon equation (with the substitution $V \longrightarrow -V$ mentioned above), the term m^2 becomes equal to the other terms only on the scale of the cosmological horizon, implying that, in this case, Jacobson's solution is also valid. Calculations of the corresponding energy flux onto the black hole yield

$$\dot{M} = -4\pi(2M)^2\dot{\phi}_\infty^2 = -64M^2m^2/3.$$

For $M_0 = M_\odot$ and $m = 10^{-33}$ eV, the effective time of the decrease in black hole mass is

$$\tau = (3/64)M^{-1}m^{-2} \sim 10^{32} \text{ yr}.$$

6. DISCUSSION AND CONCLUSIONS

In recent years, the concept of dark energy has been accepted and extensively discussed in cosmology. The possible existence of dark energy with a negative pressure leads to new cosmological scenarios, including the exotic model of the Universe in which all of the bound objects are destroyed and which dies itself as a result of the Big Rip. To determine the fate of black holes in this cosmological scenario, we considered the spherically symmetric, stationary accretion of dark energy modeled by an ideal fluid onto a black hole. We derived general equations for the accretion of an ideal fluid with the equation of state $p = p(\rho)$ onto a Schwarzschild black hole. In particular, these equations can be used to describe the accretion of thermal radiation, dark energy, and phantom energy. We also considered the accretion onto a moving and rotating black hole in the special case of an extremely hard equation of state, $p = \rho$. We calculated the change in the black hole mass through accretion. The black hole masses for $\rho + p > 0$ were found to increase, as in the usual case. However, a qualitatively new result was obtained for phantom energy, i.e., for a medium with $\rho + p < 0$. We found that the black hole masses decrease in this situation. Using this result, we solved the problem of the fate of black holes in a universe that undergoes the Big Rip. It turns out that all black holes in this Universe must decrease their masses and disappear completely by the Big Rip. We also considered the correspondence between the accretion of dark energy in the model of an ideal fluid and the accretion of a scalar field.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project nos. 02-02-16762a, 03-02-16436a, and 04-02-16757a and by the Ministry of Science of the Russian Federation, grants 1782.2003.2 and 2063.2003.2.

APPENDIX

Analytical Solutions for $\rho(x)$ and $u(x)$

In the model under consideration (Section 3), analytical solutions can be found for the dependence of the dark energy density and accretion rate on radius *r*. Using Eq. (23) for the concentration and Eq. (11) for the constant C_2, we derive the following equation for ρ_{eff} from Eqs. (9) and (10)

$$\left(\frac{\rho_{\text{eff}}}{\rho_{\text{eff},\infty}}\right)^{2\alpha/(1+\alpha)} \times\left[1-\frac{2}{x}+\frac{A^2}{x^4}\left(\frac{\rho_{\text{eff}}}{\rho_{\text{eff},\infty}}\right)^{-2/(1+\alpha)}\right] = 1. \tag{47}$$

Defining

$$y\equiv\left(\frac{\rho_{\text{eff}}}{\rho_{\text{eff},\infty}}\right)^{2/(1+\alpha)}, \tag{48}$$

we obtain the following equation from (47):

$$y\left(1-\frac{2}{x}\right)-y^{1-\alpha}+\frac{A^2}{x^4} = 0, \tag{49}$$

which can be solved analytically for certain values of α. For $\alpha = 1/3$, Eq. (49) reduces to a cubic equation:

$$z^3\left(1-\frac{2}{x}\right)-z^2+\frac{A^2}{x^4} = 0, \tag{50}$$

where $z = y^{1/3}$. Solving this equation yields the fluid density distribution for $\alpha = 1/3$:

$$\rho = \frac{\rho_0}{4}+\left(\rho_\infty-\frac{\rho_0}{4}\right)\left[z+\frac{1}{3(1-2x^{-1})}\right]^2, \tag{51}$$

where

$$z = \begin{cases} 2\sqrt{\frac{a}{3}}\cos\left(\frac{2\pi}{3}-\frac{\beta}{3}\right), & 2\le x\le 3,\\ 2\sqrt{\frac{a}{3}}\cos\frac{\beta}{3}, & x>3,\end{cases} \tag{52}$$

$$\beta = \arccos\left[\frac{b}{2(a/3)^{3/2}}\right] \tag{53}$$

and

$$a = \frac{1}{3(1-2/x)^2},\qquad b = \frac{2}{27(1-2/x)^3}-\frac{108}{(1-2/x)x^4}. \tag{54}$$

This solution corresponds to a thermalized photon gas in which the photon mean free path is much smaller than the radius of the black hole horizon, $\lambda_{fp} \ll 2M$. In this situation, the photon gas may be treated as an ideal fluid. In the opposite case, $\lambda_{fp} \gg 2M$, the photons are free particles, and their accretion rate is determined by the well-known cross section for the gravitational capture of relativistic particles by a black hole. The corresponding accretion rate is

$$\dot{M} = 27\pi M^2\rho_\infty.$$

The case of $\alpha = 2/3$ is similar to the case considered above. We obtain the following equation instead of (50):

$$z^3\left(1-\frac{2}{x}\right)-z+\frac{A^2}{x^4} = 0, \tag{55}$$

where again $z = y^{1/3}$. The fluid density distribution in this case is

$$\rho = \frac{2}{5}\rho_0+\left(\rho_\infty-\frac{2}{5}\rho_0\right)z^{5/2}, \tag{56}$$

where *z* is given by

$$z = \begin{cases} 2\sqrt{\frac{a}{3}}\cos\left(\frac{2\pi}{3}-\frac{\beta}{3}\right), & 2\le x\le 9/4,\\ 2\sqrt{\frac{a}{3}}\cos\frac{\beta}{3}, & x>9/4,\end{cases} \tag{57}$$

β is defined by Eq. (53), and

$$a = \frac{1}{1-2/x},\quad b = -\frac{2187\sqrt{3}}{128(1-2/x)x^4}.$$

For $\alpha = 1/2$, (49) is a quadratic equation and has a simple analytical solution:

$$\rho = \frac{\rho_0}{3}+\left(\rho_\infty-\frac{\rho_0}{3}\right)z^{3/2}, \tag{58}$$

where

$$z = \begin{cases} \frac{1}{2}\{1-[1-3125(1-2/x)(16x^4)^{-1}]^{1/2}\}(1-2/x)^{-1}, & 2\le x\le 5/2,\\ \frac{1}{2}\{1+[1-3125(1-2/x)(16x^4)^{-1}]^{1/2}\}(1-2/x)^{-1}, & x>5/2.\end{cases} \tag{59}$$

For $\alpha = 1$, Eq. (49) is linear in y, which gives

$$\rho = \frac{\rho_0}{2} + \left(\rho_\infty - \frac{\rho_0}{2}\right)\left(1 + \frac{2}{x}\right)\left(1 + \frac{4}{x^2}\right). \qquad (60)$$

REFERENCES

1. N. Bahcall, J. P. Ostriker, S. Perlmutter, and P. J. Steinhardt, Science **284**, 1481 (1999).
2. A. Riess, A. V. Filippenko, P. Challis, *et al.*, Astron. J. **116**, 1009 (1998).
3. S. J. Perlmutter, G. Aldering, G. Goldhaber, *et al.*, Astrophys. J. **517**, 565 (1999).
4. C. L. Bennett, M. Halpern, G. Hinshaw, *et al.*, Astrophys. J., Suppl. Ser. **148**, 1 (2003).
5. C. Wetterich, Nucl. Phys. B **302**, 668 (1988).
6. P. J. E. Peebles and B. Ratra, Astrophys. J. **325**, L17 (1988).
7. B. Ratra and P. J. E. Peebles, Phys. Rev. D **37**, 3406 (1988).
8. J. A. Frieman, C. T. Hill, A. Stebbins, and I. Waga, Phys. Rev. Lett. **75**, 2077 (1995).
9. R. R. Caldwell, R. Dave, and P. J. Steinhardt, Phys. Rev. Lett. **80**, 1582 (1998).
10. A. Albrecht and C. Skordis, Phys. Rev. Lett. **84**, 2076 (2000).
11. C. Armendariz-Picon, T. Damour, and V. Mukhanov, Phys. Lett. B **458**, 209 (1999).
12. C. Armendariz-Picon, V. Mukhanov, and P. J. Steinhardt, Phys. Rev. Lett. **85**, 4438 (2000).
13. T. Chiba, T. Okabe, and M. Yamaguchi, Phys. Rev. D **62**, 023511 (2000).
14. I. Zlatev, L. Wang, and P. Steinhardt, Phys. Rev. Lett. **82**, 896 (1999).
15. P. Steinhardt, L. Wang, and I. Zlatev, Phys. Rev. D **59**, 123504 (1999).
16. R. R. Caldwell, Phys. Lett. B **545**, 23 (2002).
17. R. R. Caldwell, M. Kamionkowski, and N. N. Weinberg, Phys. Rev. Lett. **91**, 071301 (2003).
18. B. McInnes, J. High Energy Phys. **0208**, 029 (2002); M. Bouhmadi-Lopez and J. A. J. Madrid, astro-ph/0404540.
19. U. Alam, V. Sahni, T. D. Saini, and A. A. Starobinsky, astro-ph/0311364.
20. S. Nojiri and S. D. Odintsov, Phys. Lett. B **562**, 147 (2003).
21. I. Brevik, S. Nojiri, S. D. Odintsov, and L. Vanzo, hep-th/0401073.
22. M. Visser, S. Kar, and N. Dadhich, Phys. Rev. Lett. **90**, 201102 (2003).
23. P. F. Gonzalez-Diaz, Phys. Rev. D **68**, 084016 (2003).
24. T. Jacobson, Phys. Rev. Lett. **83**, 2699 (1999).
25. R. Bean and J. Magueijo, Phys. Rev. D **66**, 063505 (2002).
26. A. Frolov and L. Kofman, J. Cosmol. Astrophys. Phys. **5**, 9 (2003).
27. W. G. Unruh, Phys. Rev. D **14**, 3251 (1976).
28. L. A. Urena-Lopez and A. R. Liddle, Phys. Rev. D **66**, 083005 (2002).
29. M. Yu. Kuchiev and V. V. Flambaum, gr-qc/0312065.
30. E. O. Babichev, V. I. Dokuchaev, and Yu. N. Eroshenko, Phys. Rev. Lett. **93**, 021102 (2004).
31. E. O. Babichev, V. I. Dokuchaev, and Yu. N. Eroshenko, Class. Quantum Grav. **22**, 143 (2005).
32. H. Bondi, Mon. Not. R. Astron. Soc. **112**, 195 (1952).
33. F. C. Michel, Astrophys. Space Sci. **15**, 153 (1972).
34. B. J. Carr and S. W. Hawking, Mon. Not. R. Astron. Soc. **168**, 399 (1974).
35. M. C. Begelman, Astron. Astrophys. **70**, 583 (1978).
36. D. Ray, Astron. Astrophys. **82**, 368 (1980).
37. K. S. Thorne, R. A. Flammang, and A. N. Zytkow, Mon. Not. R. Astron. Soc. **194**, 475 (1981).
38. E. Bettwieser and W. Glatzel, Astron. Astrophys. **94**, 306 (1981).
39. K. M. Chang, Astron. Astrophys. **142**, 212 (1985).
40. U. S. Pandey, Astrophys. Space Sci. **136**, 195 (1987).
41. L. I. Petrich, S. L. Shapiro, and S. A. Teukolsky, Phys. Rev. Lett. **60**, 1781 (1988).
42. B. J. Carr and S. W. Hawking, Mon. Not. R. Astron. Soc. **168**, 399 (1974).
43. E. Bettwieser and W. Glatzel, Astron. Astrophys. **94**, 306 (1981).
44. S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge Univ. Press, Cambridge, 1973), Chap. 4.3.
45. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
46. J. C. Fabris and J. Martin, Phys. Rev. D **55**, 5205 (1997).
47. S. M. Carroll, M. Hoffman, and M. Trodden, Phys. Rev. D **68**, 023509 (2003).
48. V. N. Lukash, Zh. Éksp. Teor. Fiz. **79**, 1601 (1980) [Sov. Phys. JETP **52**, 807 (1980)].
49. P. F. Gonzalez-Diaz, Phys. Lett. B **586**, 1 (2004).
50. M. Sami and A. Toporensky, Mod. Phys. Lett. A **19**, 1509 (2004).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 539–549.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 610–622.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Structure of the Magnetic Field of the Jovian Magnetosphere

Ya. N. Istomin

Lebedev Physical Institute, Russian Academy of Sciences, Leninskii pr. 53, Moscow, 119991 Russia
e-mail: istomin@td.lpi.ac.ru

Received October 8, 2004

Abstract—We exactly solved the problem of the interaction between the rotating magnetic field of Jupiter and the equatorial plasma disk formed by the gases flowing from the Jovian satellite Io. The disk is shown to expel the Jovian magnetic field in both directions, inward, toward Jupiter, compressing its dipole magnetic field, and outward. Jupiter spins up the disk up to velocities that correspond to nearly constant angular rotation, but with an angular frequency lower than the angular frequency of Jupiter itself. The radial velocity of the plasma in the disk approaches its azimuthal velocity. We determined the power of Jupiter's rotational energy losses. Part of this energy is transferred to the disk, and the other part goes into heating the Jovian ionosphere. We show that the Pedersen surface conductivity of the Jovian ionosphere must have a lower limit to maintain the electric current that arises in the disk–rotating magnetic field system. This current in the Jovian magnetosphere flows only along the preferential magnetic surfaces that connect the inner and outer edges of the disk to the ionosphere. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The structure of the magnetic field of the Jovian magnetosphere inside the magnetopause, i.e., the cavity produced by the flow of the solar wind around the magnetic field of Jupiter with a scale length of ~100 Jupiter radii [1], is determined by the interaction of the magnetic field frozen in the rotating central body with the equatorial plasma disk surrounding this body. In astrophysics, a similar situation arises when the strong magnetic field of a neutron star or a white dwarf interacts with an accretion disk rotating around the star. This problem has never been solved completely [2]. In the case of Jupiter, however, the disk is produced inside the magnetosphere by its satellite Io. Here, the Jovian magnetic field is strong, and its rotation together with Jupiter is faster than the Keplerian rotation of the disk in this region. This implies that the centrifugal force will accelerate the disk in the radial direction and expel it outward. On the other hand, the disk plasma as an almost ideal conductor seeks to expel the Jovian magnetic field from itself. If it is expelled completely, then there will be no interaction between the rotating magnetic field and the disk, and the plasma produced by Io will not be able to escape from the magnetosphere due to the conservation of angular momentum. The compromise lies in the fact that the disk expels the magnetic field almost completely; it becomes nearly parallel to the disk surface, leaving only a weak vertical component B_z. This component is much weaker than the dipole field of Jupiter in the equatorial plane. While interacting with the rotating disk, the magnetic field B_z generates a radial electric field E_ρ. The electric field E_ρ works on the disk ions, imparting the rotation and radial velocities to them. Clearly, the dipole structure of the Jovian magnetic field cannot be preserved near the disk; the disk will expel the dipole field in both directions, toward and away from Jupiter. Near the inner edge of the disk, the magnetic field must be strongly compressed, and the magnetic field lines will be curved in the direction opposite to the curvature of the dipole field. The magnetic field will be severely distorted near the equator, while the distortion in the remaining magnetosphere will be insignificant. The structure of the magnetic field near the disk is of great importance, since it determines the interaction of the rotating magnetic field with the conductive disk. Therefore, the assumption that the magnetic field differs only slightly from the dipole field in the magnetosphere is invalid and yields incorrect results [3, 4], as we will see below.

We consider an axisymmetric magnetosphere, since the axis of the magnetic dipole frozen in the planetary core is nearly orthogonal to the equatorial plane in which the plasma disk is located. As a result, the magnetic field in the magnetosphere produced by the dipole and by the electric currents flowing in the magnetosphere and in the disk is axisymmetric. The axisymmetric magnetic field is described in Section 2. In Section 3, we consider a cold thin plasma disk. The approximation of a thin plasma disk implies that its thickness is much smaller than its radius. This approximation is applicable not only to the plasma disk far from Io, but also to the plasma torus near Io. In Section 4, we study the closure of the electric currents that are produced by the electric field generated by the rotating magnetic field in the Jovian magnetosphere and ionosphere. The Jovian

magnetosphere is assumed to be ideal and to produce no resistance to the electric current. In contrast, the ionosphere has a finite conductivity. Finally, in Section 5, we determine the structure of the magnetic field distorted by the loop of the current that flows through the disk and the ionosphere and that is closed by the magnetospheric currents.

2. THE AXISYMMETRIC MAGNETIC FIELD

Let us first introduce convenient variables that describe the magnetic field of the Jovian magnetosphere in the simplest way. Since the magnetic dipole is inclined to the rotation axis of Jupiter at an angle of several degrees, we assume the magnetosphere to be axisymmetric. All quantities depend only on two cylindrical coordinates, the distance along the z axis and the radius ρ. Let us introduce the flux of the poloidal magnetic field $f(\rho, z)$. The magnetic field can be expressed in terms of f as

$$B_\rho = -\frac{1}{\rho}\frac{\partial f}{\partial z}, \quad B_z = \frac{1}{\rho}\frac{\partial f}{\partial \rho}, \quad B_\phi = \frac{1}{\rho}g. \tag{1}$$

The function $g(\rho, z)$ describes the toroidal magnetic field that emerges in the magnetosphere when the electric current flows along the poloidal magnetic field. The relation $f = \text{const}$ is an equation for the magnetic surfaces on which the magnetic field lines lie. It is convenient to describe the poloidal magnetic field lines not only by the relation $f = f(\rho, z)$, but also by $\rho = \rho(z, f)$, where f numbers the magnetic surfaces. The poloidal magnetic field is then

$$B_\rho = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial z}\right)_f\left(\frac{\partial \rho}{\partial f}\right)_z^{-1}, \quad B_z = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial f}\right)_z^{-1}.$$

For a dipole magnetic field, the magnetic flux is

$$f_d = \frac{B_J R_J^3 \rho^2}{(\rho^2 + z^2)^{3/2}}.$$

Here, R_J is the radius of Jupiter, and B_J is the amplitude of Jupiter's surface magnetic field on its equator:

$$B_J = |B_z(r = R_J)|_{eq}.$$

It is important to know the magnetic flux in the equatorial $z = 0$ plane, $f_0(\rho) = f(\rho, z = 0)$. The vertical magnetic field on the equator is then

$$B_z(z = 0) = \rho^{-1}\frac{\partial f_0}{\partial \rho}.$$

It is convenient to measure the magnetic field strength in units of B_J and the distances ρ and z in Jupiter radii; the unit for the magnetic flux is then $B_J R_J^2$. Relations (1) remain the same in these dimensionless variables. For a dipole field, the dimensionless magnetic flux is

$$f'_d = \frac{\rho'^2}{(\rho'^2 + z'^2)^{3/2}}, \quad f'_{0d} = \frac{1}{\rho'}.$$

In the Jovian magnetosphere, the electric currents flow along magnetic field lines:

$$\mathbf{j} = a\mathbf{B}.$$

The quantity $a(\mathbf{r})$ is an arbitrary scalar. From the poloidal components of Maxwell's equation

$$\operatorname{curl}\mathbf{B} = \mu_0\mathbf{j}$$

we then derive the relations

$$g(\rho, z) = g(f), \quad a = \mu_0^{-1}\frac{dg}{df}.$$

These imply that the toroidal magnetic field and the electric currents in an axisymmetric magnetosphere are functions of the poloidal magnetic flux f. As we will see below, there are no bulk toroidal electric currents in the Jovian magnetosphere, $j_\phi = 0$. In this case,

$$(\operatorname{curl}\mathbf{B})_\phi = 0$$

and the equation for the magnetic flux has the form of Laplace's equation:

$$\rho\frac{\partial}{\partial \rho}\left(\frac{1}{\rho}\frac{\partial f}{\partial \rho}\right) + \frac{\partial^2 f}{\partial z^2} = 0. \tag{2}$$

The function f is neither even nor odd in z due to the electric currents flowing in the plasma disk in the equatorial plane. Therefore, we expand the function f at $z > 0$ in terms of $\exp(-\lambda z)$. The solution of Eq. (2) can be represented as

$$f(\rho, z) = \int_0^\infty \exp(-\lambda z)\lambda\rho J_1(\lambda\rho)\varphi(\lambda)d\lambda, \quad z > 0, \tag{3}$$

where J_1 is the Bessel function of the first order. The arbitrary function $\varphi(\lambda)$ in the integrand is defined by the boundary conditions for Eq. (2). For a dipole magnetic field, $\varphi(\lambda) = 1$ (in dimensionless units). The deviation of φ from unity describes the distortion of the Jovian dipole magnetic field. It is convenient to express the function φ in terms of the magnetic flux on the $z = 0$ equator, $f_0(\rho)$. Since

$$f_0(\rho) = \int_0^\infty \lambda\rho J_1(\lambda\rho)\phi(\lambda)d\lambda,$$

using the inverse Bessel transform, we obtain

$$\varphi(\lambda) = \int_0^\infty f_0(\rho')J_1(\lambda\rho')d\rho'. \tag{4}$$

Equations (3) and (4) define the magnetic field via the boundary value of $f_0(\rho)$. Of course, we should also take

into account the boundary conditions at other boundaries, more specifically, at infinity and on the $\rho = 0$ axis. At these boundaries, $f(\rho, z)$ must be finite, except for the coordinate origin $\rho^2 + z^2 \longrightarrow 0$, where $f \longrightarrow f_d$.

3. THE PLASMA DISK

The volcanic activity on Io is the source of gas that continuously flows into the Jovian magnetosphere. This neutral gas is ionized near Io by solar radiation. Thus, plasma is produced inside the Jovian magnetosphere. The plasma consists mostly of singly and doubly ionized sulfur and oxygen ions and electrons. The mean ion charge is Ze. The mean ion charge-to-mass ratio is $\sim 10^{-1} e/m_p$, where m_p is the proton mass.

The Jovian strong magnetic field causes the plasma to rotate with Jupiter's angular velocity, $\omega_J = 1.8 \times 10^{-4}\ \mathrm{s}^{-1}$. The centrifugal force prevents the ions from propagating into the Jovian magnetosphere, and these form an equatorial plasma disk. The height of the disk is small compared to its radius. The disk ions have radial, v_ρ, and azimuthal, v_ϕ, velocities. We consider a stationary and axisymmetric Jovian magnetosphere. All quantities are independent of the azimuthal angle ϕ due to the axial symmetry. The continuity equation for ions is

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho n_i v_\rho) = \frac{\dot{M}}{2\pi\rho m_i}\delta(\rho - \rho_1)\delta(z). \qquad (5)$$

Here, $\dot{M}$ is the gas production rate of Io, and ρ_1 is the inner edge of the disk, $\rho_1 > \rho_{\mathrm{Io}}$ (ρ_{Io} is the orbital radius of Io). The value of ρ_{Io} is 4.2×10^5 km $\approx 5.9 R_J$ ($R_J = 7.14 \times 10^4$ km). The gas production rate by Io's volcanoes is estimated to be $\dot{M} \approx 2 \times 10^3$ kg s^{-1}. Integrating Eq. (5) over ρ and z yields

$$v_\rho \Sigma = \frac{\dot{M}}{2\pi\rho m_i}\Theta(\rho - \rho_1). \qquad (6)$$

The quantity Σ is the ion surface density

$$\Sigma = \int n_i dz.$$

$\Theta(x)$ is the Heaviside unit step function. Note the important fact that the divergence of the ion flux is zero at $\rho > \rho_1$.

Let us now write the equation of motion for ions in the disk:

$$m_i(\mathbf{v}\cdot\nabla)\mathbf{v} = Ze\mathbf{E} + Ze[\mathbf{v}\times\mathbf{B}] + \mathbf{F}. \qquad (7)$$

Jupiter exerts a gravitational force

$$F_\rho = -m_i G M_J/\rho^2$$

on the ions that must cause the ions to rotate with the Keplerian velocity

$$v_\mathrm{K} = (GM_J/\rho)^{1/2}$$

(G is the gravitational constant, and M_J is the mass of Jupiter). The component of Eq. (7) directed along ρ yields

$$v_\rho \frac{\partial v_\rho}{\partial \rho} - \frac{v_\phi^2}{\rho} = \frac{Ze}{m_i}(E_\rho + v_\phi B_z) - \frac{v_\mathrm{K}^2}{\rho}. \qquad (8)$$

We see from Eq. (8) that, in the absence of electromagnetic fields B_z and E_ρ, the ions rotate with the Keplerian velocity:

$$v_\phi = v_\mathrm{K}.$$

In contrast, in a strong magnetic field B_z, the ions move at the electric drift velocity $v_\phi = -E_\rho/B_z$. For corotation, i.e., motion with Jupiter's rotation velocity $v_\phi = \omega_J\rho$, the disk plasma must be polarized in such a way that the electric field

$$E_\rho = -\omega_J \rho B_z$$

is produced. The corotation velocity is equal to the Keplerian velocity at a distance of $\rho = 2.24 R_J$. Since the Keplerian velocity decreases with distance from Jupiter, the corotation of the plasma disk implies that the centrifugal force in the disk is larger than the gravitational force, and the plasma must be accelerated in the radial direction, $\partial v_\rho/\partial\rho > 0$. However, this is not so easy to achieve due to the conservation of angular momentum. This follows from the other component of Eq. (7) directed along φ:

$$v_\rho \frac{\partial v_\phi}{\partial \rho} + \frac{v_\rho v_\phi}{\rho} = -\frac{Ze}{m_i} v_\rho B_z. \qquad (9)$$

Note that there is no toroidal electric field, $E_\phi = 0$, in this equation, because, being stationary, the electric field must be potential. Since the radial particle velocity v_ρ is nonzero (see Eq. (6)), Eq. (9) yields

$$\rho v_\phi + \frac{Ze}{m_i} f_0(\rho) = \text{const}. \qquad (10)$$

Equation (10) represents the law of conservation of total angular momentum. The coefficient in front of the second term in Eq. (10) is much larger than unity. It is proportional to the ratio of the ion cyclotron frequency in the Jovian magnetic field $\omega_c = ZeB_J/m_i$ to Jupiter's angular frequency ω_J. We denote this ratio by $\Omega_c = \omega_c/\omega_J$. It is the ion cyclotron frequency in units of Jupiter's angular frequency. At $Ze/m_i = 0.1 e/m_p$ and $B_J = 4$ G, Ω_c is approximately equal to 2.2×10^7. For a dipole magnetic field, the two terms on the left-hand side of Eq. (10) are of the same order of magnitude far from Jupiter: $\rho/R_J \approx \Omega_c^{1/3} \approx 2.8 \times 10^2$.

The conservation of angular momentum forbids the radial motion of ions in the Jovian strong magnetic field if the latter is assumed to be close to the dipole field. The only possibility for the disk plasma to move outward is to expel a significant fraction of the magnetic

field from the disk. Only if the magnetic flux changes slowly in the disk,

$$\Delta f_0 \approx B_J \rho^2 \Omega_c^{-1},$$

does the radial motion of ions become possible. However, the small change in magnetic flux f_0 with distance ρ implies that the vertical magnetic field B_z in the equatorial plane is weak (see Eq. (1)).

As regards the electrons supplied by Io, the parameter Ω_c for these is at least a factor of 10^4 larger than that for ions. The plasma electrons are strongly magnetized and can move only along the f = const magnetic surfaces, giving rise to electric current directed toward Jupiter on the surface

$$f = f_1 = f_0(\rho = \rho_1).$$

Thus, the plasma produced by Jupiter is separated: the ions form the plasma disk in the equatorial plane, and the electrons form the surface electric current J_1. This does not imply that the plasma disk is positively charged; it is neutral. The electrons from the electron sea of the Jovian magnetosphere neutralize any electric charge.

4. THE MAGNETOSPHERIC AND IONOSPHERIC CURRENTS

We assume the magnetospheric plasma to be ideal with an infinite conductivity. Under steady-state conditions, the magnetospheric plasma can rotate around the z axis with an angular velocity $\omega(\rho, z)$. In a coordinate system rotating with this angular velocity, the electric field $\mathbf{E}'$ is zero. Therefore, the electric field in the laboratory coordinate system is

$$\mathbf{E} = -[[\omega \times \mathbf{r}] \times \mathbf{B}], \quad E_\rho = -\omega \frac{\partial f}{\partial \rho},$$

$$E_z = -\omega \frac{\partial f}{\partial z}, \quad E_\phi = 0.$$

The electric field must be potential, $\mathbf{E} = -\nabla\Psi$, where Ψ is the electric potential. This implies that the angular frequency $\omega(\rho, z)$ is only a function of the magnetic flux f, $\omega = \omega(f)$, and the electric potential is

$$\Psi = \int^f \omega(f')df'.$$

To pass the electric current J_1 through the Jovian ionosphere, the ionospheric potential Ψ_i cannot be equal to the magnetospheric potential Ψ due to the finite conductivity of the Jovian ionosphere. Thus, the angular velocity of the magnetospheric plasma ω cannot be equal to the angular velocity of the ionosphere ω_i or the angular velocity of Jupiter ω_J. In the coordinate system comoving with the ionosphere, there is an electric field,

$$E_\theta = R_J \sin\theta(\omega_i - \omega)B_r,$$

$$E_r = -R_J \sin\theta(\omega_i - \omega)B_\theta.$$

Here, it is convenient to use the spherical coordinates on the surface of Jupiter, ($r = R_J$, θ, ϕ), The magnetic field in the ionosphere is weakly distorted by the electric currents and is dipolar,

$$B_r = 2B_J \cos\theta, \quad B_\theta = B_J \sin\theta.$$

The electric field E_θ produces a surface current in the ionosphere,

$$J_{i\theta} = \Sigma_P E_\theta = 2\Sigma_P B_J R_J \sin\theta\cos\theta(\omega_i - \omega),$$

where Σ_P is the Pedersen surface conductivity of Jupiter's ionosphere in its polar region. The surface current $\mathbf{J}_i$ flowing in the ionosphere is related to the bulk current in the magnetosphere $\mathbf{j}$ by the continuity equation

$$\mathrm{div}_s \mathbf{J}_i = -j_n.$$

The divergence in this equation is taken along the surface over which the current flows, and j_n is the component of the bulk current $\mathbf{j}$ normal to this surface. A similar equation relates the surface current flowing in the disk $\mathbf{J}_d$ to the magnetospheric current $\mathbf{j}$. However, as we emphasized above (see Eq. (5)), the divergence of the ion surface current is zero. At the same time, the disk electrons can only rotate with the angular velocity of the magnetosphere $\omega(f)$, and the divergence of their flux is also zero. Thus, the bulk electric current in the magnetosphere $\mathbf{j}$ is absent everywhere, except the preferential $f = f_1$ and $f = f_3$ magnetic surfaces where the plasma disk begins and ends, respectively. Recent energetic electron measurements from the Galileo spacecraft have revealed an electric current only near Io [5]. The total electric current I produced by Io is

$$I = \frac{Ze\dot{M}}{m_i} = 1.9 \times 10^{10} \frac{Zm_p}{0.1m_i}\left(\frac{\dot{M}}{2 \times 10^3\ (\mathrm{kg\ s^{-1}})}\right)\mathrm{A}. \quad (11)$$

Half of this current flows in the northern hemisphere, and the other half flows in the southern hemisphere. The magnetospheric currents flow only along the magnetic field on the two magnetic surfaces,

$$\mathbf{j}_1 = \mathbf{B}\frac{Ze\dot{M}}{4\pi m_i}\delta(f - f_1),$$

$$\mathbf{j}_3 = -\mathbf{B}\frac{Ze\dot{M}}{4\pi m_i}\delta(f - f_3). \quad (12)$$

These currents close the surface current flowing in the ionosphere

$$J_\theta = \frac{Ze\dot{M}}{4\pi R_J m_i}\frac{1}{\sin\theta}. \quad (13)$$

This imposes the following dependence of the difference between the angular frequencies of the ionosphere

and the magnetosphere on the magnetic flux f:

$$\omega_i - \omega = \frac{ZeM}{8\pi m_i \Sigma_P} f^{-1}\left(1 - \frac{f}{B_J R_J^2}\right)^{-1/2} \times \Theta(f - f_3)\Theta(f_1 - f).$$

Here, we used the fact that on the surface of Jupiter,

$$f = B_J R_J^2 \sin^2\theta.$$

The angular frequency of the ionosphere ω_i can differ from the angular frequency of Jupiter itself ω_J. This is attributable to the Ampere forces acting on the ionospheric plasma in which the electric currents flow. The difference $\omega_J - \omega_i$ is determined by the friction between the neutral atmosphere and the ionospheric plasma. It is commonly assumed that

$$\omega_J - \omega_i = k(\omega_J - \omega),$$

where the constant $k < 1$ [6]. Introducing the effective Pedersen conductivity

$$\Sigma_P^* = \Sigma_P(1 - k),$$

we obtain the following dependence of the angular velocity of the magnetosphere on the magnetic flux:

$$\omega(f) = \omega_J - \frac{ZeM}{8\pi m_i \Sigma_P^*} f^{-1}\left(1 - \frac{f}{B_J R_J^2}\right)^{-1/2}, \quad f_1 > f > f_3. \tag{14}$$

Equation (14) shows that the corotation of the magnetic field breaks down in the entire region occupied by the disk. This makes it possible to pass the electric current I (11) through the ionosphere.

The magnetospheric electric currents $\mathbf{j}_1$ and $\mathbf{j}_3$ in the region $f_1 > f > f_3$ generate the toroidal magnetic field

$$B_\phi = \frac{g}{\rho}, \tag{15}$$

where

$$g = -\mu_0 \frac{ZeM}{4\pi m_i}.$$

The ϕ component of Maxwell's equation

$$\mathrm{curl}\mathbf{B} = \mu_0 \mathbf{j}$$

then yields an equation for the poloidal magnetic flux f:

$$\rho \frac{\partial}{\partial \rho}\left(\frac{1}{\rho}\frac{\partial f}{\partial \rho}\right) + \frac{\partial^2 f}{\partial z^2} = \frac{g^2}{2}[\delta(f - f_1) - \delta(f - f_3)]. \tag{16}$$

The right-hand side of Eq. (16) is nonzero only on the $f = f_1$ and $f = f_3$ magnetic surfaces, where the magnetospheric electric currents flow along the magnetic field. Owing to the toroidal magnetic field, the field-aligned electric currents also have toroidal components that distort the poloidal magnetic field. Outside the f_1 and f_3 surfaces, Eq. (16) is identical to Eq. (2), and its solution is described by relations (3) and (4). However, the fields on both sides of the discontinuity are different. The relationship between the magnetic fields for $f > f_1$ and $f < f_1$ as well as for $f > f_3$ and $f < f_3$ can be derived by integrating Eq. (16) over f in a small neighborhood of the $f = f_1$ and $f = f_3$ surfaces, respectively. This requires passing to different independent variables (z, f) and assuming that the magnetic surfaces are defined by the relation $\rho = \rho(z, f)$. Equation (16) takes the form

$$\frac{1}{2}\left[1 + \left(\frac{\partial \rho}{\partial z}\right)^2\right]\frac{\partial}{\partial f}\left[\left(\frac{\partial \rho}{\partial f}\right)^{-2}\right] + \left(\frac{\partial \rho}{\partial f}\right)^{-2} \times \frac{\partial}{\partial f}\left[1 + \left(\frac{\partial \rho}{\partial z}\right)^2\right] - \left(\frac{\partial \rho}{\partial f}\right)^{-1}\left(\frac{1}{\rho} + \frac{\partial^2 \rho}{\partial z^2}\right) = \frac{g^2}{2}[\delta(f - f_1) - \delta(f - f_3)]. \tag{17}$$

We assume that the shape of the magnetic surfaces is the same on both sides of the discontinuity:

$$\left.\frac{\partial \rho}{\partial z}\right|_{f = f_{1,3} + 0} = \left.\frac{\partial \rho}{\partial z}\right|_{f = f_{1,3} - 0}.$$

Integrating Eq. (17) yields

$$B_\rho^2 + B_z^2 + B_\phi^2 = \text{const}. \tag{18}$$

This expression is the only condition for the equality of the magnetic pressures on both sides of the discontinuity, and it directly follows from our equations.

5. THE STRUCTURE OF THE MAGNETIC FIELD

Let us introduce the dimensionless (primed) variables that we use below by omitting the primes:

$$\omega' = \frac{\omega}{\omega_J}, \quad \mathbf{r}' = \frac{\mathbf{r}}{R_J}, \quad \mathbf{v}' = \frac{\mathbf{v}}{R_J \omega_J}, \quad \mathbf{B}' = \frac{\mathbf{B}}{B_J}, \quad f' = \frac{f}{B_J R_J^2}, \quad \Sigma' = \frac{2\pi m_i \omega_J R_J^2 \Sigma}{\dot{M}}, \quad \mathbf{J}' = \frac{4\pi R_J \mathbf{J}}{I}. \tag{19}$$

As a result, dimensionless parameters appear. One of these, mentioned in Section 3, is the ratio of the ion cyclotron frequency to Jupiter's angular frequency:

$$\Omega_c = \frac{\omega_c}{\omega_J} = 2.16 \times 10^7 \frac{Zm_p}{0.1 m_i} \frac{B_J}{4G}. \tag{20}$$

We also introduce Hill's parameter H, first used by him in 1979 [3]:

$$H = \frac{8\pi\Sigma_P^* B_J^2 R_J^2}{\dot{M}} = 10^7\left(\frac{B_J}{4G}\right)^2 \frac{\Sigma_P^*}{1\ \text{mho}}\left(\frac{\dot{M}}{2\times 10^3\ \text{kg s}^{-1}}\right)^{-1}. \quad (21)$$

The physical meaning of H will become clear below. We also define the Keplerian parameter, the ratio of the Keplerian angular velocity on Jupiter's surface to its angular velocity:

$$\Omega_K = \left(\frac{GM_J}{R_J^3\omega_J^2}\right)^{1/2} = 3.33. \quad (22)$$

Finally, we introduce another parameter, the ratio of the magnetic field strength produced by the current $I/2$ as if it flowed over Jupiter's surface to the proper magnetic field of Jupiter B_J:

$$\delta B = \frac{Ze\dot{M}\mu_0}{4\pi m_i R_J B_J} = 6.7\times 10^{-2}\frac{Zm_p}{0.1m_i} \times \frac{\dot{M}}{2\times 10^3\ \text{kg s}^{-1}}\left(\frac{B_J}{4\ \text{G}}\right)^{-1}. \quad (23)$$

Note that the magnetic field from the magnetospheric current $J/2$ in Io's orbit is by a factor of $\delta B(\rho_{Io}/R_J)^2 \approx 2.3$ stronger than the Jovian dipole magnetic field at this location. Thus, the dipole magnetic field at $\rho < \rho_1$ must be compressed in accordance with relation (18).

The rotation velocity of the ions in the plasma disk is defined by Eq. (10),

$$v_\phi = \frac{\Omega_K}{\rho}\rho_1^{1/2} + \frac{\Omega_c}{\rho}(f_1 - f_0). \quad (24)$$

We assume that the ion rotation velocity at $\rho = \rho_1 \approx \rho_{Io}$ is equal to the Keplerian velocity v_K, which is lower than the velocity of corotation with Jupiter. Indeed, the ions are formed from the neutral gas corotating with Io. Moving subsequently in the radial direction, the ions acquire an additional toroidal velocity, approaching corotation, $v_\phi = \rho$. Of course, the energy is drawn from the work of the radial electric field E_ρ. However, the velocity v_ϕ does not exceed the corotation velocity, because the disk ends earlier, $\rho < (\Omega_c f_1)^{1/2}$. At the same time, the disk electrons rotate with the angular velocity of the magnetic field lines $\omega(f)$ defined by Eq. (14):

$$\omega\rho = \rho - \frac{\Omega_c}{H}\rho f_0^{-1}(1 - f_0)^{-1/2}. \quad (25)$$

The velocity $\omega\rho$ cannot be negative, implying that the parameter H cannot be smaller than $\Omega_c/f_1 > 10\Omega_c$. It thus follows that the Pedersen conductivity of the ionosphere cannot be low. At low conductivity Σ_P, the ionosphere cannot pass the electric current $I/2$, because Jupiter as a unipolar inductor [7] can produce only a finite voltage,

$$\Psi_{max} = \omega_J f_{eq} = B_J R_J^2 \omega_J.$$

Actually, this voltage is even lower, since the electric coupling between the disk and Jupiter passes through the polar $f = f_1$ magnetic field line, $\Psi = \Psi_{max} f_1$. The magnetospheric currents–ionosphere system is self-consistent: when the current I becomes large and the ionosphere cannot pass it, a strong electron-accelerating electric field emerges in the magnetosphere along the $f = f_1$ magnetic field lines. The fast electrons precipitating into the ionosphere produce additional ionization, causing the ionospheric conductivity to increase [8].

Thus, the toroidal electric current

$$J_\phi = 2\frac{\delta B}{\mu_0}\Sigma(v_\phi - \omega\rho) \quad (26)$$

flows in the disk; it leads to a discontinuity of the radial magnetic field on the $z = 0$ disk surface:

$$B_\rho|_{z=z+0} = -B_\rho|_{z=z-0} = \delta B\Sigma(v_\phi - \omega\rho). \quad (27)$$

The quantity $v_\phi - \omega\rho$ is a function of ρ and $f_0(\rho)$ and is defined by relations (24) and (25):

$$v_\phi - \omega\rho = -\frac{\rho^2 - \Omega_K\rho_1^{1/2}}{\rho} + \frac{\Omega_c}{\rho}(f_1 - f_0) + \frac{\Omega_c}{H}\rho f_0^{-1}(1 - f_0)^{-1/2}. \quad (28)$$

We see from Eqs. (27) and (28) that the radial magnetic field at $\rho = \rho_1$ is negative:

$$B_\rho(\rho = \rho_1, z = +0) < 0.$$

Here, we do not consider the case of a low ionospheric conductivity close to the limiting value when the parameter

$$H < \Omega_c f_1^{-1}(1 - f_1)^{-1/2}/(1 - \Omega_K/\rho_1^{3/2}) = 1.3\Omega_c f_1^{-1}(1 - f_1)^{-1/2}.$$

Below, we show that this is the instability region. For large H, the magnetic field lines are curved inward, toward Jupiter. This results from the expulsion of the Jovian magnetic field by the disk. At large distances $\rho > \rho_1$, $B_\rho(\rho, z = +0)$ changes its sign and becomes positive. This is the region where the Jovian magnetic field is expelled outward. We see that the Jovian magnetic

field is expelled in both directions and not only outward, as assumed previously [6]. It is important to know the position of the point $\rho = \rho_2$ where the radial magnetic field in the equatorial plane vanishes. The $f = f_2 = f_0(\rho_2)$ magnetic field line is nearly dipolar. All of the $f_1 < f < 1/\rho_1$ magnetic field lines are expelled from the region $\rho_1 < \rho < \rho_2$ and are pressed into the region $\rho < \rho_1$. The radial magnetic field $B_\rho(\rho, z = +0)$ is determined by the derivative of the magnetic flux with respect to the z coordinate on the equator

$$\left.\frac{\partial f}{\partial z}\right|_{z=+0} = -\delta B \Sigma \rho (v_\phi - \omega\rho). \tag{29}$$

Let us define this derivative as a function of $f_0(\rho)$. From Eqs. (3) and (4), we derive

$$\left.\frac{\partial f}{\partial z}\right|_{z=+0} = -\rho\int_0^\infty d\rho' f_0(\rho')\int_0^\infty \lambda^2 J_1(\lambda\rho)J_1(\lambda\rho')d\lambda.$$

The second integral in this equation is known [9], and we obtain

$$\left.\frac{\partial f}{\partial z}\right|_{z=+0} = \frac{2}{\pi}\rho\frac{\partial}{\partial\rho}\frac{1}{\rho}\left\{\int_0^1 \frac{dxK(x)}{dx} f_0(\rho/x)dx - \int_0^1 \frac{dK(x)}{dx} f_0(\rho x)dx\right\}. \tag{30}$$

Here, $K(x)$ is the complete elliptic integral of the first kind. It is equal to $\pi/2$ at $x = 0$, but has a logarithmic singularity for $x \longrightarrow 1$:

$$K(x) \longrightarrow \ln[16/(1-x^2)]/2.$$

Equations (28)–(30) close the system of equations that define the function $f_0(\rho)$. Equations (6) and (8) that define the disk surface density $\Sigma(\rho)$ should be added to these:

$$\rho v_\rho \Sigma = 1,$$
$$\frac{1}{2}\rho\frac{\partial v_\rho^2}{\partial\rho} = v_\phi^2 - \frac{\Omega_K^2}{\rho} + \Omega_c(v_\phi - \omega\rho)\frac{\partial f_0}{\partial\rho}. \tag{31}$$

The integral equation (30) can be simplified at $\rho \approx \rho_2 > \rho_1$. The magnetic flux $f_0(\rho)$ changes here only slightly. The function $dK(x)/dx$ increases rapidly near $x \approx 1$. In the region $0 < x < \rho_1/\rho$, the magnetic flux in the second integral of Eq. (30) may be assumed to be dipolar,

$$f_0(\rho x) = 1/\rho x.$$

As a result, we obtain the equation

$$\left.\frac{\partial f}{\partial z}\right|_{z=+0} = \frac{2}{\pi}\rho\frac{\partial}{\partial\rho}\frac{1}{\rho} \times \left\{ f_0(\rho)K\left(\frac{\rho_1}{\rho}\right) - \frac{1}{\rho_1}\left[K\left(\frac{\rho_1}{\rho}\right) - E\left(\frac{\rho_1}{\rho}\right)\right]\right\},$$
$$\rho \approx \rho_2,$$

where $E(x)$ is the complete elliptic integral of the second kind. Since $f_2 = f_0(\rho_2)$ is close to f_1, as we will see below, the equation for ρ_2 is

$$f_1\rho_1\frac{\partial}{\partial\rho}\frac{1}{\rho}K\left(\frac{\rho_1}{\rho}\right)\Big|_{\rho=\rho_2} = \frac{\partial}{\partial\rho}\frac{1}{\rho}\left[K\left(\frac{\rho_1}{\rho}\right) - E\left(\frac{\rho_1}{\rho}\right)\right]\Big|_{\rho=\rho_2}.$$

An approximate solution of this equation yields

$$\rho_2 = \rho_1(2f_1\rho_1/3)^{-1/2}.$$

However, $f \approx 1/\rho_2$, and we obtain

$$\rho_2 \approx 3\rho_1/2 \approx 8.9, \quad f_1 = 0.113.$$

Note that for a dipole,

$$f_{1d} = 0.17 > f_1.$$

The projection of the point ρ_1 along the magnetic field onto the Jovian ionosphere is a circumference in the polar region

$$\theta_1 = \arcsin(f_1^{1/2}) = 19.5^\circ.$$

At the same time, the projection of Io is

$$\theta_{Io} \approx \arcsin(\rho_1^{-1/2}) = 24.5^\circ.$$

This 5° gap between Io's footprint and the auroral oval on Jupiter is actually observed [10].

Let us now check that f_2 is very close to f_1. Indeed, the left-hand side of Eq. (28) must be equal to zero at $\rho = \rho_2$ and $f_0 = f_2$. This determines f_2:

$$f_2 = f_1 - \frac{\rho_2^2 - \Omega_K\rho_1^{1/2}}{\Omega_c} + \frac{\rho_2^2 f_1^{-1}(1-f_1)^{-1/2}}{H} < f_1.$$

The difference between the magnetic fluxes at ρ_1 and ρ_2 is very small,

$$f_1 - f_2 \approx \rho_2^2/\Omega_c \approx 10^{-5}.$$

This implies that the magnetic field lines above the plasma disk are nearly parallel to its surface, except for the region $\rho \approx \rho_2$. At the point

$$\rho = \rho_1 + 0,$$

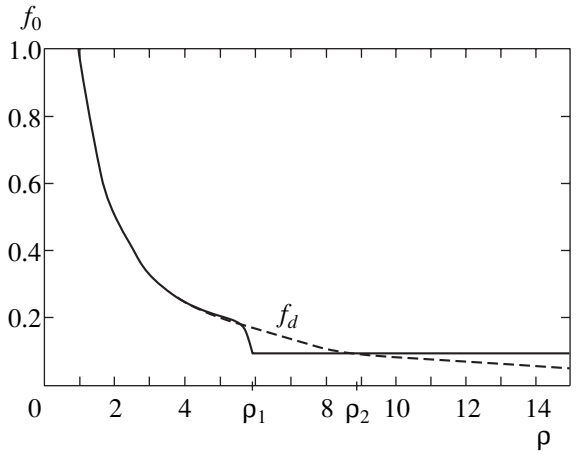

Fig. 1. Poloidal magnetic flux f_0 in the equatorial $z = 0$ plane versus radial distance ρ. The point $\rho = \rho_1$ corresponds to the inner edge of the plasma disk. The point $\rho = \rho_2$ corresponds to the "middle" point where the radial magnetic field is zero. The dashed line indicates the dipole magnetic flux $f_d = 1/\rho$.

the vertical magnetic field $B_z \approx \Omega_c^{-1}$ is weak, and it may be assumed to be zero. However, the radial magnetic field is not weak,

$$B_{\rho_1} = -\delta B \Sigma_1 \rho_1 \left[1 - \frac{\Omega_K}{\rho_1^{3/2}} - \frac{\Omega_c}{H} f_1^{-1}(1-f_1)^{-1/2}\right], \quad \Sigma_1 = \Sigma(\rho_1). \tag{32}$$

The toroidal magnetic field is also strong (see Eqs. (15) and (23)):

$$B_{\phi_1} = -\frac{\delta B}{\rho_1}.$$

The magnetic field strength is

$$B_1 = (B_{\rho_1}^2 + B_{\phi_1}^2)^{1/2} = \frac{\delta B}{\rho_1} \times \left\{1 + \Sigma_1^2 \rho_1^4 \left[1 - \frac{\Omega_K}{\rho_1^{3/2}} - \frac{\Omega_c}{H} f_1^{-1}(1-f_1)^{-1/2}\right]^2\right\}^{1/2} > \frac{\delta B}{\rho_1}. \tag{33}$$

To the left of the discontinuity, $\rho = \rho_1 - 0$, the radial magnetic field in the equatorial plane B_ρ is zero, because the magnetic field lines are symmetric; the toroidal field is also zero, $B_\phi = 0$. Then, according to the boundary condition (18),

$$B_z(\rho = \rho_1 - 0) = -B_1. \tag{34}$$

This vertical magnetic field is stronger than the dipole magnetic field at $\rho = \rho_1$, $B_{zd} = 1/\rho_1^3$, as mentioned above. For this reason, the magnetic field lines must be compressed at $\rho < \rho_1$. Thus, either the radial magnetic field at $\rho > \rho_1$ must be negative or

$$v_\phi - \omega\rho < 0.$$

At low ionospheric conductivity,

$$\Omega_c f_1^{-1}(1-f_1)^{-1/2} < H < 1.3\Omega_c f_1^{-1}(1-f_1)^{-1/2}, \tag{35}$$

the condition for the equality of the magnetic pressures on both sides of the discontinuity cannot be satisfied, and the $f = f_1$ boundary becomes unstable.

At

$$H > 1.3\Omega_c f_1^{-1}(1-f_1)^{-1/2},$$

the boundary condition (18) determines the derivative of the magnetic flux f_0 with respect to the radius ρ at $\rho = \rho_1 - 0$:

$$\left.\frac{\partial f_0}{\partial \rho}\right|_{\rho = \rho_1 - 0} = -\delta B \left\{1 + \Sigma_1^2 \rho_1^4 \left[1 - \frac{\Omega_K}{\rho_1^{3/2}} - \frac{\Omega_c}{H} f_1^{-1}(1-f_1)^{-1/2}\right]^2\right\}^{1/2}. \tag{36}$$

Under this condition and for $f_0(\rho) \longrightarrow 1/\rho$ as $\rho \longrightarrow 0$, Eqs. (28)–(31) uniquely define the function $f_0(\rho)$, which is shown in Fig. 1. The flux f_0 decreases from $1/\rho_1$ to f_1 near $\rho \approx \rho_1$. We can estimate the width of the region of the compressed magnetic field,

$$\Delta\rho = \left.\frac{1/\rho_1 - f_1}{\partial f_0/\partial\rho}\right|_{\rho = \rho_1 - 0} < \frac{1/\rho_1 - f_1}{\delta B} \approx 0.85.$$

It is rather small,

$$\Delta\rho/\rho_1 \approx 0.1.$$

To analyze the behavior of the magnetic flux on the equator $f_0(\rho)$ at a fairly large distance $\rho > \rho_2$, let us first integrate the equation of motion for ions (31). Taking into account expressions (24) and (28), we obtain

$$\mathscr{E} - \mathscr{E}_1 = \Omega_c(f_1 - f_0) - \frac{\Omega_c^2}{H}\ln\frac{f_1(1+\sqrt{1-f_0})^2}{f_0(1+\sqrt{1-f_1})^2}. \tag{37}$$

This is the energy conservation law:

$$\mathscr{E} = \frac{1}{2}v_\rho^2 + \frac{1}{2}v_\phi^2 - \frac{\Omega_K^2}{\rho}, \quad \mathscr{E}_1 = \mathscr{E}(\rho_1).$$

The right-hand side of Eq. (37) is the source of energy for the disk ions. This is the work of the radial electric field on the ions. The electric field is produced by the rotating magnetic field through Jupiter's rotation. All of

the energy is draw from Jupiter's rotation. The second term on the right-hand side of this equation represents the ohmic losses. Jupiter spends part of its rotational energy on heating the ionosphere. The total energy expenditure must be larger than the ohmic losses. Expanding the last term in Eq. (37) near the point $f_0 = f_1$, we again obtain the condition

$$H > \Omega_c f_1^{-1}(1 - f_1)^{-1/2}.$$

It is identical to the condition $\omega > 0$. Jupiter must spend part of its rotational energy on spinning up the plasma disk.

At $\rho > \rho_2$, the Keplerian rotation is insignificant, and we can write the expressions for the velocities of the disk matter as

$$v_\phi = \frac{\Omega_c}{\rho}(f_1 - f_0),$$
$$v_\rho = \left[2\Omega_c\left(1 - \frac{\Omega_c}{Hf_1}\right)(f_1 - f_0) - \frac{\Omega_c^2}{\rho^2}(f_1 - f_0)^2\right]^{1/2}. \quad (38)$$

As we show below, here we take into account the fact that the change in magnetic flux is small,

$$f_1 - f_0 \ll f_1.$$

In addition, the values of f_0 and f_1 themselves are low,

$$f_0 < f_1 \approx 0.1 \ll 1.$$

The radial velocities in the disk are always positive,

$$\Omega_c(f_1 - f_0) < 2\rho^2(1 - \Omega_c/Hf_1).$$

We now also know the disk surface density, $\Sigma = 1/\rho v_\rho$, and can solve Eqs. (29) and (30) to determine the function $f_0(\rho)$. The final equation is

$$f_0 = f_1 - \frac{\rho^2}{\Omega_c}\left(1 - \frac{\Omega_c}{Hf_1}\right)\left[1 + \frac{\rho(2Hf_1^2 + \rho^2)^{1/2}}{Hf_1^2 + \rho^2}\right], \quad (39)$$
$$\rho > \rho_2.$$

The function in square brackets is a weakly changing function of ρ; it changes from 1 at small distances to 2 at large distances. Therefore, the magnetic flux may be assumed to decrease with increasing ρ,

$$f_0 = f_1 - \text{const} \cdot \rho^2.$$

This implies that the vertical magnetic field above the disk is constant,

$$B_z(z = +0) = -2(1 - \Omega_c/Hf_1)/\Omega_c.$$

It is generated by the toroidal electric currents flowing in the disk. The radial magnetic field $B_\rho(z = +0)$ increases with distance,

$$B_\rho = \rho(\delta B/Hf_1^2),$$

because the magnetic field lines are drawn by the increasing radial plasma motion in the disk. As a result, the inclination of the poloidal magnetic field to the disk

$$\tan\alpha = -\frac{dz}{d\rho} = -\frac{B_z}{B_\rho} = \frac{2f_1}{\delta B}\left(\frac{Hf_1}{\Omega_c} - 1\right)\rho^{-1}$$

decreases with distance.

The parameters of the disk are

$$v_\phi = \left(1 - \frac{\Omega_c}{Hf_1}\right)\rho,$$
$$v_\rho = \left(1 - \frac{\Omega_c}{Hf_1}\right)\rho, \quad (40)$$
$$\Sigma = (v_\rho\rho)^{-1} = \left(1 - \frac{\Omega_c}{Hf_1}\right)^{-1}\rho^{-2}.$$

We see that the disk at $\rho > \rho_2$ rotates with an almost constant angular velocity

$$\Omega_d = 1 - \Omega_c/Hf_1 < 1,$$

which is lower than the angular velocity of Jupiter itself. This rotation was measured from Voyager 1 [11]. The radial velocity of the disk matter also increases with distance and tends to its toroidal velocity. The angular velocity of the magnetic field lines $\omega(\rho)$ (25) initially also follows this law, but subsequently begins to decrease:

$$\omega\rho = \left(1 - \frac{\Omega_c}{Hf_1}\right)\left(1 - \frac{\rho^2}{Hf_1^2}\right)\rho. \quad (41)$$

At $\rho \approx f_1H^{1/2}$, the rotation of the magnetosphere stops. This is the end of the plasma disk that produces the global currents. The more accurate expression derived from Eq. (39) is

$$\rho_3 = 0.73 f_1 H^{1/2}.$$

The corresponding expression for the magnetic flux is

$$f_3 = f_1(\Omega_c/Hf_1) < f_1.$$

At $f < f_3$, the Jovian ionosphere cannot pass the current $I/2$, because the surface current increases as $\sin^{-1}\theta$, while the voltage decreases as $\sin\theta$. Subsequently, the electric current $-I/2$ returns to the disk along the $f = f_3$ magnetic surface.

The plasma disk does not disappear after closure of the electric current. It only ceases to interact with the Jovian magnetic field. The magnetic field above the disk becomes parallel to its surface, $B_z = 0$. The disk rotation slows down due to the conservation of angular momentum,

$$v_\phi = v_{\phi_3}\frac{\rho_3}{\rho}, \quad \rho > \rho_3.$$

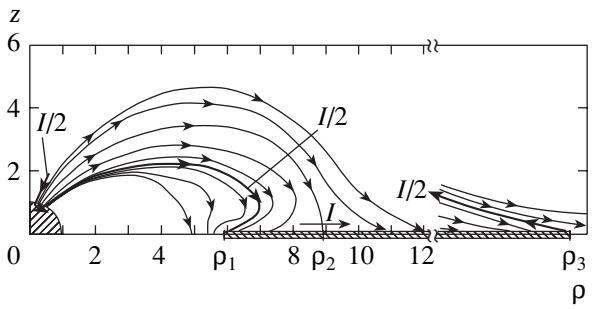

Fig. 2. Poloidal magnetic field lines. The ρ_1 and ρ_3 coordinates are the edges of the plasma disk. The heavy lines indicate the electric currents flowing in the Jovian ionosphere and magnetosphere ($I/2$) and in the plasma disk (I). The angles of inclination of the magnetic field lines to the disk are shown larger than in reality to understand the topology of the magnetic field as a whole

The radial velocity in the disk increases because of the centrifugal acceleration,

$$v_\rho = \left[v_{\rho_3}^2 + v_{\phi_3}^2\left(1 - \frac{\rho_3^2}{\rho^2}\right) - \frac{2\Omega_K^2}{\rho_3}\left(1 - \frac{\rho_3}{\rho}\right)\right]^{1/2},$$

and tends to its limiting value approximately equal to

$$(v_{\rho_3}^2 + v_{\phi_3}^2)^{1/2}.$$

The surface density Σ decreases proportionally to ρ^{-1}, and the disk disappears in interplanetary space.

The boundary condition (18) on both sides of the surface electric current flowing over the $f = f_3$ magnetic surface may be considered as the equality of the total pressures $B^2/2\mu_0 + P$ if we take into account the pressure P from the fast electrons of the return current. The magnetic pressure on the left side of the discontinuity is $(B_\rho^2 + B_\phi^2)/2\mu_0$. The toroidal magnetic field B_ϕ is given by relation (15),

$$B_\phi = -\delta B/\rho.$$

The radial magnetic field at $\rho = \rho_3$ is exactly the same value, because $\omega = 0$ and $v_\rho = v_\phi$ there (see Eq. (27)). Thus, we can satisfy the stability condition for the $f = f_3$ boundary if the electron pressure is

$$P = B_\phi^2/2\mu_0 = B_\rho^2/2\mu_0.$$

The condition

$$P = B_\rho^2/2\mu_0$$

corresponds to equipartition and seems natural enough.

The structures of the magnetic field and the electric current are shown in Fig. 2.

6. DISCUSSION

We have shown that the structure of the Jovian magnetic field can be accurately determined analytically. We used ideal approximations: an axisymmetric, highly conductive magnetosphere with a cold thin disk. These approximations are realistic. We showed that the disk expels the Jovian magnetic field both outward and inward, toward Jupiter. As a result, the magnetic field is compressed in the region $\rho < \rho_1$, where ρ_1 is the inner edge of the disk, and weakens in the remaining region occupied by the disk. Due to the field compression and the curvature of magnetic field lines, the magnetic image of the inner edge of the disk on Jupiter's surface is observed at higher latitudes than it would for a dipole magnetic field ($f_1 < 1/\rho_1$, $f_1 \approx 0.11$). This corresponds to a latitude of 70.5°, where an aurora is observed. It turned out that the electric currents in the magnetosphere could flow only along the preferential $f = f_1$ and $f = f_3$ magnetic surfaces that correspond to the inner and outer edges of the disk, respectively. No bulk currents flow in the magnetosphere. To maintain the electric current $I/2$ (11), the Jovian ionosphere must have a high Pedersen surface conductivity Σ_P. This is because the Jovian rotating magnetic field can produce only a finite electric voltage while interacting with the disk,

$$\Psi = B_J R_J^2 \omega_J f_1.$$

On the other hand, the voltage

$$U = \frac{I}{8\pi\Sigma_P^*(1 - f_1)^{1/2}}$$

must be applied to the conductive ionosphere. The condition $\Psi > U$ must be satisfied. Here, it is convenient to introduce a parameter that describes the ability of the Jovian ionosphere to pass an electric current:

$$\kappa = \frac{\Psi - U}{\Psi} = 1 - \frac{\Omega_c}{Hf_1(1 - f_1)^{1/2}} \approx 1 - \frac{\Omega_c}{Hf_1}. \quad (42)$$

As can be seen, many physical parameters of the magnetosphere and the disk contain κ (see, e.g., Eqs. (40) and (41)). Note also that κ is the part of the total energy lost by Jupiter and transferred to the disk:

$$W_d = W\kappa.$$

Accordingly, the ohmic losses in the ionosphere are

$$W_{Ohm} = W(1 - \kappa).$$

This follows from the energy conservation law (37):

$$W = IB_J R_J^2 \omega_J (f_1 - f_3)$$

$$= 6.9 \times 10^{18}(f_1 - f_3)\frac{Zm_p}{0.1m_i}\left(\frac{\dot{M}}{2 \times 10^3\ \text{kg s}^{-1}}\right)\left(\frac{B_J}{4\ \text{G}}\right)\text{W}.$$

The parameter κ is more suitable for this problem than Hill's parameter H, which has no clear physical mean-

ing. The quantity H could be treated as a dimensionless value of the Pedersen surface conductivity Σ_P, but it would be more natural to measure Σ_P in units of the ratio of the characteristic surface current

$$J = I/4\pi R_J$$

to the characteristic electric field

$$E = \Psi_{\max}/R_J.$$

The dimensionless value of Σ_P would then be

$$\Sigma_P' = H/2\Omega_c.$$

The value of H/Ω_c is on the order of unity and is not so large as H.

The parameter κ is positive. Moreover, the stability condition for the $f = f_1$ boundary dictates the inequality $\kappa > \kappa_{\min} = 0.23$. This imposes a constraint on the Pedersen surface conductivity of the Jovian ionosphere, $\Sigma_P \approx$ 10–20 mho, but the Jovian cold polar ionosphere possesses no such high conductivity. However, as calculations show [8], the ionization of the ionosphere by fast precipitating electrons increases its conductivity to these values. This implies that, actually, an electric current in the ionosphere exists only in the region where the magnetospheric current J_1 invades the ionosphere. Δf is the region occupied by the magnetospheric current. It is determined by the ionization of the gas emitted by Io. The ionospheric current flows only between the two $f = f_1$ and $f = f_3 = f_1 - \Delta f$ magnetic surfaces. The return magnetospheric current will flow along the $f = f_3$ magnetic surface. The forward and return electron currents flow almost together in the Jovian ionosphere, but then diverge significantly in the magnetosphere. This implies that the outer edge of the disk will be not at a distance $\rho_3 \approx f_1 H^{1/2}$, as we obtained previously for a constant ionospheric conductivity, but much closer:

$$\rho_3 = (\Omega_c \Delta f)^{1/2}\left(1 - \frac{\Omega_c}{H f_1}\right)^{-1/2} = \left(\frac{\Omega_c \Delta f}{\kappa}\right)^{1/2}.$$

This expression follows from Eq. (39). If Δf is $\sim 10^{-2}$, then ρ_3 is $\sim 10^2$. The projection of the entire disk along the magnetic field onto Jupiter's surface will be only 1° in this case.

In conclusion, note that the poloidal magnetic field near the disk is similar in structure to the field near a reconnecting current sheet [12]. The radial magnetic fields above and under the disk are directed oppositely. This is because the reconnection of the radial field took place in the disk; the toroidal electric current generated by the Jovian rotating magnetic field flows in it. The current loop that passes over the disk and through the ionosphere and that is closed by the magnetospheric currents emerged in order to transfer the angular momentum of Jupiter to the disk. Since the disk matter is continuously born out of the volcanic gases from Io, it must be spun up to be moved far from Jupiter. It is the Ampere force generated by the electric current flowing in the Jovian ionosphere perpendicular to the magnetic field that produces a braking torque acting on Jupiter. Therefore, the current system found that connects Jupiter and the disk is a mandatory attribute of the interaction between the plasma disk and the Jovian rotating magnetic field: the current cannot break anywhere or be reclosed.

ACKNOWLEDGMENTS

I thank S. Miller and A.D. Aylward for helpful discussions. This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16762) and Presidential Grant no. NSh-1603.2003.2.

REFERENCES

1. E. S. Belen'kaya, Itogi Nauki Tekh., Ser: Issled. Kosm. Prostranstva **33a**, 178 (2002).
2. F. K. Lamb, in *Timing Neutron Stars*, Ed. by H. Ogelman and E. P. J. van der Heuvel (Kluwer, Dordrecht, 1989).
3. T. W. Hill, J. Geophys. Res. **84**, 6554 (1979).
4. D. H. Pontius, Jr., J. Geophys. Res. **102** (A4), 7137 (1997).
5. B. H. Mauk, D. J. Williams, and A. Eviatar, J. Geophys. Res. **106** (A11), 26195 (2001).
6. S. W. H. Cowley and E. J. Bunce, Planet. Space Sci. **49**, 1067 (2001).
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1984).
8. G. Millward, S. Miller, T. Stallard, *et al.*, Icarus **160**, 95 (2002).
9. F. Oberhettinger, *Tables of Bessel Transforms* (Springer, Berlin, 1972).
10. J. T. Clarke, G. E. Ballester, J. T. Trauger, *et al.*, J. Geophys. Res. **103** (E9), 20217 (1998).
11. J. W. Belcher, in *Physics of the Jovian Magnetosphere*, Ed. by A. J. Dessler (Cambridge Univ. Press, Cambridge, 1983).
12. B. V. Somov and S. I. Syrovatskiĭ, Tr. Fiz. Inst. im. P. N. Lebedeva, Akad. Nauk SSSR **74**, 14 (1974).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 550–558.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 623–632.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Hypersound Velocity and Damping in the Critical Region of a Solution

S. V. Krivokhizha[a], I. A. Chaban[b], and L. L. Chaikov[a,*]

[a] *Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*
[b] *Andreev Institute of Acoustics, Russian Academy of Sciences, ul. Shvernika 4, Moscow, 117036 Russia*
**e-mail: chaik@sci.lebedev.ru*

Received July 28, 2004

Abstract—The results of investigation of hypersound damping in critical solutions with a closed phase-separation region at temperatures below the lower and above the upper critical points are presented. Available theories of sound propagation during phase transitions fail to provide a description of experimental results in the hypersound frequency range, in which the acoustic wavelength is comparable to fluctuations of the concentration. A new theoretical description (stemming from experimental data) proposed for hypersound damping in the vicinity of the critical points is based on the scattering of a sound wave from developed fluctuations of the concentration. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Approximately a century after the first observations and attempts at explaining critical opalescence, Smoluchowsky [1] found a correct physical interpretation of this remarkable phenomenon. He noted that the mean intensity of fluctuations of certain parameters increases in the vicinity of the critical point and, hence, the intensity of light scattered from these fluctuations also increases. Einstein [2] indicated the method for calculating the intensity of such fluctuations as well as the intensity of light scattered from these fluctuations.

Experimental and theoretical investigations of processes in the critical region is of special interest the more so because the theory of the condensed state has been developed insufficiently and the phenomena occurring in the region of the critical point and in the vicinity of second-order phase transitions in various media proceed analogously; in other words, critical phenomena exhibit isomorphism.

This important property of critical phenomena makes it possible to select a convenient object of investigation and to extend the results to other objects that have not been studied as yet. With such an approach to analyzing the problem as a whole, it is necessary to appropriately choose the order parameter, i.e., the parameter varying during a phase transition and strongly fluctuating in the critical region [3, 4]. Such a parameter for a binary solution is its concentration.

The study of propagation of ultrasound and hypersound provides important information on the dynamics of processes occurring near the critical points. Information on the velocity and damping of hypersound is usually obtained from the Mandelstam–Brillouin scattering spectra. These spectra are formed as a result of light scattering from thermal fluctuations of pressure.

In actual practice, thermal hypersound (i.e., a hypersonic Debye wave) does not attenuate on average [5, 6]. Density fluctuation or its spatial Fourier component (Debye wave) emerging at a certain instant gradually decays, giving rise to a wave with a different initial phase. Scattering of light from such waves is precisely the Mandelstam–Brillouin scattering. The intensity spectrum of this scattering is determined by the correlation function for Debye waves [7, 8]. The fluctuation-dissipative theorem states that this correlation function decreases exponentially with increasing time interval. The exponent (half-width of the Mandelstam–Brillouin lines) is equal to the product of velocity of sound V and damping coefficient α of an artificially generated acoustic wave of the same frequency [5, 7, 8]. The width of the Mandelstam–Brillouin lines had been analyzed in detail by Leontovich [5] even before the fluctuation-dissipative theorem was proved. Leontovich showed that, once appearing, Fourier component $\Delta\rho(t)$ of density fluctuations (which is characterized by wave vector $\mathbf{q}$, frequency ω, and initial phase ϕ) varies in time in accordance with the law [5]

$$\Delta\rho(t) = A\exp(-\alpha Vt)\exp[i(\omega t - \mathbf{q}\cdot\mathbf{r} + \phi)],$$

where A is the mean square amplitude of fluctuations [5]. The calculation of correlator $\langle\delta\rho(t)\delta\rho(t+\tau)\rangle$ and the spectrum of light scattered from such a wave [5] shows that the Mandelstam–Brillouin linewidth is determined by damping coefficient α. The term "hypersound damping coefficient" will be henceforth used precisely in this meaning.

Here, we report on the results of experimental investigation of the temperature dependence of hypersound velocity in critical solutions of guaiacol–glycerol with a phase-separation region (i.e., with two critical phase-separation points) and the temperature dependence of the hypersound damping coefficient, which was determined from the Mandelstam–Brillouin scattering linewidth. A new theory describing the behavior of the hypersound damping coefficient near the critical points is developed and the results of approximation of the experimental data using the formulas of the new theory are considered.

2. EXPERIMENTAL RESULTS

We chose as the object of investigation a guaiacol–glycerol solution with a closed phase-separation region and two critical points. This solution is convenient due to the fact that, first, its critical temperatures are close to room temperature and, second, it has a broad critical region of several degrees, while the entire critical region for single crystals and solutions with a single critical point spans only over fractions of a degree [4]. The coexistence of two critical points of phase separation (upper T_U and lower T_L) for this solution leads to an abnormally fast increase in correlation radius r_c of concentration fluctuations and in characteristic time τ of their dispersal. For example, a value of $r_c = 1000$ Å near the liquid–vapor critical point is attained at a temperature exceeding the critical temperature by approximately 0.013°C, while the same value of the correlation radius for the guaiacol–glycerol solution is attained for $|T - T_L| \approx 1.34$°C or $|T - T_U| \approx 1.34$°C. This makes it possible to approach the critical state in such a solution very closely (i.e., experiments can be carry out in a medium with a larger correlation radius of the order parameter).

A guaiacol ($CH_3OC_6H_4OH$)–glycerol ($C_3H_8O_3$) solution with high-purity components is homogeneous in the entire T–c phase plane (c is the concentration). If, however, we add a small amount of a third substance to such a solution (a water molecule for 27 molecules of the solution or a molecule of CCl_4 for 170 molecules of the solution), a closed region (loop) appears on the phase plane of the solution; the solution is heterogeneous (separated in two phases) inside the loop and homogeneous outside this loop. The size of the closed phase-separation region depends on the amount of the third substance added to the solution.

Figure 1 schematically shows the 3D phase diagram of such a solution [9, 10] in the following coordinates: temperature T, concentration c, and concentration C_x of the third substance. The diagram has the shape of an inverted dome. The solution is separated into phases inside the dome and is homogeneous outside it. The cross sections of the dome by the T–c planes form phase-separation regions with two critical points (upper and lower). These points form the lines of critical points

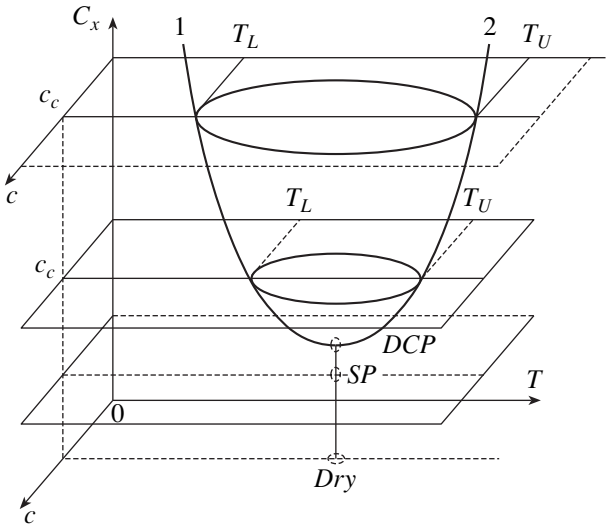

Fig. 1. Schematic 3D phase diagram of a solution with a closed phase-separation region: c is the concentration, T is the temperature, C_x is the concentration of the admixture, c_c is the critical concentration, T_L and T_U are the lower and upper critical temperatures, *Dry* is the plane corresponding to the solution without an admixture, *DCP* is the double critical point, and *SP* is the peculiar point.

on the surface of the dome. The lines of the upper and lower critical points merge at the vertex of the dome, forming a double critical point.

At concentrations C_x of the third substance for which phase-separation region does not emerge in the solution any longer, each T–c plane contains a point near which we observe an increase in the scattered light intensity and in the correlation radius of concentration fluctuations. Such a point on the T–c plane, which lies on the vertical that drops from the double critical point onto this plane and is closest to the double critical point in the 3D phase space, is referred to as a peculiar point. Peculiar points form a line of peculiar points between the double critical point and the peculiar point of the T–c plane corresponding to $C_x = 0$.

We analyzed the temperature dependences of the velocity and damping coefficient of hypersound above the upper critical point and below the lower critical point in a solution with a phase-separation region as well as in solutions with a double critical point and with peculiar points [11–14].

The velocity of hypersound was determined from the displacement of the Mandelstam–Brillouin components in the spectrum of molecular scattering of light, while the damping coefficient was determined from the widths of these components. The spectra were studied on a setup with a multipass piezoscanned Fabry–Perot interferometer manufactured at Burleigh [11].

The temperature dependences of the velocity and damping coefficient of hypersound [11–14] as well as

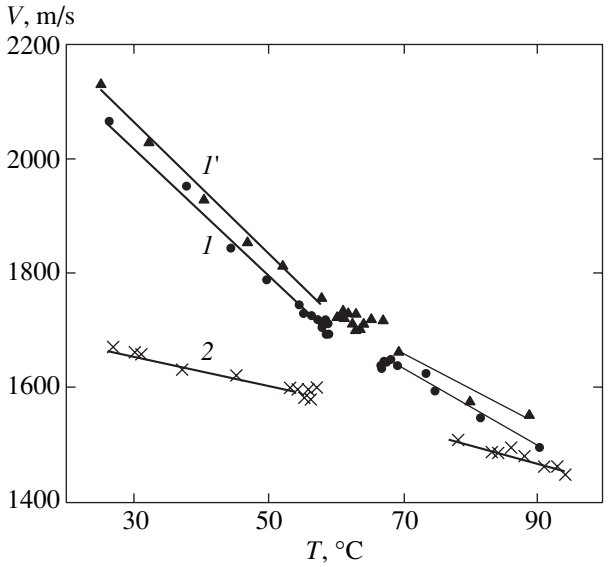

Fig. 2. Temperature dependences of the velocity of hypersound ($\Delta T = 7.28$°C (*1*) and $\Delta T = 0.062$°C (*1'*) [11, 13]) and ultrasound with frequency 2.8 MHz ($\Delta T = 7.2$°C (*2*) [15]) in guaiacol–glycerol solutions with different phase-separation regions.

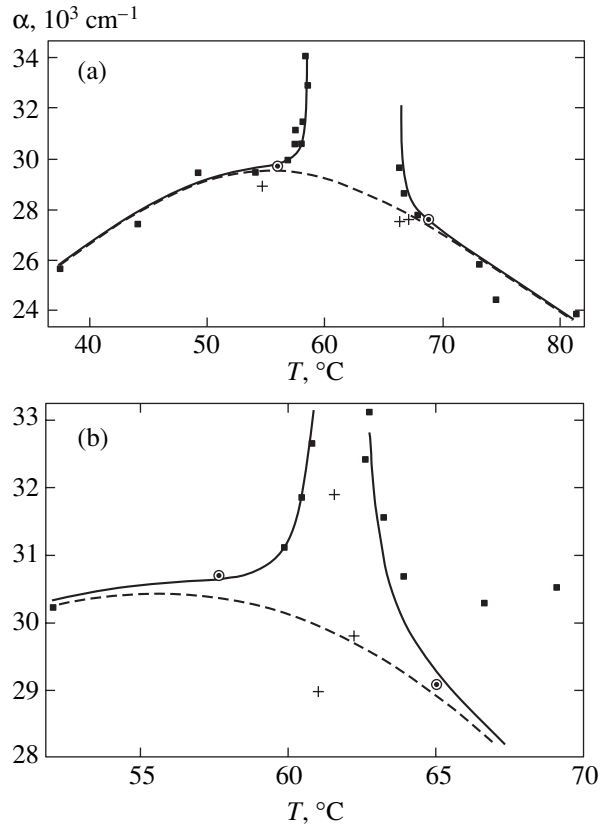

Fig. 3. Results of approximation of experimental curves $\alpha(T)$ by formulas (31)–(36) for solutions with $\Delta T = 7.28$°C [11] (a) and $\Delta T = 0.062$°C (double critical point) [11] (b). Solid curves are the results of fitting and the dashed curves describe background damping α_B. Points marked by crosses are not taken into account in the approximation. Points bounding the regions of approximation are encircled.

the velocity of propagation of ultrasound in such a solution [13, 15] exhibit a number of peculiar features.

(i) The value of the temperature coefficient dV/dT of the hypersound velocity for $T > T_U$ is half as large as for $T < T_L$.

(ii) A temperature region exists in the vicinity of the double critical point and peculiar points, where $dV/dT = 0$.

(iii) The temperature coefficients of the velocity of ultrasound for $T > T_U$ and $T < T_L$ are identical.

(iv) For $T < T_L$, anomalously strong dispersion of sound attaining 22% is observed (Fig. 2).

(v) The hypersound damping coefficient sharply increases in the vicinity of the upper, lower, and double critical points. This increase is observed against the background of a broad and smooth attenuation peak. Such a background damping is determined by features of the main components of the solution and is not associated with critical phenomena (Fig. 3).

(vi) In the vicinity of the peculiar point, an increase in the damping coefficient in the form of a narrow peak against the background of a broad damping peak is also observed (Fig. 4).

All these clearly manifested features characterizing the propagation of hypersound in a solution with a closed phase-separation region have not been explained theoretically so far. Attempts at using the existing theories for this purpose were unsuccessful.

3. THEORY

Propagation of ultra- and hypersound in the vicinity of critical points and phase transitions was studied experimentally and theoretically a long time ago [16–18].

Fixman [17] and Kawasaki [18] described the behavior of the velocity and absorption coefficient of ultrasound in the vicinity of the critical points of phase separation on the basis of interacting mode theories. The formulas of these theories are in satisfactory agreement with the experimental data obtained in the ultrasonic frequency range, in which $\Omega\tau \leq 1$ (Ω is the frequency of sound and τ is the critical fluctuation relaxation time).

However, in the high-frequency range in which $\Omega\tau > 1$, the formulas of the above-mentioned theories give a decrease in the damping coefficient of sound in the vicinity of the critical points, while our experiments reveal a sharp increase in its value.

One of the authors of this paper [19, 20] developed a theory of propagation of sound in the vicinity of the critical points of phase separation taking into account nonlocality; in combination with the theory of interacting modes, this theory led to expressions correctly describing the experimental data even for $\Omega\tau > 1$. Comparison with the experiment was carried out in the frequency range $0 \leq \Omega\tau \leq 50$ (for frequencies of ultrasound up to 75 MHz).

In Chaban's theory [19, 20], fluctuations of the concentration are expanded into a Fourier spectrum in wavevector **k**. In this case, thermodynamic potential Φ (per unit volume) associated with fluctuations of the concentration has the form [21]

$$\Phi - \Phi_0 = \frac{B}{(2\pi)^3}\int d\mathbf{k}\frac{\overline{|c_\mathbf{k}|^2}}{\chi(kr_c)}.$$

Here, $\overline{|c_\mathbf{k}|^2}$ is the spectral intensity of concentration fluctuations with wavevector **k**, $\chi(kr_c) = 1/[1 + (kr_c)^2]$ is the Ornstein–Zernicke function, and r_c is the correlation radius for concentration fluctuations. The value of B for critical concentrations tends to zero as we approach the critical points. We write here the expressions for B in the vicinity of an ordinary critical point as well as the critical points of the closed phase-separation region, the double critical point, and the peculiar point:

$$B = B_1'[|T - T_c| + d_1(c - c_c)^2]^\gamma, \quad (1)$$

$$B = B_1[(T - T_U)(T - T_1) + d_1(c - c_c)^2]^\gamma, \quad (2)$$

$$B = B_1^0[(T - T_0)^2 + d_0(c - c_c)^2]^\gamma, \quad (3)$$

$$B = B_1^{PP}[(T - T_{PP})^2 + a_1(c - c_{PP})^2 + a_2]^\gamma. \quad (4)$$

Here, γ is the critical index of the generalized susceptibility; c is the mean concentration; T_c, c_c, T_0, c_0, and T_{PP}, c_{PP} are the absolute temperatures and concentrations of the ordinary critical, double critical, and peculiar points, respectively; and d, d_1, d_0, a_1, a_2 and B_1, B_1', B_1^0, B_1^{PP} are constants. Here, we disregard the smoothing of the binodal vertex in the vicinity of the critical point, which leads to the dependence $B \propto (c - c_c)^3$. In the case of large closed phase-separation regions, $\gamma = 1.25$ for the critical points. Scattered light intensity I, correlation radius r_c diffusion coefficient D, and characteristic time $\tau = r_c^2/D$ of fluctuation relaxation can be expressed in terms of B:

$$I = A\left(\frac{d\varepsilon}{dc}\right)^2\frac{k_BT}{2B}, \quad (5)$$

$$r_c = r_{c1}(B_1/B)^{1/2}, \quad (6)$$

$$D = [6\pi\eta r_c(B_1/B)^{1/2}]^{-1}, \quad (7)$$

$$\tau = \tau_1(B_1/B)^{3/2}. \quad (8)$$

Here, ε is the dielectric constant; k_B is the Boltzmann constant; η is the shear viscosity; and r_{c1}, τ_1, and A are constants. For the double critical and peculiar points, we must substitute B_1^0 and B_1^{PP} for B_1 in these formulas.

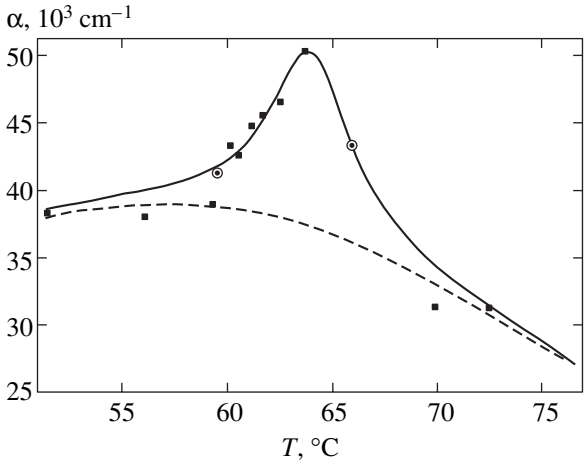

Fig. 4. Result of approximation of experimental curves $\alpha(T)$ by formulas (32)–(37) for a solutions with a peculiar point [14]. The solid curve is the result of fitting and the dashed curve describes background damping α_B. Points bounding the regions of approximation are encircled.

It was shown in [19, 20] that the ultrasound absorption coefficient divided by the square of frequency Ω as a function of temperature and concentration is described by the formula

$$\frac{\alpha}{\Omega^2} = \frac{k_BT\rho V_0\tau}{4\pi^2 r_c^3 B^2}\left(\frac{\partial B}{\partial p}\right)^2 F_1(\Omega\tau). \quad (9)$$

Here, ρ is the density of the solution, V_0 is the velocity of sound extrapolated from a region far from the critical point, and $F_1(\Omega\tau)$ is a function of $\Omega\tau$, which is given in [20] in integrated form. The plot of this function in the interval $1 < \Omega\tau \le 40$ is also given in [19, 20]. This theory makes it possible to correctly describe the behavior of the dispersion in the velocity and absorption coefficient of ultrasound in the vicinity of the critical points up to $\Omega\tau \approx 50$.

However, attempts at applying the theory developed in [19, 20] for calculating the hypersound damping coefficient in a solution with a phase-separation dome were not successful [22]. The experimental data [11–14] pertain to the range $10^4 \le \Omega\tau \le 10^{10}$. In this range, the upper integration limit for function $F_1(\Omega\tau)$ cannot be set at infinity, but should be assumed to be equal to the cutting factor $b = 5$ introduced in [20]. This value corresponds approximately to the condition of equality between the correlation radius and wavelength of the Fourier component of the concentration fluctuations. For such an upper limit in the range $10^4 \le \Omega\tau \le 10^{10}$, function $F_1(\Omega\tau)$ is found to be proportional to $(\Omega\tau)^{-2}$. This dependence leads to a sharp decrease in the damping coefficient with decreasing distance from the criti-

cal and peculiar points. The damping coefficient is found to be proportional to $B^{(-2+3\gamma)/\gamma}$ and decreases in the critical region.

The unsuccessful attempt at describing the behavior of the damping coefficient of hypersound using the theory developed in [19, 20] stimulated the search for a new mechanism responsible for the experimental facts established in [11–14].

It should be noted that the above-mentioned theories [17–20] described precisely the absorption of ultrasound, i.e., losses in an acoustic wave due to the conversion of its energy into heat. Such a conversion is due to a delay in the change in the density of the medium relative to the change in the pressure of the medium upon the relaxation of the medium (relaxation in concentration fluctuations in our case). However, attenuation of the acoustic wave can be due to absorption or scattering of sound from inhomogeneities. We assumed that the main mechanism of hypersound damping in a medium with developed fluctuations for large values of $\Omega\tau$ is the scattering of a hypersonic wave from concentration fluctuations. The damping coefficient associated with scattering of the acoustic wave describes the width of the Mandelstam–Brillouin components in analogy to the absorption coefficient in [5]. In the case of ultrasound, this mechanism is insignificant since the ultrasound wavelength is much larger than the correlation radius under the standard experimental conditions. The hypersound wavelength in the experiments [11–14] is comparable to the correlation radius. This leads to strong scattering of hypersound, which is associated with an incoherent composition of scattered waves (their intensities are added).

Let us calculate the hypersound damping coefficient associated with scattering of hypersound from concentration fluctuations. These calculations repeat in many respects the calculations of the intensity of light scattered from concentration fluctuations in the vicinity of the critical points which were carried out by Fabelinskiĭ [8].

Compressibility β of a liquid and its density change upon a variation of the concentration. To simplify the analysis, we assume that the relative change in the compressibility is much larger than the relative change in the density. The final formulas derived here can easily be generalized to the case when the change in the density is taken into account. Let us suppose that a medium contains a spherical region of radius a, in which compressibility β_1 differs from the volume-averaged compressibility β. The pressure in an acoustic wave scattered in this region is given by the known expression [23]

$$p_{sc} = p_0 \frac{q^2 a^3}{3} \frac{\beta_1 - \beta}{\beta} \frac{\exp(iqr)}{r}, \qquad (10)$$

where p_0 is the pressure in the incident plate acoustic wave, q is the wavenumber of the acoustic wave, and r is the distance from the center of this region to the point under investigation. We assume that intensity I_{sc} of the scattered wave is equal to the time-averaged energy flux through a sphere of radius L with its center at the center of the region studied here,

$$I_{sc} = 4\pi L^2 \overline{\rho v_s}\big|_{r=L} \frac{2\pi p_0^2}{pV_0}\left(\frac{q^2 a^3}{3}\frac{\beta_1 - \beta}{\beta}\right)^2, \qquad (11)$$

where v_s is the velocity of displacement of particles in the acoustic wave.

Let us write the ratio $(\beta_1 - \beta)/\beta$ in the form

$$\frac{\beta_1 - \beta}{\beta} = \frac{\partial\beta}{\partial c}\frac{\Delta c}{\beta},$$

where Δc is the deviation of the concentration from the mean value in the region studied. In the subsequent analysis, we assume that a region of radius a is a region with correlation radius r_c. The intensity of hypersound scattered from fluctuations of the concentration per unit volume is described by the expression

$$I_{sc} = I_0 \frac{2(2\pi)^3 r_c^3}{9\Lambda^4}\left(\frac{\partial\beta}{\partial c}\right)^2 \frac{\overline{(\Delta c)^2}}{\beta^2}. \qquad (12)$$

Here, I_0 is the intensity of the incident acoustic wave, Λ is the hypersound wavelength, and $\overline{(\Delta c)^2}$ is the volume-averaged square of deviation of the concentration from the mean value,

$$\overline{(\Delta c)^2} = \frac{2\pi k_B T}{B}\int_0^{b/r_c} \chi(kr_c)k^2 dk = \frac{2\pi k_B T}{Br_c^3}(b - \arctan b). \qquad (13)$$

Here, as before, b is the cutoff factor. Substituting expression (13) into (12) and setting $b = 5$, we obtain

$$I_{sc} = I_0 G/B,$$

where

$$G = \frac{7.3(2\pi)^4 k_B T(\partial\beta/\partial c)^2}{9\Lambda^4\beta^2}.$$

After a time interval of 1 s, the intensity of the hypersound wave becomes

$$I = I_0(1 - GV_0/B), \qquad (14)$$

which corresponds to the following hypersound damping coefficient per unit length:

$$\alpha = G/B. \qquad (15)$$

Substituting into this formula the expressions for B in the vicinity of the ordinary critical points, critical

points of a solution with a closed phase-separation region, and double critical and peculiar points, we obtain, respectively,

$$\alpha = \frac{G}{B_1'}[|T - T_c| + d_1(c - c_c)^2]^{-\gamma}, \quad (16)$$

$$\alpha = \frac{G}{B_1}[(T - T_U)(T - T_L) + d_1(c - c_c)^2]^{-\gamma}, \quad (17)$$

$$\alpha = \frac{G}{B_1^0}[(T - T_0)^2 + d_0(c - c_c)^2]^{-\gamma}, \quad (18)$$

$$\alpha = \frac{G}{B_1^{PP}}[(T - T_{PP})^2 + a_{PP}(c - c_{PP})^2 + a_2]^{-\gamma}. \quad (19)$$

These formulas mainly differ by the expressions in the brackets. Allowance for variation of the density with the concentration only leads to a certain change in the factors in front of the brackets.

As the temperature approaches the double critical point, formula (18) for $c = c_c$ gives the following temperature dependence of α:

$$\alpha \propto (T - T_0)^{-2\gamma}. \quad (20)$$

Let us now discuss which value of γ must be chosen for this formula. For ordinary critical phase-separation points, we have $\gamma = 1.25$. For the double critical point, the value of γ is apparently different. It was shown in [24, 25] that the critical index of the correlation radius in the vicinity of the double critical point is equal approximately to 0.5. In this region, $r_c \propto B^{-1/2} \propto (T - T_0)^{-1}$; consequently, $\gamma = 1$ and

$$\alpha \propto (T - T_0)^{-2}. \quad (21)$$

For large closed regions of phase separation, the hypersound damping coefficient is described by formula (17), where $T - T_L \approx$ const (or $T - T_U \approx$ const); for $c = c_c$, the dependence $\alpha(T)$ in the vicinity of temperature T_c has the form

$$\alpha \propto |T - T_c|^{-\gamma},$$

where $\gamma = 1.25$. For small phase-separation regions, it is convenient to represent the expression in the brackets in formula (17) in the form

$$\alpha \propto B^{-1} \propto [(T - T_0)^2 - \Delta T^2/4 + d(c - c_c)^2]^{-\gamma}, \quad (22)$$

where $\Delta T = T_U - T_L$, while T_0 is, as before, the temperature corresponding to the double critical point and coinciding with the center of the phase-separation region. Thus, for small phase-separation regions, for T_0 $c = c_c$, we have

$$\alpha \propto [(T - T_0)^2 - \Delta T^2/4]^{-\gamma}, \quad (23)$$

while for the peculiar point we can write

$$\alpha \propto [(T - T_{PP})^2 + a_2]^{-\gamma}. \quad (24)$$

It remains unclear which value of γ should be taken in these expressions. For closed phase-separation regions of various widths, this value must vary from 1 to 1.25.

4. COMPARISON OF EXPERIMENTAL RESULTS WITH THE THEORY

Let us first consider the behavior of the velocity of hypersound in the vicinity of the critical points and the peculiar point and then compare the experimental data on damping with the theoretical dependences obtained in the previous section.

Scattering of hypersound from concentration fluctuations does not change the velocity of hypersound. However, the mechanism considered in [19, 20] leads to the following expression for the velocity of sound:

$$V = V_0\left[1 - \frac{k_B T\rho V_0^2}{4\pi^2 r_c^3 B^2}\left(\frac{\partial B}{\partial p}\right)^2 F_2(\Omega\tau) + \frac{k_B T\rho V_0^2}{8\pi^2 r_c^3 B}\frac{\partial^2 B}{\partial p^2}(b - \arctan b)\right]. \quad (25)$$

Function $F_2(\Omega\tau)$ is given in [20] in integrated form. For $\Omega\tau$ varying in the range 10^4–10^{10}, function $F_2(\Omega\tau)$, as well as $F_1(\Omega\tau)$, is proportional to $(\Omega\tau)^{-2}$. According to estimates, the second term in the brackets is much smaller than the last term. Consequently, we can write

$$\frac{V - V_0}{V_0} = \frac{k_B T\rho V_0^2}{8\pi^2 r_c^3 B}\frac{\partial^2 B}{\partial p^2}(b - \arctan b). \quad (26)$$

The dependence of B^{-1} on the temperature and concentration for small phase-separation regions has the form (22); the only term depending on p in this expression is ΔT. The dependence $\Delta T(p)$ was not measured for our solution, but for an analogous solution of α-picoline-D_2O, this dependence has the form

$$\Delta T = A_p\sqrt{p_0 - p}, \quad (27)$$

where $p_0 = 208$ atm and $A_p = (2.01 \pm 0.12) \times 10^{-3}$ °C cm dyne$^{-1/2}$ [22, 26]. Note that the measured dependence $\Delta T(C_x)$ has an analogous form in our case [9, 10].

Assuming that the $\Delta T(p)$ dependence for a guaiacol–glycerol solution has a form similar to (27), we obtain

$$\frac{1}{B}\frac{\partial^2 B}{\partial p^2} = \frac{A_p^4}{16}\gamma(\gamma-1) \times\left[(T-T_0)^2 - \frac{\Delta T^2}{4} + d(c-c_c)^2\right]^{-2}. \quad (28)$$

Noting that

$$r_c^{-3} = r_0^{-3}\left[(T-T_0)^2 - \frac{\Delta T^2}{4} + d(c-c_c)^2\right]^{3\nu}, \quad (29)$$

and substituting expressions (28) and (29) into (26), we find that

$$\frac{V-V_0}{V_0} = \frac{k_B T\rho V_0^2}{128\pi^2 r_0^3}A_p^4\gamma(\gamma-1)(b-\arctan b) \times\left[(T-T_0)^2 - \frac{\Delta T^2}{4} + d(c-c_c)^2\right]^{3\nu-2}, \quad (30)$$

where ν is the critical exponent of the correlation radius, $3\nu - 2 < 0$.

Thus, in accordance with relation (30), a peak at the critical concentration must be observed on the concentration dependence of the velocity of hypersound at constant temperature against the background of a smooth variation of the velocity from its value in glycerol to the value in guaiacol. The width of this peak decreases and the height increases as the temperature approaches the critical value. The qualitative concentration dependences of the velocity of hypersound observed earlier [27] were precisely of this type.

The results of measurements of the damping coefficient of hypersound [11–14] enabled us to compare the above theoretical description with experiment not only qualitatively, but also quantitatively. For this purpose, we used formulas (17)–(24) to approximate the experimentally measured temperature dependence of hypersound damping in guaiacol–glycerol solutions with phase-separation regions $\Delta T = 7.28$°C and $\Delta T = 0.062$°C (i.e., with a double critical point) and with a peculiar point [11, 14] at the critical concentration.

For convenience of approximation of measured dependences $\alpha(T)$ by formulas (17)–(24), we introduced the reduced temperature,

$$\varepsilon = \frac{(T-T_L)(T-T_U)}{T_0^2} = \frac{(T-T_0)^2 - (\Delta T/2)^2}{T_0^2}. \quad (31)$$

Then $B = B_0\varepsilon^\gamma$, where $B_0 = B_1 T_0^{2\gamma}$ is a constant, and we obtain the following expression on account of the fact that $\beta = 1/\rho V^2$ and that total damping coefficient $\alpha(T)$ is the sum of the component associated with critical fluctuations and described by formula (16) and the background damping coefficient $\alpha_B(T)$:

$$\alpha(\varepsilon) = \alpha_B(T) + 0.811 k_B T q^4 \rho^2 V_0^4\left(\frac{\partial\beta}{\partial c}\right)^2\frac{\varepsilon^{-\gamma}}{B_0}. \quad (32)$$

It should be noted that, although the background damping is usually defined as damping in the absence of critical fluctuations, in the present case we must use effective background damping coefficient for $\alpha_B(T)$. It must contain, in addition to the damping coefficient associated with mechanisms not related to critical fluctuations (such as relaxation in a viscous liquid), the absorption coefficient associated with all other mechanisms. For this reason, the effective background damping coefficient $\alpha_B(T)$ outside the critical region (i.e., for $T - T_U > 7$°C and for $T_L - T > 7$°C) is plotted directly through the experimental points (see Figs. 3 and 4).

The experimental data were approximated using the least-squares method for a linear dependence of the logarithm of normalized critical damping coefficient,

$$\ln\frac{\alpha(T)-\alpha_B(T)}{0.811 k_B T q^4 \rho^2 V_0^4 (\partial\beta/\partial c)^2}, \quad (33)$$

on the logarithm of reduced temperature (31).

The values of ρ, V_0, and refractive index n required for calculating the values of q, which appear in expression (33), are known for various solutions and temperatures [11–14] and amount to (subscript 7.28 corresponds to the width of the phase-separation region)

$$\rho_{7.28}[\text{g/cm}^2] = 1.1892 - (T - 20°\text{C})\times 8.912\times 10^{-4},$$

$$\rho_{DCP}[\text{g/cm}^2] = 1.1892 - (T - 20°\text{C})\times 8.9075\times 10^{-4},$$

$$n_{7.28} = 1.497 - (T - 50°\text{C})\times 4.6\times 10^{-4},$$

$$n_{DCP} = 1.497 - (T - 50°\text{C})\times 4.593\times 10^{-4},$$

$$V_{0L}[\text{m/s}] = 2410 - 11.83T[°\text{C}],$$

$$V_{0U}[\text{m/s}] = 2129 - 6.77T[°\text{C}]$$

for a solution with $\Delta T = 7.28$°C and

$$V_{0L}[\text{m/s}] = 2410 - 11.39T[°\text{C}],$$

$$V_{0U}[\text{m/s}] = 2046 - 5.67T[°\text{C}]$$

for a solution with a double critical point. Here, V_{0L} and V_{0U} are the velocities of hypersound at $T < T_L$ and $T > T_U$, respectively.

The temperature dependence of quantity $\partial\beta/\partial c$ is defined in terms of quantities $\partial V/\partial c$ and $\partial\rho/\partial c$ for various temperatures:

$$\frac{\partial\beta}{\partial c} = -\frac{1}{\rho^2}\frac{\partial\rho}{\partial c}\frac{1}{V_0^2} - \frac{2}{\rho V_0^3}\frac{\partial V_0}{\partial c}. \quad (34)$$

It was found that, if concentration c is measured in volume fractions, the temperature dependence of quantity $\partial V_0/\partial c$ obtained from the data presented in [27] in the entire temperature interval 33°C < T < 85°C of interest is correctly described by the expression

$$\frac{\partial V_0}{\partial c}(T)\left[\frac{\text{cm}}{\text{s}}\right] = -1.4644\times10^6 + 7.9280\times10^4 T - 1783T^2 + 18.306T^3 - 0.0708T^4 \quad (T \text{ in } ^\circ\text{C}). \tag{35}$$

Under the same condition, expression (34) can be written in terms of the difference in the densities of the solution components, $\Delta\rho = 0.126$ g/cm^3:

$$\frac{\partial\beta}{\partial c} = \left(\frac{0.126}{\rho} - 2\frac{\partial V_0}{\partial c}\frac{1}{V_0}\right)\frac{1}{\rho V_0^2}. \tag{36}$$

Substituting expression (35) as well as the values of $\rho(T)$ given above into this formula, we obtain the empirical expression for the temperature dependence of quantity $\partial\beta/\partial c$, which is used in normalization (33).

The values of γ and B_0 were defined by approximation by formula (32) of the experimental data presented in Fig. 3a for the solution with ΔT = 7.28°C and in Fig. 3b for the solution with a double critical point.

For the solution with a peculiar point, the approximation was carried out in the same way, the only difference being that the reduced temperature had the form

$$\varepsilon_{PP} = \frac{(T - T_{PP})^2 + a_2}{T_{PP}^2} \tag{37}$$

instead of (31). Parameter a_2, having the meaning of the squared distance from the peculiar point to the double critical point along the C_x axis (see Fig. 1), was initially unknown and was determined by minimizing the residual root-mean-square error of approximation: $\sqrt{a_2}$ = 2.39°C. The values of ρ and n were the same as for the solution with a double critical point; the velocity of hypersound below and above the critical points was defined as $V_{0L} = 2332 - 11.26T$[°C] and V_{0U}[m/s] = $2008 - 6.167T$[°C] [14]. The result of approximation was given above in Fig. 4.

It should be noted that formulas (10)–(19) describe the attenuation of sound due to scattering from spherical fluctuations and are valid as long as $r_c \ll \Lambda$. For scatterers of size $r \geq \Lambda/4$, the increase in the scattering coefficient with r is substantially slowed; the scattering coefficient attains its maximum value at $r \approx 1.5\Lambda$ and then decreases [28]. For this reason, the experimental points for which r_c > 2500 Å were disregarded in approximation (see Fig. 3). It should be noted that the dependence of the scattering coefficient on the particle size, which was theoretically obtained in [28], also explains nonmonotonicity in the dependence $\alpha(T)$ in the immediate vicinity of the upper and lower critical phase-separation points, where the value of r_c becomes comparable with Λ.

As a result of approximation, we obtained the following critical indices γ: $\gamma = 1.14 \pm 0.21$ (or $\gamma = 1.22 \pm 0.22$ in the case of minimization of the residual approximation error α by varying the values of T_U and T_L) for the solution with ΔT = 7.28°C, $\gamma = 0.90 \pm 0.21$ for the solution with a double critical point, and $\gamma = 1.03 \pm 0.10$ for the solution with a peculiar point. These values of γ correspond to the theoretical prediction given by formulas (17)–(19), (23), (24), and (32).

5. CONCLUSIONS

We have studied the behavior of the velocity and damping coefficient of hypersound in solutions with a closed phase-separation region, a double critical point, and a peculiar point and proposed an explanation for the sharp increase in the damping coefficient of hypersound, which is observed in the vicinity of the critical and peculiar points. A theory describing the behavior of hypersound in the critical region, based on the mechanism of hypersound scattering from concentration fluctuations, has been proposed. A comparison of our formulas with the experimental data has revealed good agreement.

ACKNOWLEDGMENTS

The authors are deeply grateful to the late I.L. Fabelinskiĭ for his persistent interest and fruitful discussions of the results of this research.

This study was financially supported by the Russian Foundation for Basic Research (project nos. 02-02-16118 and 03-02-17499).

REFERENCES

1. M. Smoluchowsky, Ann. Phys. (Leipzig) **25**, 205 (1908).
2. A. Einstein, Ann. Phys. (Leipzig) **33**, 1275 (1910).
3. M. A. Anisimov, *Critical Phenomena in Liquids and Liquid Crystals* (Nauka, Moscow, 1987), p. 47 [in Russian].
4. O. A. Shustin, Pis'ma Zh. Éksp. Teor. Fiz. **3**, 491 (1966).
5. M. A. Leontovich, in *Selected Works: Theoretical Physics* (Nauka, Moscow, 1965), p. 100 [in Russian].
6. I. L. Fabelinskiĭ, Dokl. Akad. Nauk **377**, 180 (2001) [Dokl. Phys. **46**, 149 (2001)].
7. M. L. Levin and S. M. Rytov, *Theory of Equilibrium Thermal Fluctuations in Electrodynamics* (Nauka, Moscow, 1967) [in Russian].
8. I. L. Fabelinskiĭ, *Molecular Scattering of Light* (Nauka, Moscow, 1965; Plenum, New York, 1968).
9. S. V. Krivokhizha, O. A. Lugovaya, I. L. Fabelinskiĭ, and L. L. Chaĭkov, Zh. Éksp. Teor. Fiz. **89**, 85 (1985) [Sov. Phys. JETP **62**, 48 (1985)].

10. R. G. Johnston, M. A. Clark, P. Wiltzius, and D. S. Cannall, Phys. Rev. Lett. **54**, 49 (1985).
11. K. V. Kovalenko, S. V. Krivokhizha, I. L. Fabelinskiĭ, and L. L. Chaĭkov, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 395 (1993) [JETP Lett. **58**, 404 (1993)].
12. K. V. Kovalenko, S. V. Krivokhizha, I. L. Fabelinskiĭ, and L. L. Chaĭkov, Dokl. Akad. Nauk **347**, 327 (1996) [Phys. Dokl. **41**, 98 (1996)].
13. S. V. Krivokhizha, I. L. Fabelinskiĭ, and L. L. Chaĭkov, Akust. Zh. **47**, 238 (2001) [Acoust. Phys. **47**, 194 (2001)].
14. K. V. Kovalenko, S. V. Krivokhizha, and L. L. Chaĭkov, Kratk. Soobshch. Fiz., No. 2, 40 (2001) [Bull. Lebedev Phys. Inst., No. 2, 34 (2001)].
15. S. V. Krivokhizha, I. L. Fabelinskiĭ, L. L. Chaĭkov, and A. A. Shubin, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 166 (1996) [JETP Lett. **64**, 177 (1996)].
16. Y. Garrabos, G. Zalczer, and D. Beysens, Phys. Rev. A **25**, 1147 (1982).
17. M. Fixman, J. Chem. Phys. **36**, 1961 (1962).
18. K. Kawasaki, in *Phase Transition and Critical Phenomena*, Ed. by M. S. Green and C. Dome (Academic, New York, 1976), Vol. 5A, p. 165.
19. I. A. Chaban, Akust. Zh. **21**, 104 (1975) [Sov. Phys. Acoust. **21**, 64 (1975)].
20. I. A. Chaban, Akust. Zh. **21**, 286 (1975) [Sov. Phys. Acoust. **21**, 177 (1975)].
21. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1979).
22. S. V. Krivokhizha and L. L. Chaĭkov, Kratk. Soobshch. Fiz., No. 1, 32 (2004) [Bull. Lebedev Phys. Inst., No. 1, 27 (2004)].
23. M. A. Isakovich, *General Acoustics* (Nauka, Moscow, 1973) [in Russian].
24. L. L. Chaikov, I. L. Fabelinskii, S. V. Krivokhizha, *et al.*, J. Raman Spectrosc. **25**, 463 (1994).
25. S. V. Krivokhizha, O. A. Lugovaya, I. L. Fabelinskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **103**, 115 (1993) [JETP **76**, 62 (1993)].
26. C. W. Garland and K. Nichigaku, J. Chem. Phys. **65**, 5298 (1976).
27. K. V. Kovalenko, S. V. Krivokhizha, I. L. Fabelinskiĭ, and L. L. Chaĭkov, Kratk. Soobshch. Fiz., No. 11, 19 (2002) [Bull. Lebedev Phys. Inst., No. 11, 19 (2002)].
28. K. S. Shifrin, *Light Scattering in Turbid Medium* (Gostekhizdat, Moscow, 1951) [in Russian].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 559–567.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 633–642.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

The Vortex State of an Antiferromagnet with Uniaxial Anisotropy

Yu. I. Dzhezherya, M. V. Sorokin, and E. A. Bubuk

Institute of Magnetism, National Academy of Sciences of Ukraine and Ministry of Education and Science of Ukraine, Kiev, 03142 Ukraine

e-mail: sorockin@ukr.net

Received July 12, 2004

Abstract—For the model of an antiferromagnet with easy-axis anisotropy, we have determined the multisoliton state that corresponds to the domain boundary on the surface of which antiferromagnetic vortices of various topological charges are localized. We analyze the pattern of interaction between the vortices. We show that repulsive and attractive forces equivalent to the forces of electrostatic interaction between point charges in a plane act between the vortices with like and unlike topological charges, respectively. However, there is a steady state in this case, since these vortices in the model of a uniaxial antiferromagnet have an infinite effective mass and do not change their relative positions. We give a general solution that describes the vortex structures on the surface of the domain boundary. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Planar domain boundaries (DBs) and isolated stripe domains are among the best studied elements of domain structures. Their mathematical images are one-dimensional soliton solutions of the Landau–Lifshitz equations in the form of kinks and breathers, respectively.

The magnetic and antiferromagnetic solitons in two- and three-dimensional systems are less well understood than their one-dimensional counterparts. The first multidimensional solution that describes the vortex states in an isotropic ferromagnet (FM) was found by Belavin and Polyakov [2]. Khodenkov [2, 3] pointed out the existence of a class of particular solutions to the Landau–Lifshitz equations in the model of an easy-axis FM, considered the solution describing a single magnetic vortex [2] and a pair of interacting vertices [3], and surmised that such vortex structures exist in easy-axis antiferromagnets (AFMs).

When the domain structure in collinear antiferromagnetic systems is considered, no problem of allowing for the magnetodipole contributions exists. The equations of motion for the antiferromagnetic vector are invariant to the Lorentz transformation and admit the existence of a dynamical soliton state. In this paper, we find a three-dimensional solution of the Landau–Lifshitz equations that describes a family of topological solitons for the model of an easy-axis AFM. The solution admits a dynamical generalization and incorporates the influence of the magnetic field directed along the anisotropy axis.

We discuss the interaction and stability of a system of antiferromagnetic vortices in actual antiferromagnetic systems.

Let us consider the model of an unbounded two-sublattice AFM with easy-axis anisotropy whose direction coincides with the z axis of the coordinate system. The AFM is assumed to be in a uniform magnetic field directed along the easy axis in z. The expression for the energy density of this system is

$$W = \frac{1}{2}A\mathbf{M}^2 + \frac{\alpha_1}{2}(\nabla L_i)^2 + \frac{\alpha_2}{2}(\nabla M_i)^2 - \frac{\beta_1}{2}L_z^2 - \frac{\beta_2}{2}M_x^2 - M_z H, \quad (1)$$

where A is the uniform exchange constant; α_1, α_2 are the nonuniform exchange constants; β_1, β_2 are the uniaxial anisotropy constants; $i = x, y, z$; $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$ is the magnetization vector; $\mathbf{M}_1$, $\mathbf{M}_2$ are the magnetizations of the AFM sublattices; $\mathbf{L} = \mathbf{M}_1 - \mathbf{M}_2$ is the antiferromagnetic vector, the main parameter of the system; and

$$\nabla = \mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z}.$$

Based on the natural (for an AFM) assumption that the relativistic interaction energy is negligible compared to the exchange energy, the equations of motion for the vectors $\mathbf{M}$ and $\mathbf{L}$ can be reduced to one equation for the antiferromagnetic vector $\mathbf{L}$. For a system with the energy density (1), this equation is [4]

$$\left[\mathbf{l}\times\left(\nabla^2\mathbf{l} - \frac{1}{c^2}\frac{\partial^2\mathbf{l}}{\partial t^2}\right)\right] - \frac{2\omega_0}{c^2}\frac{\partial\mathbf{l}}{\partial t}(\mathbf{l}\cdot\mathbf{h}) + \frac{\omega_0^2}{c^2}(1-\mathbf{h}^2)[\mathbf{l}\times\mathbf{e}_z](\mathbf{l}\cdot\mathbf{e}_z) = 0, \quad (2)$$

1063-7761/05/10003-0559$26.00

where $\mathbf{l} = \mathbf{L}/2M_0$ is the reduced antiferromagnetic vector, $\mathbf{h} = \mathbf{H}/2M_0\sqrt{A\beta_1}$ is the reduced magnetic field directed along the anisotropy axis parallel to the z axis, $c = (4\mu_0 M_0/\hbar)\sqrt{A\alpha_1}$ is the characteristic velocity, $\omega_0 = c/\delta$ ($\delta = \sqrt{\alpha_1\beta_1}$) is the characteristic length, and μ_0 is the Bohr magneton.

This procedure for deriving the equations of motion for AFMs can be familiarized with, for example, in [4] and [5].

Equation (2) has a fairly wide validity range defined by the relations

$$H \ll AM_0, \quad \mathbf{M}^2 \ll \mathbf{L}^2. \tag{3}$$

The components of the vector $\mathbf{M}$ can be expressed in terms of the vector $\mathbf{L}$ and its derivatives [5]:

$$\mathbf{M} = \frac{\hbar}{8\mu_0 AM_0^2}\frac{\partial \mathbf{L}}{\partial t}\times\mathbf{L} + \frac{1}{4AM_0^2}\mathbf{L}\times[\mathbf{H}\times\mathbf{L}]. \tag{4}$$

Since the length of the antiferromagnetic vector is virtually constant in approximation (3), for the subsequent calculations, it is appropriate to introduce the angular variables θ and φ related to the components of the vector $\mathbf{L}$ by

$$\mathbf{l} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta). \tag{5}$$

Equation of motion for AFM (2) in terms of the angular variables takes the form

$$\begin{aligned} &-\frac{\partial}{\partial t}\sin^2\theta(\dot\varphi - \omega_H) + c^2\mathrm{div}[(\nabla\varphi)\sin^2\theta] = 0,\\ &\ddot\theta - c^2\nabla^2\theta + [\omega_0^2 + c^2(\nabla\varphi)^2 - (\dot\varphi - \omega_H)^2]\\ &\qquad\times\sin\theta\cos\theta = 0, \end{aligned} \tag{6}$$

where $\omega_H = 2\mu_0 H/\hbar$ and the dots over the functions denote time differentiation.

In order not to restrict the generality of our analysis, let us pass to a different orthogonal coordinate system with unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. In doing so, we assume that the new coordinate system is oriented arbitrarily with respect to the original one.

Equations (6) remain unchanged, with the only difference that

$$\nabla = \mathbf{e}_1\frac{\partial}{\partial\xi} + \mathbf{e}_2\frac{\partial}{\partial\eta} + \mathbf{e}_3\frac{\partial}{\partial\zeta}.$$

The radius vector of a point in space in the new coordinate system is

$$\boldsymbol{\rho} = \mathbf{e}_1\xi + \mathbf{e}_2\eta + \mathbf{e}_3\zeta.$$

Let us consider the localized antiferromagnetic states that would be homogeneous at infinity. Since we examine various antiferromagnetic configurations, the explicit form of the boundary conditions is concretized below.

Assuming that the antiferromagnetic configuration moves at constant velocity v along the ξ axis, we represent the polar and azimuthal angles of the antiferromagnetic vector as

$$\begin{aligned} \theta(\boldsymbol{\rho}) &= 2\arctan\exp[u(\xi^*, \eta, \zeta)],\\ \varphi &= \psi(\xi^*, \eta, \zeta) + \omega' t, \end{aligned} \tag{7}$$

where $\xi^* = (\xi - vt)/\sqrt{1 - v^2/c^2}$ and ω' is the antiferromagnetic-vector precession frequency.

Introducing the variable ξ^* essentially implies the passage to a comoving coordinate system that uniformly moves together with the antiferromagnetic configuration.

Given that

$$\frac{\partial\theta}{\partial u} = \sin\theta, \quad \frac{\partial^2\theta}{\partial u^2} = \sin\theta\cos\theta,$$

Eqs. (6) in the comoving coordinate system transform into

$$\begin{aligned} &2\left(c^2(\nabla^* u)(\nabla^*\psi) + \frac{v\omega}{\sqrt{1 - v^2/c^2}}\frac{\partial u}{\partial\xi^*}\right)\\ &\qquad\times\sin\theta\cos\theta = -c^2(\nabla^{*2}\psi)\sin\theta,\\ &\left(c^2(\nabla^* u) - c^2(\nabla^*\psi)^2 - \omega_0^2 + \omega^2 - 2\frac{v\omega}{\sqrt{1 - v^2/c^2}}\frac{\partial\psi}{\partial\xi^*}\right)\\ &\qquad\times\sin\theta\cos\theta = -c^2(\nabla^{*2} u)\sin\theta, \end{aligned} \tag{8}$$

where

$$\nabla^* = \mathbf{e}_1\frac{\partial}{\partial\xi^*} + \mathbf{e}_2\frac{\partial}{\partial\eta} + \mathbf{e}_3\frac{\partial}{\partial\zeta}, \quad \omega = \omega' - \omega_H.$$

In particular, it follows from Eqs. (8) that these become identities if ψ and u are harmonic functions and satisfy the equations

$$(\nabla^{*2} u)\sin\theta = 0, \quad (\nabla^{*2}\psi)\sin\theta = 0. \tag{9}$$

At the same time, these quantities must satisfy the additional conditions

$$\begin{aligned} &c^2(\nabla^* u)(\nabla^*\psi) + \frac{v\omega}{\sqrt{1 - v^2/c^2}}\frac{\partial u}{\partial\xi^*} = 0,\\ &c^2(\nabla^* u)^2 - c^2(\nabla^*\psi)^2 - \omega_0^2 + \omega^2\\ &\qquad - 2\frac{v\omega}{\sqrt{1 - v^2/c^2}}\frac{\partial\psi}{\partial\xi^*} = 0. \end{aligned} \tag{10}$$

Let us determine the localized antiferromagnetic configurations that satisfy Eqs. (9) and (10).

2. VORTEX STATES IN THE MODELS OF EASY-AXIS AND ISOTROPIC AFMs

(1) Let us require, for example, that $\theta = 0$ or $\theta = \pi$, $\partial\theta/\partial\xi^* = 0$, and the derivative $\partial\psi/\partial\xi^*$ be bounded for $\xi^* \longrightarrow \pm\infty$.

In the simplest case, these conditions can be satisfied by choosing ψ and u in the form of linear functions:

$$u = (\xi^* - \xi_0)/\delta^*, \quad \psi = k\xi^* + \psi_0,$$

$$\delta^* = \delta\sqrt{\frac{1-(v/c)^2}{1-(\omega/\omega_0)^2-(v/c)^2}}, \qquad (11)$$

$$k = -\frac{v}{c}\frac{\omega}{\sqrt{c^2-v^2}},$$

where ψ_0 and ξ_0 are the integration constants.

Configuration (7), (11) describes a 180-degree DB whose uniform motion is accompanied by a precession of the antiferromagnetic vector. The dynamical properties of this structure were studied in detail in [5]. Basically, ξ_0 defines the DB coordinate.

Our objective is to find a three-dimensional multisoliton solution. We seek this solution in the form of a DB with the coordinate ξ_0 and the azimuthal angle ψ_0 modulated in space by assuming that

$$\begin{aligned} u(\mathbf{\rho}, t) &= \xi^*/\delta^* - \xi_0(\eta, \zeta)/\delta^*, \\ \psi(\mathbf{\rho}, t) &= k\xi^* + \psi_0(\eta, \zeta). \end{aligned} \qquad (12)$$

Substituting (12) into (9), we find that the sought functions ξ_0 and ψ_0 must satisfy the equations

$$\begin{aligned} \sin\theta\left(\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\zeta^2}\right)\psi_0(\eta, \zeta) &= 0, \\ \sin\theta\left(\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\zeta^2}\right)\xi_0(\eta, \zeta) &= 0. \end{aligned} \qquad (13)$$

The additional conditions (10) for these quantities take the form

$$\begin{aligned} \left(\frac{\partial\xi_0}{\partial\eta}\right)^2 + \left(\frac{\partial\xi_0}{\partial\zeta}\right)^2 &= \delta^{*2}\left(\frac{\partial\psi_0}{\partial\eta}\right)^2 + \delta^{*2}\left(\frac{\partial\psi_0}{\partial\zeta}\right)^2, \\ \frac{\partial\psi_0}{\partial\eta}\frac{\partial\xi_0}{\partial\eta} + \frac{\partial\psi_0}{\partial\zeta}\frac{\partial\xi_0}{\partial\zeta} &= 0. \end{aligned} \qquad (14)$$

For Eqs. (13) to turn into an identity in the region of space where $\sin\theta \neq 0$, $\psi_0(\eta, \zeta)$ and $\xi_0(\eta, \zeta)$ must be the solutions of the Laplace equations and harmonic functions. The additional conditions (14) are satisfied automatically if these functions are related by the Cauchy–Riemann conditions:

$$\begin{aligned} \frac{\partial\xi_0(\eta, \zeta)}{\partial\zeta} &= \delta^*\frac{\partial\psi_0(\eta, \zeta)}{\partial\eta}, \\ \frac{\partial\xi_0(\eta, \zeta)}{\partial\eta} &= -\delta^*\frac{\partial\psi_0(\eta, \zeta)}{\partial\zeta}. \end{aligned} \qquad (15)$$

Thus, the coordinate $\xi_0(\eta, \zeta)$ of the DB and the azimuthal angle $\psi_0(\eta, \zeta)$ of the magnetization vector are arbitrary conjugate harmonic functions.

Such a result also holds for the FM model with easy-axis anisotropy [2, 3]. It is easy to show that this relationship between $\xi_0(\eta, \zeta)$ and $\psi_0(\eta, \zeta)$ holds for the soliton configurations of isotropic FMs and AFMs.

Certain boundary conditions should be specified to unambiguously determine the AFM configuration. In this paper, from the entire variety of conjugate harmonic functions, we choose those that describe the antiferromagnetic vortices by assuming that

$$\begin{aligned} &\xi_0(\eta, \zeta) \\ &= \delta^*\sum_{k=1}^{N}\nu_k\ln\left[\frac{1}{\delta^*}\sqrt{(\eta-\eta_k)^2 + (\zeta-\zeta_k)^2}\right] + C_1, \end{aligned} \qquad (16)$$

where C_1 is an arbitrary constant that defines the DB location along the ξ axis, N is the number of vortices, (η_k, ζ_k) is the coordinate of the core of vortex k, and ν_k is a number that defines its topological charge. Below, we return to the question of ν_k values.

We find from the Cauchy–Riemann conditions (15) that, to an arbitrary constant, the function $\psi_0(\eta, \zeta)$ conjugate to $\xi_0(\eta, \zeta)$ is

$$\psi_0(\eta, \zeta) = \sum_k \nu_k \arctan\left(\frac{\eta-\eta_k}{\zeta-\zeta_k}\right) + C_2, \qquad (17)$$

where C_2 is the integration constant.

The coordinate $\xi_0(\eta, \zeta)$ of the DB and the azimuthal angle $\psi_0(\eta, \zeta)$ of the magnetization vector (see (16) and (17)) are harmonic functions on the entire $\eta\zeta$ plane, except for the coordinates (η_k, ζ_k) of the vortex cores, since the Laplace equation for $\xi_0(\eta, \zeta)$ at these points does not turn into an identity.

Nonetheless, Eqs. (16) and (17) are still the solution of Eqs. (13), since $\sin\theta$ becomes zero at the singular points (η_k, ζ_k) where $\xi_0(\eta, \zeta)$ and $\psi_0(\eta, \zeta)$ are not harmonic functions.

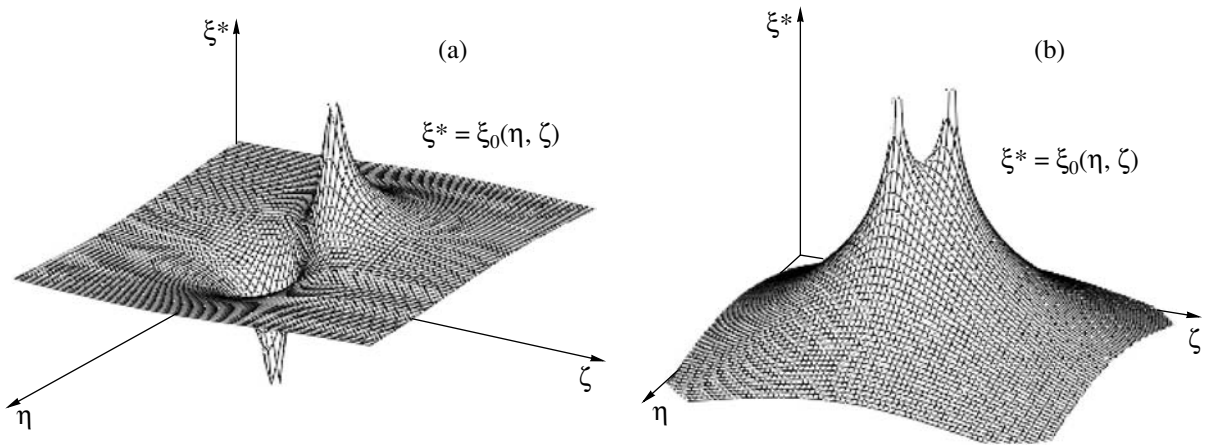

Fig. 1. Image of the surface of the AFM DB with two vortices at (a) zero ($\nu_1 = -\nu_2$) and (b) nonzero ($\nu_1 = \nu_2 < 0$) total topological charges of the vortices.

Thus, we found a three-dimensional solution of the equations of motion for AFM:

$$\tan\frac{\theta(\mathbf{\rho}, t)}{2} = \exp\left(\pm\frac{\xi^* - \xi_0(\eta, \zeta)}{\delta^*}\right),$$
$$\varphi(\mathbf{\rho}, t) = \sum_k^N \nu_k \arctan\left(\frac{\eta - \eta_k}{\zeta - \zeta_k}\right) + k\xi^* + \omega' t + C_2. \quad (18)$$

This solution describes the family of static antiferromagnetic vortices localized on the surface of a moving DB. The vortex axes can be oriented arbitrarily with respect to the anisotropy axis.

An important result of our study is the conclusion that the presence of vortices on the surface of the DB does not affect the dynamical parameters of the latter, since the DB velocity, the precession frequency, and the wavenumber k do not depend on the topological charge of the individual vortices and the vortex structure as a whole.

It follows from our qualitative analysis of configuration (18) that the points in space where the condition $\xi^* = \xi_0(\eta, \zeta)$ is satisfied form the DB surface in passing through which the antiferromagnetic vector component $l_z = \mathbf{1} \cdot \mathbf{e}$ changes its sign. Figure 1 shows fragments of such surfaces for various AFM configurations.

Clearly, if the total charge of the system of vortices, $\sum_k \nu_k$, is zero, then the DB surface far from it is flat. If, alternatively, this charge is nonzero, then the DB shape is close to the surface formed by the rotation of the $\xi^* = (\sum_k \nu_k)\ln\eta$ curve around the ξ^* axis.

In the easy-axis AFM model, the DB surface asymptotically extends to infinity along the vortex axis, contracting exponentially. In fact, the validity range of this result is determined by the restrictions in using the continual approach of the phenomenological theory of antiferromagnetism. A natural restriction of the antiferromagnetic vortex diameter is the lattice parameter a.

(2) The next localized configuration whose existence is allowed by the easy-axis AFM model is the family of vortices moving at constant velocity over the surface of a 180-degree DB.

To determine this structure, let us specify the boundary conditions by requiring that $\theta = 0$ or $\theta = \pi$, $\partial\theta/\partial\zeta = 0$, and the derivative $\partial\psi/\partial\zeta$ be bounded for $\zeta \longrightarrow \pm\infty$. As in the previous case, we seek a solution in form (7) by assuming that the functions $u(\mathbf{\rho}, t)$ and $\psi(\mathbf{\rho}, t)$ satisfy Eqs. (9) and the additional conditions (10).

The simplest soliton solution of Eqs. (9) and (10) corresponding to the specified boundary conditions is a planar 180-degree DB:

$$\theta = 2\arctan(e^u), \quad u = (\zeta - \zeta_0)/\delta,$$
$$\varphi = \omega_H t + \psi_0, \quad (19)$$

where ζ_0 and ψ_0 are the integration constants.

We seek a three-dimensional multisoliton configuration in the form of DB (19) with the coordinate ζ_0 and the azimuthal angle ψ_0 modulated in space by assuming that

$$u(\mathbf{\rho}, t) = \frac{1}{\delta}[\zeta - \zeta_0(\xi^*, \eta)],$$
$$\psi_0 = \psi_0(\xi^*, \eta). \quad (20)$$

Substituting (20) into (9), we find that the sought functions ζ_0 and ψ_0 must satisfy the equations

$$\sin\theta\left(\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\xi^{*2}}\right)\psi_0(\xi^*, \eta) = 0,$$
$$\sin\theta\left(\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\xi^{*2}}\right)\zeta_0(\xi^*, \eta) = 0. \quad (21)$$

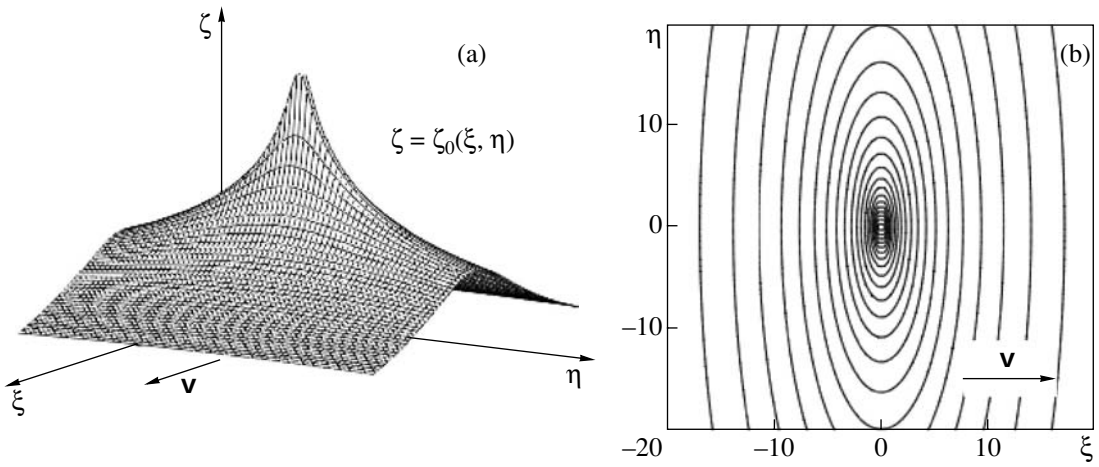

Fig. 2. (a) Image of the surface of the AFM DB with one vortex moving at constant velocity v parallel to the ξ axis and (b) the isolines formed by the sections of this surface by planes perpendicular to the ζ axis. The ratio $v^2/c^2 = 0.9$.

The additional conditions (10) for these quantities take the form

$$\left(\frac{\partial\zeta_0}{\partial\eta}\right)^2 + \left(\frac{\partial\zeta_0}{\partial\xi^*}\right)^2 = \delta^2\left(\frac{\partial\psi_0}{\partial\eta}\right)^2 + \delta^2\left(\frac{\partial\psi_0}{\partial\xi^*}\right)^2,$$
$$\frac{\partial\psi_0}{\partial\eta}\frac{\partial\zeta_0}{\partial\eta} + \frac{\partial\psi_0}{\partial\xi^*}\frac{\partial\zeta_0}{\partial\xi^*} = 0. \quad (22)$$

As in the previous case, the system of equations and additional conditions (21) and (22) has a solution that describes the family of vortices. However, these vortices execute simultaneous motion at constant velocity along the surface of the DB at rest.

The expression that describes this configuration is

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left(\pm\frac{\zeta - \zeta_0(\xi^*, \eta)}{\delta}\right),$$
$$\varphi(\boldsymbol{\rho}, t) = \sum_k^N \nu_k \arctan\left(\frac{(\eta - \eta_k)\sqrt{1 - v^2/c^2}}{\xi - vt - \xi_k}\right) + \omega_H t + C_2, \quad (23)$$

where

$$\zeta_0(\xi^*, \eta) = \delta\sum_k^N \nu_k \ln\left(\frac{1}{\delta}\sqrt{(\eta - \eta_k)^2 + \frac{(\xi - vt - \xi_k)^2}{1 - v^2/c^2}}\right) + C_1.$$

The points in space where the condition $\zeta = \zeta_0(\xi^*, \eta)$ is satisfied from the DB surface, in passing through which the antiferromagnetic vector component $l_z = \mathbf{l}\cdot\mathbf{e}_z$ changes its sign. Figure 2a shows a fragment of this surface that contains an antiferromagnetic vortex moving at constant velocity. Figure 2b display the isolines formed by the section of the vortex by planes perpendicular to its axis. Figure 2 illustrates the Lorentz transformation of the structure of the vortex caused by its motion: the linear sizes of the vortex in the direction of its motion reduce by a factor of $1/\sqrt{1 - v^2/c^2}$.

(3) By arranging the vortices in space in a certain way, we can obtain various AFM configurations. Below, we give two of these defined by elementary functions.

For example, if vortices (23) with identical topological charges $\nu_n = \nu$ are arranged in the form of a periodic chain along the η axis, then the distribution of the angle φ can be represented as

$$\varphi(\boldsymbol{\rho}, t) = \nu\sum_{n=-\infty}^{\infty} \arctan\left(\frac{\eta - dn}{\xi^*}\right) + \omega_H t + C_2, \quad (24)$$

where d is the vortex period. The result of the summation of this series can be expressed in terms of elementary functions [6]:

$$\varphi(\boldsymbol{\rho}, t) = \nu\arctan\left[\tan\left(\frac{\pi\eta}{d}\right)\coth\left(\frac{\pi\xi^*}{d}\right)\right] + \omega_H t + C_2. \quad (25)$$

Based on the expression for the function $\varphi(\boldsymbol{\rho}, t)$, we find the function $u(\boldsymbol{\rho}, t)$ conjugate to it; the polar angle of the antiferromagnetic vector can then be written as

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left[\pm\frac{\zeta - \zeta_0(\xi^*, \eta)}{\delta}\right], \quad (26)$$

where

$$\zeta_0(\xi^*, \eta) = \delta\frac{\nu}{2}\ln\left[\sinh^2\left(\frac{\pi\xi^*}{d}\right) + \sin^2\left(\frac{\pi\eta}{d}\right)\right] + C_1.$$

In a similar way, we can show that the configuration of the antiferromagnetic vector for a chain of vortices with alternating signs of the topological charge is defined by

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left[\pm\frac{\zeta - \zeta_0(\xi^*, \eta)}{\delta}\right],$$
$$\varphi(\boldsymbol{\rho}, t) = \nu\arctan\left[\frac{\sin(\pi\eta/d)}{\sinh(\pi\xi^*/d)}\right] + \omega_H t + C_2, \quad (27)$$

where

$$\zeta_0(\xi^*, \eta) = \delta\frac{\nu}{2}\ln\left[\frac{\cosh(\pi\xi^*/d) - \cos(\pi\eta/d)}{\cosh(\pi\xi^*/d) + \cos(\pi\eta/d)}\right] + C_1.$$

We will not analyze the distribution of the antiferromagnetic vector for these configurations and give only the images of the surfaces that separate the domains with the opposite signs of the antiferromagnetic vector component l_z (Fig. 3). As we see from Fig. 3b, the DB surface is flat far from the chain of vortices with alternating signs of the topological charge. If, however, the signs are identical, then the DB far from the chain of vortices consists of two flat regions oriented at an angle χ relative to each other (Fig. 3a). This angle depends on the topological charge ν and the vortex period d and is defined by

$$\chi = \pi - 2\arctan\frac{\pi\nu\delta}{d}.$$

(4) The transformation of the derived configurations when passing to the isotropic AFM model is of considerable interest.

The configuration that corresponds to the family of vortices in an isotropic AFM can be obtained without additional calculations from relations (18), (23), (26), and (27) by passing to the limit $\beta_1 \longrightarrow 0$ ($\delta \longrightarrow \infty$).

Thus, for an AFM whose structure is described by (23), passing to the limit yields

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \prod_k\left(q\sqrt{(\eta - \eta_k)^2 + (\xi^* - \xi_k^*)^2}\right)^{\nu_k},$$
$$\varphi(\boldsymbol{\rho}, t) = \sum_k \nu_k\arctan\left(\frac{\eta - \eta_k}{\xi^* - \xi_k^*}\right) + \omega_H t + C_2, \quad (28)$$

where q is an arbitrary constant in units of reciprocal length.

Relations (28) define a configuration with two-dimensional modulation of the antiferromagnetic vector in the $\xi\eta$ plane. In the static case, this structure closely coincides with the family of Belavin–Polyakov solitons [1] and may be considered as a special case of the soliton configurations of easy-axis anisotropic systems.

When passing to an isotropic AFM, expressions (25) and (26), which describe the chain of vortices with like charges, transform into

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \tilde{C}_1\left[\sinh^2\left(\frac{\pi\xi^*}{d}\right) + \sin^2\theta\left(\frac{\pi\eta}{d}\right)\right]^{\nu/2},$$
$$\varphi(\boldsymbol{\rho}, t) = \nu\arctan\left[\tan\left(\frac{\pi\eta}{d}\right)\coth\left(\frac{\pi\xi^*}{d}\right)\right] + \omega_H t + C_2, \quad (29)$$

where $\tilde{C}_1$ and C_2 are arbitrary constants.

For the chain of vortices with alternating signs of the topological charge, the distribution of the antiferromagnetic vector in an isotropic system can be written as

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \tilde{C}_1\left(\frac{\cosh(\pi\xi^*/d) - \cos(\pi\eta/d)}{\cosh(\pi\xi^*/d) + \cos(\pi\eta/d)}\right)^{\nu/2},$$
$$\varphi(\boldsymbol{\rho}, t) = \nu\arctan\left(\frac{\sin(\pi\eta/d)}{\sinh(\pi\xi^*/d)}\right) + \omega_H t + C_2. \quad (30)$$

It follows from Eqs. (28)–(30) that the vortices in the isotropic AFM model have a two-dimensional structure and differ fundamentally from the vortices in an AFM with easy-axis anisotropy. Our qualitative analysis of the results obtained shows that including the easy-axis anisotropy modifies the two-dimensional distribution of the antiferromagnetic vector for vortex structures typical of the isotropic AFM model and causes their modulation in the third direction.

(5) Returning to the topological charge, note that Khodenkov [3] discussed the possibility of the existence of vortices in an AFM with half-integer values of $\nu_k = n_k + 1/2$, where $n_k = 0, \pm1, \pm2, \ldots$. At such ν_k, there must be a geometrical boundary that separates the regions with $\mathbf{l}$ and $-\mathbf{l}$. Since the AFM states with $\mathbf{l}$ and $-\mathbf{l}$ are totally equivalent, the nonuniqueness of the vector $\mathbf{l}$ is believed to be unrelated to any discontinuity in spin density and shows that peculiar nonuniform distributions of the vectors $\mathbf{M}_1$ and $\mathbf{M}_2$ can exist. This is valid in the macroscopic theory based on the requirement that the magnetic sublattices $\mathbf{M}_1$ and $\mathbf{M}_2$ be indistinguishable [5]. At the same time, as we show in the next subsection, the macroscopic continual approximation is not self-sufficient. For example, when the energy characteristics of the vortex structure are determined, one has to impose constraints on the sizes of the domain of integration and to use the crystal cell parameter. In this case, the assertion that the magnetic sublattices are indistinguishable becomes incorrect. In addition, although the $\mathbf{l}$ and $-\mathbf{l}$ states are equivalent in energy terms and with respect to the external fields, this equivalency may turn out to break at the banks of the geometrical boundary where there is a contact between the

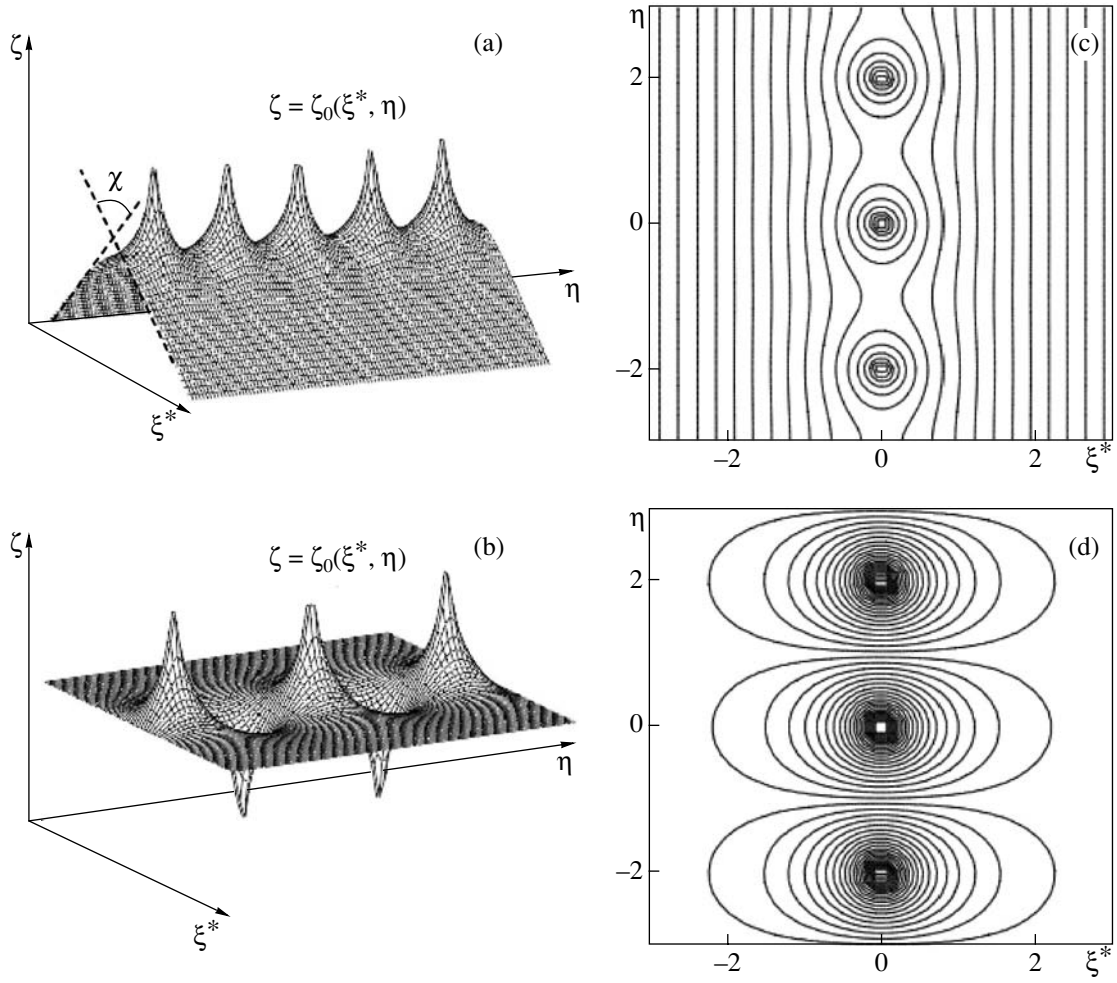

Fig. 3. Images of the surfaces of the AFM DB with chains of vortices with (a) identical and (b) alternating signs of the topological charge arranged periodically along the η axis (the structures are described by Eqs. (27), (28), and (29), respectively) and the isolines formed by the sections of these surfaces by planes perpendicular to the ζ axis.

regions with **l** and –**l** and a layer with ferromagnetic coupling can be formed.

The discussion of this question is beyond the scope of our study. We only emphasize that a soliton configuration with an integer topological charge corresponds to the continuity condition for both the antiferromagnetic vector and the sublattice magnetization vectors, $\mathbf{M}_1$ and $\mathbf{M}_2$. Therefore, below, we discuss AFM configurations with $\nu_k = 0, \pm 1, \pm 2, \ldots$.

(6) Let us consider the influence of dissipative properties of the medium on the dynamics of vortex structures. We take into account the dissipation by means of the dissipative function whose density in the simples form (equivalent to the Hilbert–Landau form for ferromagnets) for the angular variables is proportional to

$$\frac{1}{2\tau}\left[\left(\frac{\partial\theta}{\partial t}\right)^2 + \left(\frac{\partial\varphi}{\partial t}\right)^2 \sin^2\theta\right],$$

where τ is a phenomenological parameter that characterizes the relaxation time of the system.

Given the dissipative terms, the equations of motion (6) take the form

$$-\frac{\partial}{\partial t}(\dot{\varphi} - \omega_H)\sin^2\theta - \frac{1}{\tau}\dot{\varphi}\sin^2\theta + c^2\mathrm{div}[(\nabla\varphi)\sin^2\theta] = 0,$$

$$\ddot{\theta} + \frac{1}{\tau}\dot{\theta} - c^2\nabla^2\theta + [\omega_0^2 + c^2(\nabla\varphi)^2 \tag{31}$$

$$-(\dot{\varphi} - \omega_H)^2]\sin\theta\cos\theta = 0.$$

For simplicity, we exclude the magnetic field from our analysis by assuming that $\omega_H = 0$ and consider the motion of the family of vortices on the surface of the DB at rest by including the dissipation in the nonrelativistic limit $V^2/c^2 \longrightarrow 0$. It is easy to show that the sys-

tem of equations (31) in this case has the solution

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left(\pm\frac{\zeta - \zeta_0(\boldsymbol{\rho}, t)}{\delta}\right),$$
$$\varphi(\boldsymbol{\rho}, t) = \sum_k^N \nu_k \arctan\left(\frac{\eta - \eta_k}{\xi - \xi_k(t)}\right) + C_2, \quad (32)$$

where

$$\zeta_0(\boldsymbol{c}, t) = \delta\sum_k^N \nu_k \ln\left(\frac{1}{\delta}\sqrt{(\eta - \eta_k)^2 + [\xi - \xi_k(t)]^2}\right) + C_1,$$

$\xi_k(t) = \xi_k(0) + v_0\tau(1 - e^{-t/\tau})$ is the coordinate of the vortex core, and v_0 is the initial velocity.

Thus, the motion of the core of an antiferromagnetic vortex is similar in pattern to the motion of a particle in a dissipative medium. The velocity decreases exponentially. In the travel time, the vortices displace by the distance $v_0\tau$.

3. THE INTERACTION BETWEEN ANTIFERROMAGNETIC VORTICES. INSTABILITY OF THE VORTEX STATE

Let us turn to the interaction between antiferromagnetic vortices (18) and (23) and to the stability of such states. For simplicity, we will consider configuration (23) in the static case ($v = 0$, $\omega_H = 0$).

If we write the volume integral of the energy density (1) for AFM in the static case for configuration (23) and perform integration over the variable ζ, then, given condition (22), we will obtain the following result:

$$U = U_0 + \varepsilon\int\int d\xi d\eta (\nabla_\perp \zeta_0)^2, \quad (33)$$

where

$$\nabla_\perp = \mathbf{e}_1\frac{\partial}{\partial\xi} + \mathbf{e}_2\frac{\partial}{\partial\eta}, \quad \varepsilon = 8\sqrt{\alpha_1\beta_1}M_0^2,$$

ε is the energy per unit surface area of the 180-degree AFM DB, $U_0 = -2\beta_1 M_0^2 V + \varepsilon S$ is the energy of the AFM with the DB, and S is the cross-sectional area of the AFM sample in the $\xi\eta$ plane; the second term in (33) represents the intrinsic energy of the system of vortices.

We perform integration in the $\xi\eta$ plane over a circular area containing the family of antiferromagnetic vortices. We assume that the center of the circle lies within the region where the antiferromagnetic vortex cores are concentrated and that its radius R is much larger than the size of this region ($R \gg |\xi_k|, |\eta_k|$). In what follows, it is necessary to pass to the limit by letting R approach the AFM boundaries.

Through integration by parts and based on the divergence theorem, expression (34) for the energy of the vortex system can be transformed into

$$U - U_0 = \varepsilon\oint dl(\zeta_0\nabla_\perp\zeta_0)_n - \varepsilon\int\int d\xi d\eta \zeta_0\nabla_\perp^2\zeta_0. \quad (34)$$

The first integral in (34) is calculated over a circular path with a large radius R, dl is the element of length of the path. The subscript n means that the projection of the vector $\zeta_0\nabla_\perp\zeta_0$ onto the normal to the element of length of the path is taken in the integrand.

Following the distribution of the vector **l** far from the vortex structure, we note that it coincides with the distribution for one vortex with a charge equal to the total charge of the system. Indeed, given the condition $R \gg |\xi_k|, |\eta_k|$, expression (23) can be represented as

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left(\pm\frac{\zeta - \zeta_0(\xi, \eta)}{\delta}\right),$$
$$\varphi(\boldsymbol{\rho}, t) = \sum_k \nu_k \arctan\left(\frac{\eta}{\xi}\right) + C_2, \quad (35)$$

where

$$\zeta_0(\xi, \eta) = \delta\sum_k \nu_k \ln\left(\frac{1}{\delta}\sqrt{\eta^2 + \xi^2}\right) + C_1,$$

whereupon the integral in question can be easily calculated. Calculating the second integral in (34) causes no difficulty either, since

$$\nabla_\perp^2\zeta_0 = 2\pi\sum_k \nu_k\delta(\xi - \xi_k)\delta(\eta - \eta_k).$$

Finally, we obtain

$$U - U_0 = E\left(\sum_k \nu_k\right)^2 \ln\left(\frac{R}{\delta}\right) - E\sum_{i,k}\nu_i\nu_k\ln\left(\frac{R_{ik}}{\delta}\right), \quad (36)$$

where

$$E = 2\pi\delta^2\varepsilon, \quad R_{ik} = \sqrt{(\xi_i - \xi_k)^2 + (\eta_i - \eta_k)^2},$$

R_{ik} is the separation between the vortex axes.

At $i \neq k$, the second term in (36) defines the vortex interaction energy. Clearly, this energy does not depend on the influence of the remaining neighbors. The force of interaction has the same form as that for point charges in a plane:

$$\mathbf{F}_{ik} = -\frac{\partial U}{\partial \mathbf{R}_{ik}} = 2E\nu_i\nu_k\frac{\mathbf{R}_{ik}}{R_{ik}^2}. \quad (37)$$

Accordingly, the vortices repel and attract each other for like and unlike topological charges, respectively.

Expression (36) for the energy of the vortex system shows that the continual approximation of the macroscopic theory, according to which the intrinsic energy of an individual vortex is an unbounded quantity, since $R_{ik} = 0$ at $i = k$, is limited. This explains the stationary configurations (18) and (23) that exist despite the interaction between individual vortices. Indeed, since the intrinsic energy in the continual approximation, as well as the effective vortex mass, is infinitely high, the influence of the finite (in magnitude) interactions between the vortices cannot change their relative positions.

The situation will change if the constraints determined by the period of the crystal structure are imposed on the diameter of the vortex core. Therefore, when writing R_{ii} in (36), we assume that $R_{ii} = aK$, where K is an integer. The uncertainty related to the introduction of parameter K for materials with $\delta/a \gg K$ is of no fundamental importance, since it leads only to relatively small energy corrections that do not depend on the sample sizes and the vortex arrangement. Consequently, the parameter K can generally be excluded from our analysis.

Dividing the system's energy (36) by the intrinsic energy of the vortices and the interaction energy, we write

$$U - U_0 = E\left(\sum_k \nu_k\right)^2 \ln\left(\frac{R}{a}\right) - E\sum_{i,k} \nu_i \nu_k \ln\left(\frac{R_{ik}}{a}\right). \quad (38)$$

It is easy to see that the energy of the family of vortices with a zero total charge is always bounded. If the total topological charge of the system is nonzero, $\sum_k \nu_k \neq 0$, then the vortex energy is determined by the AFM sizes and increases infinitely as $R \longrightarrow \infty$, just as the disclination energy in AFM [5].

In conclusion, let us compare the potential energy $U(\nu_1, \nu_2)$ of a pair of vortices with charges ν_1 and ν_2 with the energy $U(\nu)$ of one vortex whose charge is equal to the total charge $\nu = \nu_1 + \nu_2$ of this pair. As follows from relation (36), the difference between these energies is

$$\Delta U = U(\nu) - U(\nu_1, \nu_2) = 2E\nu_1\nu_2 \ln(R_{12}/a). \quad (39)$$

This result indicates that the vortex is unstable against its decay into components with an elementary charge. At the same time, if the neighboring vortices have unlike charges, their merger is energetically favorable.

Thus, in actual physical systems, the stable existence of vortex structures is possible only in the case of artificial nucleation and stabilization. For example, vortices of this type can be nucleated in easy-axis ferromagnetic systems near ferromagnetic dots with a small diameter (much less than the DB thickness parameter δ) in a ferromagnetic medium if there is a magnetic exchange interaction between the dot material and the medium.

4. CONCLUSIONS

Note that all of the configurations considered are a special case of a more general distribution of the antiferromagnetic vector:

$$\tan\frac{\theta(\boldsymbol{\rho}, t)}{2} = \exp\left(\pm\frac{\zeta - \zeta_0(u, v)}{\delta}\right),$$
$$\varphi(\boldsymbol{\rho}, t) = \sum_k \nu_k \arctan\left(\frac{u - u_k}{v - v_k}\right) + \omega_H t + C_2, \quad (40)$$

where

$$\zeta_0(u, v) = \delta\sum_k \nu_k \ln\left(\sqrt{(u - u_k)^2 + (v - v_k)^2}\right) + C_1,$$

$u(\xi, \eta)$, $v(\xi, \eta)$ are any conjugate harmonic functions; u_k, v_k are arbitrary constants. Note that expression (16) also solves the Landau–Lifshitz equations in the FM model with easy-axis anisotropy.

ACKNOWLEDGMENTS

We are grateful to V.G. Bar'yakhtar and Yu.I. Gorobets for support and fruitful discussions.

REFERENCES

1. A. A. Belavin and A. M. Polyakov, Pis'ma Zh. Éksp. Teor. Fiz. **22**, 503 (1975) [JETP Lett. **22**, 245 (1975)].
2. G. E. Khodenkov, Fiz. Met. Metalloved. **54**, 644 (1982).
3. H. E. Khodenkov, Phys. Status Solidi B **127**, K121 (1985).
4. V. G. Bar'yakhtar and B. A. Ivanov, Fiz. Nizk. Temp. **5**, 759 (1979) [Sov. J. Low Temp. Phys. **5**, 361 (1979)].
5. A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, *Nonlinear Waves of Magnetization: Dynamical and Topological Solitons* (Naukova Dumka, Kiev, 1983) [in Russian].
6. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series. Elementary Functions* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 568–575.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 643–651.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Multilayer Structures of the Iron/Chromium Type with Almost Ideal Interfaces in an External Magnetic Field

V. N. Men'shov and V. V. Tugushev

Russian Research Center Kurchatov Institute, pl. Akademika Kurchatova 1, Moscow, 123182 Russia

e-mail: vnmenshov@mail.ru; vvtugushev@mail.ru

Received July 21, 2004

Abstract—The magnetic phase diagram of the Fe/Cr/Fe three-layer structure with almost ideal interlayer boundaries was constructed. The effective interlayer interaction in this structure was described by the "half-angle coupling" model. Various system configurations were analyzed taking into account crystalline anisotropy, and the ground state of the system was determined. The behavior of the structure in an external magnetic field applied along easy and hard magnetic axes was studied. The magnetization curves $M(H)$ characteristic of structures with various interface roughness parameter and interlayer exchange values were described and analyzed. The experimental situation is discussed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Interest in multilayer structures of the Fe/Cr type has not lessened for almost two decades and has been continuously stirred up by new experimental results. For these structures, such striking effects as giant magnetoresistance, the long- and short-period oscillations of the effective interaction of ferromagnetic Fe layers depending on the thickness of antiferromagnetic Cr spacers between them, the phase slip of this interaction, the noncollinear magnetization configurations of neighboring ferromagnetic layers, etc. (in particular, see reviews [1–4]), have been observed for the first time. Studies of the special features of the magnetic coupling of ferromagnetic layers in the structures under consideration allow conclusions to be drawn on two factors that play a determining role. First, this is specific spin ordering, called spin density wave, in the antiferromagnetic chromium interlayer [3]. The second factor is inevitable roughness of boundaries between iron and chromium layers determined by the technological conditions of multilayer structure growth [1, 5]. These two factors were simultaneously taken into account at a qualitative level in the "bilinear-biquadratic" model [6], which presupposed the division of the Fe layers into ferromagnetic domains, the absence of deformation of the antiferromagnetic structure in the Cr spacer, and stiff exchange coupling at the Fe/Cr interface. This model gives the dependence of the effective exchange energy $E(\psi)$ on the angle ψ between the magnetizations of neighboring ferromagnetic layers in the form of the sum of terms proportional to $\cos\psi$ and $\cos^2\psi$, along with the coefficients that depend on the temperature and spacer thickness. Remarkably, the estimates obtained in [7] on the assumption different from that made in [6], that is, on the assumption of a weak spin structure deformation in the antiferromagnetic spacer and uniformly magnetized ferromagnetic plates, gave the $E(\psi)$ functional dependence of magnetic coupling energy similar in form to that reported in [6].

Note that, for the first time, the term proportional to $\cos^2\psi$ was introduced into the interlayer interaction energy purely phenomenologically, to explain a noncollinear magnetic configuration of the Fe/Cr/Fe three-multilayer structure [8]. Since then, the model suggested in [6] has been used in virtually all experimental works to interpret the data on the magnetic properties of Fe/Cr-type systems. Sometimes, this was done fairly successfully [4, 7]. Nevertheless, many facts do not fit such interpretation. This is especially true of structures prepared using optimum technologies and having high-quality interlayer boundaries. For instance, the data on the reflection of polarized neutrons and the magnetooptical Kerr effect obtained in [9] for [Fe/Cr](001) superlattices grown at 250°C, with an estimated noncorrelated interface roughness of order 10 nm, led the authors to seek an explanation of the behavior of interlayer coupling in terms of the "proximity magnetism" model [10], different in principle from that suggested in [6]. This model is, however, more suitable for describing systems with strong exchange interaction between local moments at the ferromagnet/antiferromagnet boundary such as Fe/Mn structures [11] rather than Fe/Cr-type systems in which this interaction is weak [12, 13] and spin density is strongly delocalized in spacers.

The recent data obtained in [14] in studying Fe/Cr/Fe(001) sandwiches by Kerr magnetometry and Mandelstam–Brillouin light scattering show that the character of interlayer interaction changes substantially depending on the structure and quality of the Fe/Cr

interface. According to [14], the experimental data on structures with weakly rough interfaces can be interpreted more correctly using the proximity magnetism model [10] or the so-called half-angle coupling model than the model suggested in [6]. The half-angle coupling model will be considered in more detail in what follows.

The inhomogeneous antiferromagnetic state of the type of spin density wave formed in Cr spacers over a broad range of thickness and temperature values is extremely sensitive to the quality and geometry of the interface. A self-consistent description of this state (charge-induced spin density wave) was suggested in [12, 13]; it was based on the theory of a localized spin density wave [15] suggested more than 15 years ago to describe the properties of dilute chromium-based alloys. The approach developed in [12, 13] allowed several unusual properties of Fe/Cr-type magnetic nanostructures to be explained. First and foremost, this relates to the description of phase diagrams (temperature–spacer thickness) for superlattices of various orientations and compositions ([Fe/Cr](001) [1–3], [Fe/$Cr_{1-x}Fe_x$](001) [16], [Fe/Cr](011) [17], and [$Cr_{1-x}Mn_x$/Cr](001) [18]). The widely known phenomenon of the phase slip of effective interlayer interaction in a whisker-grown Fe/Cr/Fe(001) three-layer structure with a wedge-like spacer [1–3] and the recently observed phenomenon of the rearrangement of the structure of spin density wave caused by the introduction of δ-doping submonolayers of some metals [19] also fit this scheme well.

A new mechanism of interlayer magnetic coupling was suggested in [20–22] for the Fe/Cr/Fe(001) three-layer structure with rough interlayer boundaries and thick iron layers. In this mechanism, the antiferromagnetic spacer acquires substantial exchange stiffness at high (compared with the Néel point T_N) temperatures because of the formation of a charge-induced spin density wave. The approach suggested in [21, 22] was used to substantiate the existence of noncollinear states and obtain an equation for the energy $E(\psi)$ whose value and form substantially depended on the quality of Fe/Cr interfaces. The $E(\psi)$ function takes the traditional form of bilinear-biquadratic interaction only in the limit of a high concentration of monatomic steps (we will use the term "monostep") at interfaces. In the opposite situation, when the relief of the Fe/Cr boundary surface is modulated by broad terraces, the $E(\psi)$ analytic dependence obtained in [21, 22] has no phenomenological analog.

In this work, we study the behavior of the magnetization **M** of a Fe/Cr/Fe(001) three-layer structure in an external magnetic field **H**. We confine our consideration to a structure with a low concentration of monosteps at the boundaries between the antiferromagnetic spacer and ferromagnetic plates, which are assumed to be sufficiently thick for considering them uniformly magnetized. Precisely in this (of course idealized) situation do the special features of the state with charge-induced spin density wave and of the **M**(**H**) dependence for the Fe/Cr structure with an almost ideal interface explicitly manifest themselves. The opposite situation with a high concentration of monosteps at interfaces is characteristic of systems with a bilinear-biquadratic interlayer coupling in an external field, which have repeatedly been analyzed theoretically [4].

2. THE HALF-ANGLE COUPLING MODEL FOR INTERLAYER MAGNETIC COUPLING IN A Fe/Cr/Fe THREE-LAYER STRUCTURE

Recall the main characteristics of the model [22]. The temperature range $T_N < T < T_0$ corresponding to short-range antiferromagnetic order in the chromium spacer is considered; here, $T_0 \ll T_C$, where T_C is the Curie temperature of the iron plates. The thickness of the ferromagnetic layers is assumed to be sufficiently large, and the magnetization density **S** within the ferromagnetic layers can be considered uniform and temperature-independent at $T_0 \ll T_C$. At the same time, the thickness L of the antiferromagnetic spacer can vary over a fairly wide range provided the condition $L > 2\xi_0$, where ξ_0 is the spin density wave coherence length, is satisfied. The magnetization density of the antiferromagnetic sublattices $\boldsymbol{\sigma}(\mathbf{r})$ can be very inhomogeneous and strongly depend on temperature over the temperature range under consideration, [13]. A detailed description of the system with the Ginzburg–Landau expansion of its thermodynamic potential is presented in [12, 13, 20, 21, 22] in terms of the antiferromagnetic order parameter, that is, the spin density wave amplitude in the spacer $\boldsymbol{\Delta}(\mathbf{r}) = U\boldsymbol{\sigma}(\mathbf{r})$, where U is the effective spin density wave potential. The reason why $\boldsymbol{\Delta}(\mathbf{r})$ is nonzero above the Néel temperature of bulk chromium is an increase in the electronic spin susceptibility close to the interface because of charge flowing between the Fe and Cr layers. This effect can be formally described as a local increase in the coupling constant U, which determines the characteristic temperature $T_0(L)$ of the appearance of short-range antiferromagnetic order in the spacer on the scale of the amplitude correlation length $\xi(T) \sim \xi_0|1 - T/T_N|^{-1/2}$.

In spite of its comparatively small contribution to the amplitude of spin density wave, the exchange interaction component of order $\mathbf{S}\boldsymbol{\Delta}$ at the Fe/Cr interface to a substantial extent determines the details of the spatial dependence of the $\boldsymbol{\Delta}(\mathbf{r})$ vector and its orientation with respect to ferromagnetic layer magnetization. The role played by the exchange component becomes especially important for a multilayer structure with nonideally smooth (rough) interfaces, because charge and exchange potential fluctuations are inevitable near such surfaces. Unlike the Coulomb interaction, the exchange contribution to the surface energy sharply changes sign as the thickness of the spacer decreases by as little as one chromium monolayer. Such a dependence of the

energy on whether the number N of monolayers is even or odd allows long-wave spacer thickness fluctuations for a three-layer structure with rough surfaces to be modeled in a fairly simple way [22]. The equilibrium magnetic configuration of the system, namely, the spin density wave structure $\mathbf{\Delta}(\mathbf{r})$ and the angle ψ between the moments of the ferromagnetic layers, is determined by balance between the surface exchange energy at the Fe/Cr interface and the energy of a nonuniform spin density wave deformation within the spacer. For a thin spacer, $L \ll D$, where D is the interpolation length [13], the spin density wave amplitude $\mathbf{\Delta}(\mathbf{r})$ is almost constant, and the problem reduces to considering static spin density wave orientation fluctuations. This problem was solved exactly by representing the spacer as a set of fragments with even and odd numbers of chromium monolayers N and lengths l_e and l_o, respectively. These fragments are taken to alternate periodically along one of the directions in the Fe/Cr interface plane. The characteristic scale of vector $\mathbf{\Delta}(\mathbf{r})$ rotations caused by exchange potential jumps at monosteps is the angle correlation length $\zeta(L, T) \sim (\Delta L)^{1/2}$.

A general equation for the interlayer coupling energy $E(\psi)$ was obtained in [22]. In the limit of strong roughness ($\rho\zeta \gg 1$), where $\rho = 2(l_e + l_o)^{-1}$ is the density of monosteps, the $E(\psi)$ dependence has the form

$$E(\psi) = A_1\cos\psi + A_2\cos^2\psi. \quad (1)$$

In the weak roughness limit ($\rho\zeta \ll 1$), that is, for the interface with broad terraces, we have

$$E(\psi) = B_1\cos\left(\frac{\psi}{2}\right) + B_2\sin\left(\frac{\psi}{2}\right). \quad (2)$$

Here, $A_{1,2}$ and $B_{1,2}$ are the coefficients depending on the temperature T, spacer thickness L, and density of monosteps ρ and also on the interface roughness parameter $b = l_o/l_e$. The magnetic phase diagram of the model Fe/Cr/Fe three-layer structure with bilinear-biquadratic interlayer coupling (Eq. (1)) was constructed in [23]. A similar problem for the half-angle coupling model (Eq. (2)) has not been studied and will be solved in this work.

Let us consider a symmetrical Fe/Cr/Fe(100) three-layer structure (called trilayer below) taking into account the fourfold crystallographic axes in the iron layers and assuming that the external magnetic field $\mathbf{H}$ lies in the interface plane (y, z). We also assume that the magnetization $\mathbf{M}$ is uniformly distributed over the volume of each ferromagnetic plate, and its amplitude is independent of the external field. The magnetic moment of each plate rotates as a whole in the (y, z) plane, because the demagnetizing field is much stronger than the anisotropy field K. The energy of the magnetic structure under consideration per unit area of its cross section can be written as

$$E = -dHM(\cos\varphi_1 + \cos\varphi_2)$$
$$-\frac{dKM}{16}(\cos 4(\varphi_1 + \alpha) + \cos 4(\varphi_2 + \alpha)) \quad (3)$$
$$+ B_1\cos\left(\frac{\varphi_1 - \varphi_2}{2}\right) + B_2\sin\left(\frac{|\varphi_1 - \varphi_2|}{2}\right).$$

The meaning of the terms in (3) is obvious. Here, d is the thickness of the ferromagnetic layer, α is the angle between the easy axis $\mathbf{n}_z$ and external field $\mathbf{H}$, and $\varphi_{1,2}$ is the angle between the vector $\mathbf{H}$ and magnetic moment $\mathbf{M}_{1,2}$ of the right or left plate. Below, we only consider two variants, when the external field $\mathbf{H}$ is applied either along the easy ($\alpha = 0$) or hard ($\alpha = \pi/4$) magnetic axis of the ferromagnetic layers.

Let us introduce the dimensionless values

$$\varepsilon = \frac{E}{dKM}, \quad h = \frac{H}{k},$$
$$I = \frac{a}{dKM}, \quad \Lambda = \frac{b}{1+b} = \frac{l_o}{l_o + l_e}, \quad (4)$$

where a, $b > 0$, $a = -(B_1 + B_2)$, $b = B_1/B_2$, and I is the dimensionless exchange coupling amplitude between the iron and chromium layers. For an ideal Fe/Cr(100) interface, we would have $\Lambda = 0$ or 1. It follows that the Λ parameter ($0 < \Lambda < 1$) determines the degree of interface roughness. Let us use the new variables

$$m = \cos\left(\frac{\varphi_1 - \varphi_2}{2}\right), \quad n = \cos\left(\frac{\varphi_1 + \varphi_2}{2}\right). \quad (5)$$

They describe the polarization and orientation, respectively, contributions to the applied field dependence of the total trilayer magnetization $\mu = nm$ (measured in saturation magnetization $2M$ units). In the new denotations, we have

$$\varepsilon(m, n) = -2hmn \mp \frac{1}{8}f(m)f(n)$$
$$-I(\Lambda m + (1-\Lambda)\sqrt{1-m^2}), \quad (6)$$
$$f(p) = 8p^4 - 8p^2 + 1, \quad p = (n, m),$$

where minus corresponds to the angle $\alpha = 0$ and plus to $\alpha = \pi/4$.

Let us consider the ground state of the system (m_0, n_0) at $H = 0$. In all cases except two trivial situations with $\Lambda = 0$ and $\Lambda = 1$, functional (6) is minimum for noncollinear configurations of the plate magnetic

moments ($m_0 \neq 0, 1$), which can be oriented strictly symmetrically with respect to the easy ($n_0 = 0, 1$) or hard ($n_0 = 1/\sqrt{2}$) magnetic axis. These configurations will be denoted by NCE and NCH, respectively. Irrespective of the interlayer coupling I value, energy equilibrium between them is attained if $\Lambda = 1/\sqrt{2}$ or $\Lambda = 1 - 1/\sqrt{2}$. In the interval $1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}$, the ground state has the NCH configuration, and in the $0 < \Lambda < 1 - 1/\sqrt{2}$ and $1/\sqrt{2} < \Lambda < 1$ intervals, its configuration is NCE. The boundaries of the absolute and relative instability of the magnetic states of the trilayer are shown in Fig. 1 in the (I, Λ) coordinates. The same figure contains some curves corresponding to a constant ψ_0 angle between the $\mathbf{M}_1$ and $\mathbf{M}_2$ vectors. It follows that the magnetic state of a Fe/Cr/Fe-type multilayer structure is very sensitive to the integral roughness parameter Λ of interlayer boundaries and changes jumpwise at $\Lambda = 1/\sqrt{2}$ and $\Lambda = 1 - 1/\sqrt{2}$.

The equation that explicitly describes the equilibrium magnetization configuration for weak interlayer coupling $I \ll 1$ has the form

$$n_0 = 0, 1, \quad m_0 = I\Lambda/2, \quad 0 < \Lambda < 1 - 1/\sqrt{2},$$

$$n_0 = \frac{1}{\sqrt{2}}, \quad m_0 = \frac{1}{\sqrt{2}} + \frac{I(2\Lambda - 1)}{4},$$

$$1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}, \qquad (7)$$

$$n_0 = 0, 1, \quad m_0 = 1 - \frac{I^2(1-\Lambda)^2}{8},$$

$$1/\sqrt{2} < \Lambda < 1.$$

If interlayer coupling is strong ($I \gg 1$), we have

$$m_0 = \frac{\Lambda}{(\Lambda^2 + (1-\Lambda)^2)^{1/2}} \pm \frac{2}{I}\frac{\Lambda(1-\Lambda)^2(2\Lambda - 1)}{(\Lambda^2 + (1-\Lambda)^2)^3}, \quad (8)$$

where "+" is used if $0 < \Lambda < 1 - 1/\sqrt{2}$ or $1/\sqrt{2} < \Lambda < 1$ and "–," if $1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}$. The square of the equilibrium $m = m_0$ value as a function of the Λ parameter is shown in Fig. 2 for some intermediate I values $I \sim 1$.

A characteristic feature of our model with almost ideal plane interfaces is the absence of collinear states in the zeroth order with respect to the $(\rho\zeta) \ll 1$ parameter. It could be shown that a region of collinear states appears on the (Λ, I) phase plane as surface roughness increases, and this region broadens as the $(\rho\zeta)$ parameter grows.

3. THE BEHAVIOR OF THE TRILAYER IN A MAGNETIC FIELD APPLIED ALONG THE EASY MAGNETIC AXIS

Because of the presence of not only the ground but also metastable states in the structure under consideration at $\mathbf{H} = 0$, orientation transitions between various

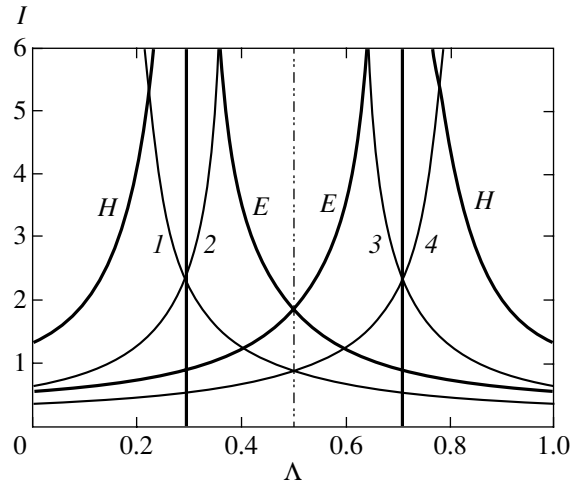

Fig. 1. Phase diagram of the Fe/Cr/Fe(100) three-layer structure in the (I, Λ) coordinates. The vertical lines $\Lambda = 1/\sqrt{2}$ and $\Lambda = 1 - 1/\sqrt{2}$ are the boundaries between the NCH and NCE states. The lines H and E are the boundaries of the absolute stability of the NCH and NCE states, respectively. The numbered curves are the states with constant ψ_0 angle values: (*1*) $7\pi/8$, (*2*) $5\pi/8$, (*3*) $3\pi/8$, and (*4*) $\pi/8$.

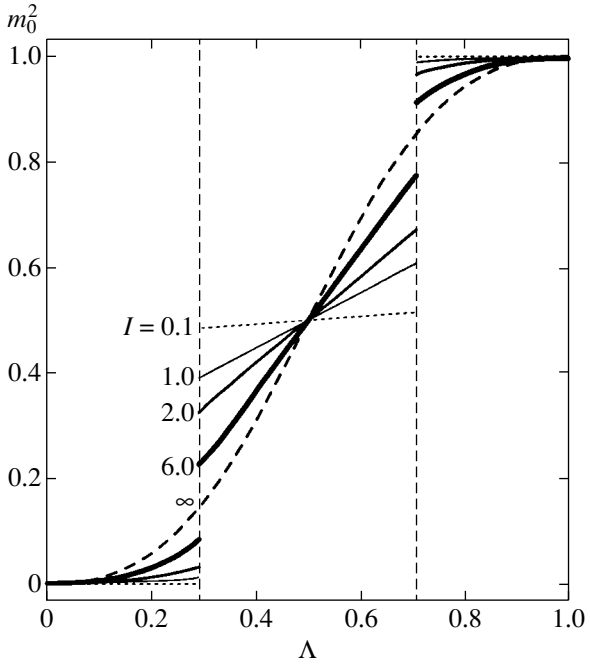

Fig. 2. Dependence $m_0^2(\Lambda)$ at various interlayer coupling parameter values: $I = 0.1$, 1.0, 2.0, 6.0, and ∞.

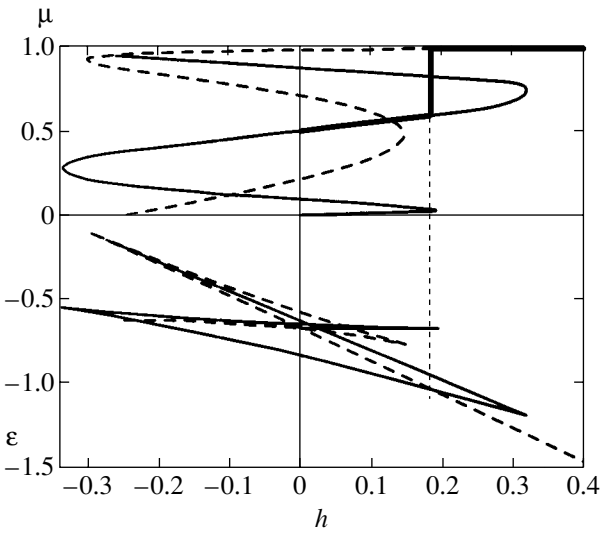

Fig. 3. Magnetization curves of a three-layer structure with the parameters $\Lambda = 1/2$ and $I = 1$ in symmetrical $\mu_S(h)$ (dashed line) and asymmetric $\mu_N(h)$ (solid line) configurations in a field applied along the easy axis. The energies corresponding to these magnetic configurations $\varepsilon_S(h)$ and $\varepsilon_N(h)$ are shown at the bottom. At the $h^* = 0.182$ point, an orientation transition occurs.

magnetic configurations can occur when an external field is switched on. Let us first consider a system characterized by the Λ parameter from the interval $1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}$ in field **H** applied along the easy magnetic axis ($\alpha = 0$). The transition from the NCH state with $n_0 = 1/\sqrt{2}$ to the NCE state with $n_0 = 1$ occurs as the field increases through some intermediate configuration of plate moments $M_{1,2}$ canted with respect to the anisotropy field; that is, $n_0 \neq 0, 1, 1/\sqrt{2}$. This asymmetric state can only be described analytically in the limit of $|\Lambda - 1/2| \ll 1$,

$$n_0 = \frac{1}{\sqrt{2}}\left(1 + \frac{h}{2}\right),$$
$$m_0 = \frac{1}{\sqrt{2}}\left(1 + \frac{\sqrt{2}h + 2\Lambda - 1}{I + 2\sqrt{2}}\right). \quad (9)$$

It exists in the interval $0 < h < h^* \ll 1$. In the weak coupling approximation ($I \ll 1$), we obtain

$$h^* = I\left(\frac{1}{\sqrt{2}} - \Lambda\right).$$

When the field value exceeds the critical value h^*, an NCE-type state symmetrical with respect to the vector **H** becomes more energetically favorable,

$$n_1 = 1, \quad m_0 = 1 - \frac{I(1-\Lambda)}{2(2(1+h) + I\Lambda)}, \quad (10)$$
$$1 - m^2 \ll 1.$$

In a high external field, (10) describes gradual trilayer magnetization saturation. Formally, the asymptotic transition to the ferromagnetic state as $h \longrightarrow \infty$ and the occurrence of a sharp orientation transition at the h^* point make the behavior of the Fe/Cr structure with an almost ideal interlayer boundary in an external magnetic field substantially different from that of a structure with fairly rough boundaries [24, 25].

In order to obtain a general picture of the behavior of our system in an applied field, we numerically studied functional (6) over wide ranges of I and Λ variations. The field dependence of the total magnetization $\mu(h)$ of the three-layer structure with the parameters $I = 1$ and $\Lambda = 1/2$ is shown in Fig. 3. The $\mu_N(h)$ curve, which meets the extremum condition for two-parameter functional (6),

$$\frac{\partial\varepsilon(m, n)}{\partial m} = \frac{\partial\varepsilon(m, n)}{\partial n} = 0,$$

corresponds to the trilayer magnetic configuration asymmetric with respect to the external field direction, whereas the $\mu_S(h)$ curve obtained from the equation $\partial\varepsilon(m, 1)/\partial m = 0$ corresponds to the NCE configuration. The $\varepsilon_N(h)$ and $\varepsilon_S(h)$ energies of these configurations are shown at the bottom of Fig. 3. This picture is typical of a system that has the NCH state at $h = 0$; that is, it has Λ in the range $1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}$. The $\mu_N(h)$ and $\mu_S(h)$ functions are not single-valued, and only those of their branches along which magnetization increases with the field are stable. It follows that the sequence of the magnetization of a three-layer structure is as follows. First, in the interval $0 < h < h^*$, magnetization increases under asymmetric configuration conditions from the remanent magnetization $\mu_R = \mu_N(h = 0)$ to $\mu_N(h^*)$. Next, at the $h = h^*$ point, where $\varepsilon_N(h) = \varepsilon_S(h)$, the system switches to symmetrical NCE configuration conditions, and magnetization jumpwise reaches a value of $\mu_S(h^*)$ and gradually attains saturation ($1 - \mu \sim h^{-1} \longrightarrow 0$) as the field increases in the interval $h^* < h < \infty$. Figure 2 in essence illustrates changes in the remanent magnetization μ_R (recall that $\mu = mn$) as a function of I and Λ.

The behavior of the structure with Λ from the interval $0 < \Lambda < 1 - 1/\sqrt{2}$ in a field applied along the easy axis also attracts attention. At $H = 0$, this structure is in the NCE state with almost antiparallel M_1 and M_2 moments ($135° < \psi_0 < 180°$). In a weak field ($0 < h < h_1^*$), the NCE state is retained; that is, the plate moments are symmetrical with respect to the direction

of H, and the magnetization of the trilayer increases linearly as h grows, $\mu = \mu_R + h$, where $\mu_R = I\Lambda/2$. At the point $h_1^* = I(1 - 1/\sqrt{2} - \Lambda)$, the system switches to an asymmetric state, in which

$$n_0 = \frac{1}{\sqrt{2}}\left(1 + \frac{h}{2}\right),$$
$$m_0 = \frac{1}{\sqrt{2}} + \frac{\sqrt{2}h + I(2\Lambda - 1)}{4}. \quad (11)$$

When the field increases to $h_2^* = I(1/\sqrt{2} - \Lambda)$, the system, however, returns to the symmetrical NCE state and then reaches saturation as $h \longrightarrow \infty$,

$$1 - \mu \sim h^{-1} \longrightarrow 0.$$

To summarize, two first-order orientation transitions sequentially occur in the system as the applied field strength increases. A similar phenomenon was observed experimentally [26].

4. THE BEHAVIOR OF THE TRILAYER IN A FIELD APPLIED ALONG THE HARD MAGNETIC AXIS

Let us briefly describe the special features of the behavior of the Fe/Cr/Fe(100) three-layer structure in a magnetic field applied along the hard axis ($\alpha = \pi/4$). An analysis shows that the symmetrical state is more favorable energetically than asymmetric in a system with the integral roughness parameter $1 - 1/\sqrt{2} < \Lambda < 1/\sqrt{2}$ at any field value $h > 0$. In other words, the symmetrical structure of the NCH state ($n = 1$) is retained in an external field. The magnetization $\mu(h) = \mu_S(h)$ monotonically increases from the remanent value μ_R (see Fig. 2) to saturation: $\mu^2 = 1 - I(1 - \Lambda)/2h$ if $h \gg (1, I)$. The $\mu(h)$ dependence for $\Lambda = 1/2$ is shown in Fig. 4.

As has been mentioned, the three-layer structure with the Λ parameter satisfying the condition $1/\sqrt{2} < \Lambda < 1$ is in the NCE state at $h = 0$. A weak magnetic field applied along the hard axis changes the state of this system to asymmetric. The system then experiences a first-order transition at the point $h = h^*(I, \Lambda)$ to the NCH state. Given $I \sim 1$, h^* is of several tenths according to numerical estimates, and the magnetization of the trilayer at the point $h = h^*$ increases by a value of about saturation magnetization.

The structure with Λ from the interval $0 < \Lambda < 1 - 1/\sqrt{2}$ behaves similarly. It also undergoes a first-order phase transition at point $h = h^*$, but the corresponding magnetization jump is then comparatively insignificant.

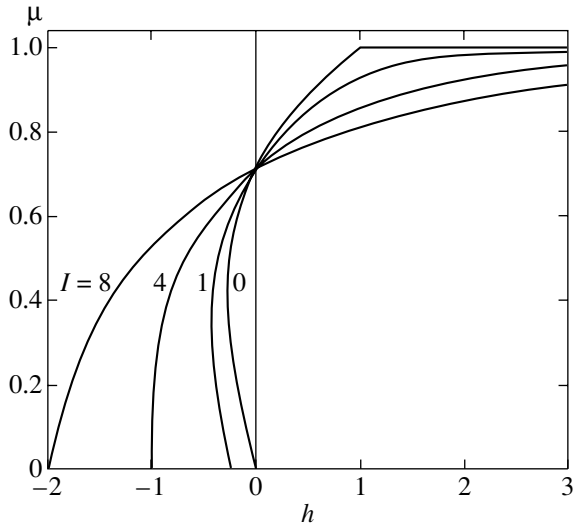

Fig. 4. Field dependences of the magnetization of a three-layer structure with the parameter $\Lambda = 1/2$ at various interlayer coupling values ($I = 0$, 1, 4, and 8) in a field applied along the hard axis.

5. CONCLUSIONS

Let us formulate the principal results of this work.

(1) We constructed the (Λ, I) phase diagram of the Fe/Cr/Fe three-layer structure with effective interlayer interaction described by the half-angle coupling model. Various configurations of the system were analyzed taking into account crystalline anisotropy, and the ground state of the system was found.

(2) The behavior of the structure in a magnetic field applied along either the easy or hard axis was studied. The $M(H)$ magnetization curves were described and analyzed for structures with different Λ and I parameters.

Interlayer interaction in the Fe/Cr/Fe system with a wedgelike spacer and thick (of order 100 Å) plates was studied in detail depending on the quality of interfaces in works [7, 14] already mentioned in the Introduction. A characteristic sharp minimum of the frequency of the optical branch of spin waves in the saturation field and a well-defined linear region, which, with a kink, transforms into saturation along the magnetization curve constructed in the Arrott coordinates, are observed for samples with strongly rough interfaces. These special features are well reproduced by the bilinear-biquadratic exchange model. However, for samples with smoother interfaces, the magnetization curve reaches saturation asymptotically, without a kink, and the magnetic field dependence of the frequency of the optical branch of spin waves has virtually no minimum. Moreover, over the whole range of spacer thick-

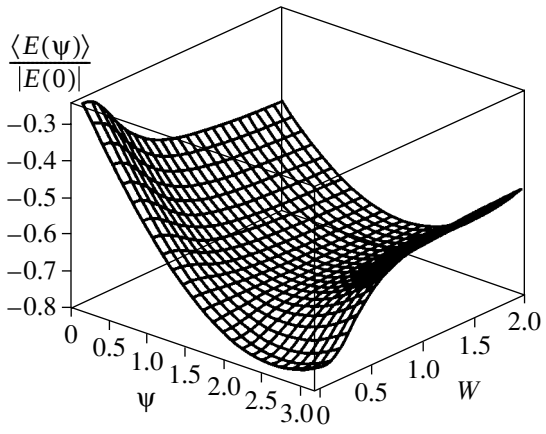

Fig. 5. Dependences $\langle E(\psi)\rangle$ of the energy of a three-layer structure at various variance W values obtained for $\langle\Lambda\rangle = 1/4$ and $I = 1$.

nesses $L = 0$–40 Å and at all measurement temperatures $T = 77$–473 K, the half-angle coupling model describes the experimental magnetization curves $M(H)$ of the samples with smooth interfaces no worse than the proximity magnetism model.

We attach great importance to the studies performed in [5], where the morphology of interlayer boundaries in the Fe/Cr/Fe(001) three-layer structure with 5 nm thick plates grown on the Ag(001)/Fe/GaAs(001) substrate was analyzed in detail. The optimum temperature conditions for layer-by-layer structure growth were found to produce high-quality interfaces between iron and chromium layers with a longitudinal roughness correlation length of about 20 nm. The optimized samples manifested well-defined two-monolayer oscillations of interlayer interaction in magnetooptical measurements. Their $M(H)$ hysteresis loops at an $L > 1.2$ nm spacer thickness had definitely flat regions at $M \approx \pm M_R$, $M_R \approx M_S = (M_1 + M_2)/2$ in low field H (an almost 90° coupling between the ferromagnetic plates) and almost jumpwise reached the saturation state at $H \approx \pm 10$–20 kA/m.

Also note that the half-angle coupling model qualitatively satisfactorily reproduces the special features of the $M(H)$ dependences obtained for [Fe(14 Å)/Cr(74 Å)](001) superlattices [25]. When a field is applied along the hard axis, magnetization in the first hysteresis loop quadrant gradually decreases as H lowers from the M_S saturation to remanent magnetization value $M_R \approx 0.7M_S$. When a field is applied along the easy axis, $M(H)$ drops in a low field $H \le 100$ Oe from $0.9M_S$ to $0.54M_S$.

Let us consider the applicability of our model with regard to the temperature and spacer thickness. The interlayer coupling effect in Fe/Cr-type structures is related just to the phase diagram region $T_0(L) < T < T_N(L)$ where a commensurate spin density wave (we use the terminology of [2]) exists in the spacer, and the $T_N(L)$ line is a boundary between this phase and a phase with an incommensurate spin density wave. The calculations in [12, 13] were performed on the assumption that $L \gg 2\xi_0$, but the $E(\psi)$ angular dependence of interlayer interaction (and its particular case (2)) was obtained using a general-type functional for orientation spin density wave fluctuations. We may therefore hope that (2) will be valid also for thinner spacers $L \sim 2\xi_0$; naturally, the $B_{1,2}(L, T)$ dependence will then be different from that obtained in [22]. The $E(\psi)$ dependence in form (2) was found on the assumption that the mean distance between monosteps $\sim\rho^{-1}$ and the angular correlation length ζ were no less than the thickness of the spacer itself, $L \ll (\zeta, \rho^{-1})$. Note that a particular case of (2) for $B_1 = B_2$ was obtained in terms of the model of interlayer coupling in ferromagnet/antiferromagnet structures [27] different from that suggested in [22].

The $E(\psi)$ dependence was derived in [22] with modeling Fe/Cr interface roughness by a one-dimensional periodic structure of monosteps parallel to each other. This model is, of course, far from reality. Spacer fragments with a constant N value that have different shapes and areas make their contribution to the effective coupling between the ferromagnetic plates. Nevertheless, the theoretical estimates obtained in [22] show that the major contribution to effective interlayer exchange is made by fragments with the largest area, which is in agreement with experiment [5].

For a structure with almost ideal interlayer boundaries, that is, in the limit $\zeta\rho \ll 1$, we can qualitatively estimate the effect of fluctuations of the distance between neighboring monosteps by averaging interlayer exchange energy (2) over some statistical Λ parameter distribution. For instance, suppose that Λ has a Gaussian distribution with the variance W and geometric mean $\langle\Lambda\rangle$. The dependence of the energy of the trilayer $\langle E(\psi)\rangle$ on W for $\langle\Lambda\rangle = 1/4$ and $I = 1$ is shown in Fig. 5. The equilibrium ψ_0 angle value shifts from the value corresponding to the mean $\langle\Lambda\rangle$ value ($\psi_0 \approx 143°$ at $W = 0$) to $\psi_0 = 90°$ (as $W \longrightarrow \infty$).

This approach to the interpretation of experimental data appears to be reasonable if the cross size of the region from which the signal is read is much larger than the mean distance between monosteps. Note that a laser beam is focused on the surface of a sample into a spot with a diameter of order 0.1 mm when the magnetization curves of three-layer Fe/Cr/Fe structures are measured using the magnetooptical Kerr effect [14].

In our view, in most experiments performed with Fe/Cr/Fe(001) trilayers and [Fe/Cr](001) superlattices, effective interlayer exchange coupling was accompanied by orientation changes (deformations) in spin density wave in the chromium spacer, which were induced by monatomic steps at interlayer boundaries. The appearance of ferromagnetic domain walls can most likely be expected in structures with very thin ($d <$ 10–15 Å) iron layers. The parameter region where

magnetization nonuniformity in iron layers should be taken into account can be qualitatively estimated as $d\gamma < L\delta$, where γ is the exchange stiffness of iron and $\delta \sim \Delta^2$ is the exchange stiffness of the chromium layer determined in [22] in terms of the charge-induced spin density wave model.

ACKNOWLEDGMENTS

The authors are deeply indebted to N.M. Kreines, D.I. Kholin, and A.B. Drovosekov for discussions and a detailed account of the experimental situation. This work was financially supported by the Russian Foundation for Basic Research (project no. 04-02-16938).

REFERENCES

1. D. T. Pierce, J. Unguris, R. J. Celotta, and M. D. Stiles, J. Magn. Magn. Mater. **200**, 290 (1999).
2. H. Zabel, J. Phys.: Condens. Matter **11**, 9303 (1999).
3. R. S. Fishman, J. Phys.: Condens. Matter **13**, R235 (2001).
4. S. O. Demokritov, J. Phys.: Appl. Phys. **31**, 925 (1998).
5. C. M. Schmidt, D. E. Burgler, D. M. Schaller, *et al.*, Phys. Rev. B **60**, 4158 (1999).
6. J. C. Slonczewski, Phys. Rev. Lett. **67**, 3172 (1991).
7. S. O. Demokritov, A. B. Drovosekov, N. M. Kreĭnes, *et al.*, Zh. Éksp. Teor. Fiz. **122**, 1233 (2002) [JETP **95**, 1062 (2002)].
8. M. Ruhrig, R. Schafer, A. Hubert, *et al.*, Phys. Status Solidi A **125**, 635 (1991).
9. A. Schreyer, J. F. Anker, Th. Zeidler, *et al.*, Phys. Rev. B **52**, 16066 (1995).
10. J. C. Slonczewski, J. Magn. Magn. Mater. **150**, 13 (1995).
11. D. A. Tulchinsky, J. Unguris, and R. J. Celotta, J. Magn. Magn. Mater. **212**, 91 (2000); D. T. Pierce, A. D. Davies, J. A. Stroscio, *et al.*, J. Magn. Magn. Mater. **222**, 13 (2000).
12. M. Avignon, V. Men'shov, and V. Tugushev, Europhys. Lett. **56**, 132 (2001).
13. V. N. Men'shov and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **120**, 899 (2001) [JETP **93**, 786 (2001)].
14. N. M. Kreĭnes, D. I. Kholin, S. O. Demokritov, and M. Rikart, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 1124 (2003) [JETP Lett. **78**, 627 (2003)].
15. V. V. Tugushev, in *Electronic Phase Transitions*, Ed. by W. Hanke and Yu. V. Kopaev (North-Holland, Amsterdam, 1992), Modern Problems in Condensed Matter Sciences, Vol. 32, p. 239.
16. E. E. Fullerton, C. H. Sowers, and S. D. Bader, Phys. Rev. B **56**, 5468 (1997).
17. H. Fritzsche, S. Bonn, J. Hauschild, *et al.*, Phys. Rev. B **65**, 144408 (2002).
18. E. E. Fullerton, J. L. Robertson, A. R. E. Prinsloo, *et al.*, Phys. Rev. Lett. **91**, 237 201 (2003).
19. K. Mibu, M. Takeda, J. Suzuki, *et al.*, Phys. Rev. Lett. **89**, 287202 (2002); K. Mibu and T. Shinjo, J. Phys.: Appl. Phys. **35**, 2359 (2002).
20. V. N. Men'shov and V. V. Tugushev, Fiz. Tverd. Tela (St. Petersburg) **44**, 1650 (2002) [Phys. Solid State **44**, 1727 (2002)].
21. V. N. Men'shov and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **122**, 1044 (2002) [JETP **95**, 901 (2002)].
22. V. N. Men'shov and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **125**, 136 (2004) [JETP **98**, 123 (2004)].
23. V. V. Kostyuchenko and A. K. Zvezdin, Phys. Rev. B **57**, 5951 (1998).
24. A. B. Drovosekov, O. V. Zhotikova, N. M. Kreĭnes, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1817 (1999) [JETP **89**, 986 (1999)].
25. S. Adenwalla, G. P. Felcher, E. E. Fullerton, *et al.*, Phys. Rev. B **53**, 2474 (1996); E. E. Fullerton, S. Adenwalla, G. P. Felcher, *et al.*, Physica B (Amsterdam) **221**, 370 (1996).
26. A. Azevedo, C. Chesman, S. M. Rezende, *et al.*, Phys. Rev. Lett. **76**, 4837 (1996).
27. A. I. Morozov and A. S. Sigov, Fiz. Tverd. Tela (St. Petersburg) **41**, 1240 (1999) [Phys. Solid State **41**, 1130 (1999)]; **46**, 385 (2004) [**46**, 395 (2004)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 576–596.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 652–673.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

High-Frequency Nonlinear Response of Thin Superconducting Films with a Mixed *d* and *s* Symmetry of the Order Parameter

V. V. Kurin* and A. A. Utkin**

Institute for Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, GSP-105, 603600 Russia
**e-mail: kurin@ipm.sci-nnov.ru*
***e-mail: tolik@ipm.sci-nnov.ru*
Received September 8, 2004

Abstract—Third-harmonic generation during reflection of electromagnetic radiation from a thin superconducting film with a mixed *d*- and *s*-order parameter is studied theoretically. The dependence of the third harmonic intensity on the temperature and amplitude of an incident wave is calculated in the framework of the Ginzburg–Landau theory for superconductors with a two-component order parameter, and its behavior in the vicinity of transitions between phases with different symmetries is analyzed. It is shown, in particular, that the third harmonic intensity in the vicinity of the temperature corresponding to the $d \rightleftharpoons d + s$ phase transition substantially increases and is a nonanalytic function of the amplitude of the incident wave, while no singularity in the nonlinear response is observed for the $d \rightleftharpoons d + is$ transition. The linear reflection coefficient is found to be virtually insensitive to these phase transitions. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In spite of the fact that cuprate superconductors (HTSCs) with a superconducting transition temperature an order of magnitude higher than in normal superconductors were discovered 18 years ago, the problem of symmetry of the superconducting order parameter remains at the focus of researchers' attention. It has been almost commonly accepted that the order parameter with *d* symmetry dominates in HTSC materials [1, 2]. At the same time, some experiments on Raman scattering [3], measurements of surface impedance [4], tunneling between an HTSC and a normal superconductor [5], and angular resolution photoemission [1, 6] indicate that the *s* component of the order parameter is also present along with the order parameter with *d* symmetry.

The concept of mixed $s + e^{i\theta}d$ symmetry of the order parameter for high-T_c superconductors was apparently discussed for the first time in [7, 8] and was used, for example, for analyzing the experimental data on nuclear magnetic resonance in YBCO [9] and for interpreting Josephson measurements in HTSCs [10]. Many microscopic models proposed for describing the HTSC mechanism in cuprates have led to the conclusion that mixed symmetry of the order parameter is possible for high-T_c superconductors. Superconductors with a mixed-symmetry order parameter can be phenomenologically described using the Ginzburg–Landau theory including two equations for interacting order parameters and the Maxwell equations supplemented with an expression for supercurrent, which is the sum of not only the individual currents of the *s* and *d* components, but also the interference component [11–13].

Here, we concentrate our attention on nontrivial electrodynamic properties of such superconductors, in particular, on a peculiar nonlinear response associated with the presence of the interference current. Calculating the generation of the third harmonic of incident radiation in a thin superconducting film using the Ginzburg–Landau theory for a two-component order parameter by way of example, we will show that the nonlinear response is extremely sensitive to possible phase transitions between superconducting phases with different symmetries and that this response may substantially increase in the vicinity of phase transitions. In particular, we will show that the third harmonic intensity noticeably increases in the vicinity of the temperature corresponding to the $d \rightleftharpoons d + s$ phase transition and is a nonanalytic function of the amplitude of the incident wave, while no singularity is observed in the nonlinear response during the $d \rightleftharpoons d + is$ transition. On the other hand, the linear reflection coefficient is found to be virtually insensitive to these phase transitions. Thus, experimental investigation of nonlinear characteristics of superconductors is an interesting tool for studying transitions between phases with different symmetries of the order parameter.

The paper has the following structure. In Section 2, the electrodynamic part of the problem is considered. Under the assumption of weak nonlinearity and local nature of the relation between the vector potential and the current in the superconductor, the problem of incidence of a plane wave on a thin superconducting film is

solved. The relationship between the amplitude of the incident wave and the amplitudes of reflected and transmitted waves at the frequencies of the incident wave and its third harmonic is determined. In Section 3, the expressions connecting the electric current with the vector potential at various temperatures corresponding to phases with different symmetries of the order parameter are derived by solving the Ginzburg–Landau equations for superconductors with a mixed *s* and *d* order parameter. In Section 4, the results are analyzed and the potentialities of the linear and nonlinear electrodynamic diagnostics of the symmetry properties of the order parameter in superconductors are compared.

2. REFLECTION OF AN ELECTROMAGNETIC WAVE FROM A THIN SUPERCONDUCTING FILM

Let us consider the problem of normal incidence of a linearly polarized electromagnetic wave with circular frequency ω and with electric field amplitude $\boldsymbol{E}_{\text{inc}}$ on a thin homogeneous superconducting film of thickness *d* (Figs. 1 and 2). We must find the electromagnetic field distribution outside and inside the film. We write the Maxwell equations

$$\text{curl}\mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j}, \quad \text{div}\mathbf{B} = 0,$$

$$\text{curl}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \text{div}\mathbf{E} = 4\pi\rho.$$

Here, **E** and **B** are the electric and magnetic fields and ρ and **j** are the charge and current densities. We introduce in the usual manner the vector and scalar potentials **A** and ϕ so that $\mathbf{B} = \text{curl}\mathbf{A}$ and $\mathbf{E} = -c^{-1}\partial\mathbf{A}/\partial t - \nabla\phi$. Substituting these expressions into the system of the Maxwell equations and choosing the calibration $\text{div}\mathbf{A} = 0$, we obtain the following equation for **A**:

$$\Delta\mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c}\mathbf{j}. \qquad (1)$$

Considering that expression $\text{div}\,\mathbf{E} = 0$ holds in our problem in view of the transverse nature of electromagnetic waves, we find that potential ϕ satisfies the Laplace equation, which in our case has a solution ϕ = const. Thus, the system of Maxwell equations is reduced to a single equation for **A**.

To close this equation, we must find the dependence of current density **j** on the electromagnetic field. We will refer to this dependence as the constitutive relation. The current density in a superconductor is an explicit function of the vector potential and the order parameter. In the case of a London superconductor, this function is linear and local in **A**. We assume that the order parameter has two components ψ_d and ψ_s associated with the *d* and *s* ordering, respectively. Then the constitutive relation in this calibration can be written in the form

$$\mathbf{j} = \overleftrightarrow{f}(\psi_d, \psi_s)\mathbf{A}, \qquad (2)$$

where $\overleftrightarrow{f}$ is a certain rank two tensor. It is well known that the superconducting order parameter is suppressed by a magnetic field. In our case, this effect leads to the dependence of ψ_d and ψ_s on the vector potential, $\psi_{d,s} = \psi_{d,s}(\mathbf{A})$. Consequently, after the substitution of $\psi_{d,s}(\mathbf{A})$ into Eq. (2), dependence $\mathbf{j}(\mathbf{A})$ becomes nonlinear. Equation (1) will be also nonlinear in **A**, which leads to generation of harmonics in our problem. We will

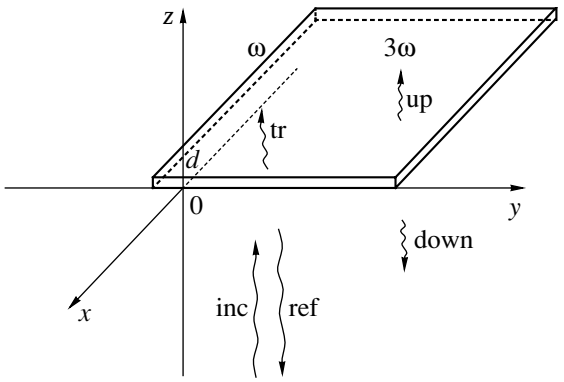

Fig. 1. Incidence of an electromagnetic wave on a superconducting film (3D projection). The incident, reflected, and transmitted waves of frequency ω, which are re-emitted upwards and downwards at tripled frequency 3ω, are indicated. The lower and upper surfaces of the film coincide with the planes $z = 0$ and $z = d$, respectively. The incident wave propagates along the *z* axis. The *x* and *y* axes coincide with the crystallographic axes of the crystal.

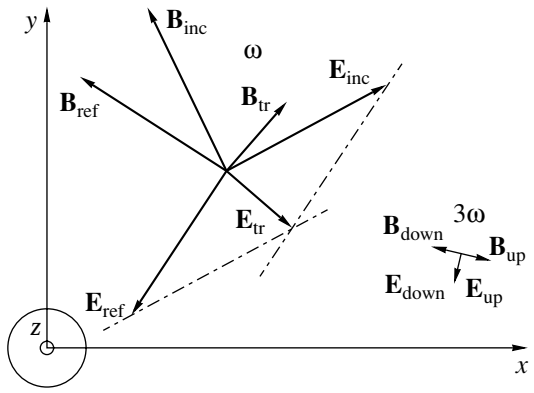

Fig. 2. Incidence of an electromagnetic wave on a superconducting film (top view). The electric and magnetic field vectors in the incident, reflected, and transmitted waves of frequency ω, as well as the fields in the waves re-emitted upwards and downwards at tripled frequency, are indicated. All fields are shown for the same instant. The fields are shown for $z = 0$ for waves propagating in the lower half-space and for $z = d$ for waves propagating in the upper half-space.

assume that the nonlinearity is weak in the sense that the changes in the order parameter associated with the effect of the magnetic field are small as compared to the characteristic value of the order parameter itself in zero magnetic field. The dependences $\psi_{d,s}(\mathbf{A})$ can then be presented in the form $\psi_{d,s} = \psi^0_{d,s} + G_{d,s}(\mathbf{A})$, where $\psi^0_{d,s}$ are the values of the order parameter components in zero magnetic field and $G_{d,s}(\mathbf{A})$ are the deviations caused by the magnetic field, $|G_{d,s}(\mathbf{A})| \ll |\psi^0_{d,s}|$. Substituting these relations into Eq. (2) and expanding function $\overleftrightarrow{f}(\psi_d, \psi_s)$ into a series in the vicinity of point (ψ^0_d, ψ^0_s), we can obtain the relation for the current density in the film,

$$\mathbf{j} = -c(4\pi)^{-1}\overleftrightarrow{1/\lambda^2}\mathbf{A} + \overleftrightarrow{G}(\mathbf{A})\mathbf{A}, \tag{3}$$

where $\overleftrightarrow{1/\lambda^2} = -4\pi c^{-1}\overleftrightarrow{f}(\psi^0_d, \psi^0_s)$ is the tensor connecting $\mathbf{j}$ and $\mathbf{A}$ in the linear approximation,

$$\overleftrightarrow{G}(\mathbf{A}) = \frac{\partial \overleftrightarrow{f}}{\partial \psi_d}(\psi^0_d, \psi^0_s)G_d(\mathbf{A}) + \frac{\partial \overleftrightarrow{f}}{\partial \psi_s}(\psi^0_d, \psi^0_s)G_s(\mathbf{A}) + \text{c.c.}$$

To simplify calculations, we will assume in this section that tensor $\overleftrightarrow{1/\lambda^2}$ can be represented in the form $1/\lambda^2 \times \delta_{ik}$, where δ_{ik} is a unit tensor (at a later stage, we will find $\overleftrightarrow{f}(\psi_d, \psi_s)$ and see in which cases this assumption is valid). This means that $\mathbf{j}$ and $\mathbf{A}$ are parallel if we disregard the term nonlinear in $\mathbf{A}$ in expression (3). We will henceforth assume that the values of $1/\lambda^2$ and $\overleftrightarrow{G}(\mathbf{A})$ are known and analyze Eq. (1) with expression (3) for the current density in the film. Considering that the current flows only in the film, we can write this equation in the form

$$\frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = \begin{cases} \dfrac{\mathbf{A}}{\lambda^2} - \dfrac{4\pi}{c}\overleftrightarrow{G}(\mathbf{A})\mathbf{A}, & 0 \le z \le d, \\ 0, & z > d \ \text{ or } \ z < 0. \end{cases} \tag{4}$$

We assume that film thickness d is much smaller than the characteristic scale of variation of $\mathbf{A}$ in the film. Then we can assume that the electromagnetic field and current distributions in the film are uniform in coordinate z. Taking into account the fact that the film is infinitely large in the xy plane, we can assume that the current and electromagnetic field distributions in the film are uniform. It should be noted that the London penetration depth in HTSC materials is on the order of 2000 Å; consequently, the film thickness must be much smaller than this value. In fact, the experimental thickness of the films is of the same order of magnitude. In our estimates, we will assume that the film thickness is 1000 Å (i.e., half as large as the London penetration depth). In this case, the uniform current approximation naturally cannot be treated as completely satisfactory; however, calculations show that the error will be not large; in addition, the error cannot influence the qualitative effects we are going to predict. For this reason, we will use the thin film approximation. In this case, Eq. (4) can be written in the form

$$\frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = d\delta(z)\left\{\frac{1}{\lambda^2}\mathbf{A} - \frac{4\pi}{c}\overleftrightarrow{G}(\mathbf{A})\mathbf{A}\right\}. \tag{5}$$

It would be recalled that nonlinear effects are regarded as weak. This means that

$$\left|\frac{4\pi}{c}\overleftrightarrow{G}(\mathbf{A})\mathbf{A}\right| \ll \left|\frac{\mathbf{A}}{\lambda^2}\right|.$$

Then Eq. (5) can be solved using perturbation theory. We will seek the solution in the form $\mathbf{A} = \mathbf{A}_l + \mathbf{A}^{(1)}$, where $\mathbf{A}_l$ is the solution to Eq. (5) disregarding the nonlinear term and $\mathbf{A}^{(1)}$ is a correction determined by the nonlinear term in the equation. In fact, the zeroth approximation corresponds to the linear problem of reflection of an electromagnetic wave from the film, while the first approximation corresponds to the problem of generation of harmonics as well as the influence of nonlinear effects on the solution in the linear approximation.

Let us first consider the zeroth order of perturbation theory. It follows from Eq. (5) that the equation for $\mathbf{A}_l$ has the form

$$\frac{\partial^2 \mathbf{A}_l}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}_l}{\partial t^2} = d\delta(z)\frac{1}{\lambda^2}\mathbf{A}_l. \tag{6}$$

This equation describes the reflection of an incident wave from the film. Solving this equation, we find that the reflection coefficient is given by

$$R_\omega = -\frac{icd/2\omega\lambda^2}{1 + icd/2\omega\lambda^2}$$

and the transmission coefficient is

$$T_\omega = \frac{1}{1 + icd/2\omega\lambda^2}.$$

Let us estimate the value of quantity $cd/\omega\lambda^2$ appearing in the reflection and transmission coefficients. If the wave frequency $\omega \sim 10^9$ Hz, then the value of $cd/\omega\lambda^2$

for an HTSC film with $\lambda \sim 2 \times 10^{-5}$ cm is on the order of 10^5 even for a film of thickness on the order of 10^5. Consequently, we must assume that $cd/\omega\lambda^2 \gg 1$. In this case, $|1 + R_l| \ll 1$ and $|T| \ll 1$; i.e., the wave is reflected almost completely. Then we can assume to a high degree of accuracy that the amplitude $\mathbf{j}_0$ of the current density in the film is connected only with $\mathbf{E}_{\text{inc}}$ via the simple relation $\mathbf{j}_0 = c\mathbf{E}_{\text{inc}}/2\pi d$. Consequently, we must assume in the subsequent solution that the linear current $\mathbf{j}_l = \mathbf{j}_0 \sin\omega t$ in the film is preset (the preset current approximation); i.e., we have

$$\mathbf{j}_l = \mathbf{j}_0 \sin\omega t = \frac{c}{2\pi d}\mathbf{E}_{\text{inc}} \sin\omega t. \tag{7}$$

For the subsequent calculation in the first order of perturbation theory, we must know $\mathbf{A}_l(z = 0)$ since it is quantity $\mathbf{A}$ that appears in the nonlinear part of Eq. (5). This quantity can be determined most easily from Eq. (3) in the linear approximation. This gives

$$\mathbf{A}_l(z = 0) = -\frac{4\pi}{c}\lambda^2 \mathbf{j}_l, \tag{8}$$

and linear current $\mathbf{j}_l$ is directly connected with the known quantity $\mathbf{E}_{\text{inc}}$.

Let us now consider the first order of perturbation theory. Equation (5) leads to the following equation for $\mathbf{A}^{(1)}$:

$$\frac{\partial^2 \mathbf{A}^{(1)}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}^{(1)}}{\partial t^2} = d\delta(z)\left\{\frac{1}{\lambda^2}\mathbf{A}^{(1)} - \frac{4\pi}{c}\overleftrightarrow{G}(\mathbf{A}_l)\mathbf{A}_l\right\}. \tag{9}$$

In fact, this equation describes the emission of electromagnetic waves by extraneous current,

$$\mathbf{j}_e = \overleftrightarrow{G}(\mathbf{A}_l)\mathbf{A}_l\big|_{z=0} = -\overleftrightarrow{G}\left(-\frac{4\pi}{c}\lambda^2 \mathbf{j}_l\right) \times \frac{4\pi}{c}\lambda^2 \mathbf{j}_l, \tag{10}$$

flowing through the film. Let us consider the frequency spectrum of this radiation. It is clear from the symmetry of the problem that the film emits identically in both directions; consequently, the spectra of radiation in the upward and downward directions coincide. Since $\overleftrightarrow{G}(\mathbf{A})\mathbf{A}$ is a vector function of the vector argument and a constant magnetic field is absent in the problem, it is clear from symmetry considerations that the expression for the extraneous current,

$$-\frac{4\pi}{c}\overleftrightarrow{G}\left(-\frac{2\lambda^2}{d}\mathbf{E}_{\text{inc}} \sin\omega t\right) \times \mathbf{E}_{\text{inc}} \sin\omega t,$$

can contain only odd harmonics of frequency ω. Consequently, the spectrum of radiation generated by this current will contain only such harmonics. We will be interested in third harmonic radiation since it is must stronger than radiation at higher harmonics. The amplitude of the wave emitted at the third harmonic will be referred to as the nonlinear response. To single out the part describing the process at frequency 3ω from Eq. (9), we derive the equation for $\mathbf{A}^{(1)}_{3\omega}$, viz., the coefficient of $\sin 3\omega t$ in the Fourier series for $\mathbf{A}^{(1)}$. For this purpose, we multiply Eq. (9) by $\pi^{-1}\omega \sin 3\omega t$ and integrate over the period $2\pi\omega^{-1}$ of the incident wave. This gives

$$\frac{\partial^2 \mathbf{A}^{(1)}_{3\omega}}{\partial z^2} + \frac{(3\omega)^2}{c^2}\mathbf{A}^{(1)}_{3\omega} = d\delta(z)\left\{\frac{1}{\lambda^2}\mathbf{A}^{(1)}_{3\omega} - \frac{4\pi}{c}\mathbf{j}_{e3\omega}\right\}, \tag{11}$$

where

$$\mathbf{j}_{e3\omega} = \frac{\omega}{\pi}\int_0^{2\pi/\omega} \mathbf{j}_e(t) \sin 3\omega t dt. \tag{12}$$

Solving this equation, we obtain the vector potential amplitude in the emitted wave in the form

$$\mathbf{A}^{\text{rad}}_{3\omega} = R_{3\omega} 4\pi c^{-1}\lambda^2 \mathbf{j}_{e3\omega},$$

where

$$R_{3\omega} = -\frac{icd/2(3\omega)\lambda^2}{1 + icd/2(3\omega)\lambda^2}$$

is the coefficient of reflection of a wave with frequency 3ω from the film. Considering that $cd/\omega\lambda^2 \gg 1$, we obtain the electric field amplitude in the nonlinear response,

$$\mathbf{E}^{\text{rad}}_{3\omega} = -i\frac{12\pi\omega\lambda^2}{c^2}\mathbf{j}_{e3\omega}. \tag{13}$$

It is this formula that will be used in subsequent analysis since we will usually determine precisely $\mathbf{j}_{e3\omega}$; however, to analyze the result, we will now proceed a little further. Formulas (10) and (12) imply that quantity $\mathbf{j}_{e3\omega}$ can be represented in the form

$$\mathbf{j}_{e3\omega} = \overleftrightarrow{h}(\lambda^2, \mathbf{j}_0)\mathbf{j}_0, \tag{14}$$

where

$$\overleftrightarrow{h}(\lambda^2, \mathbf{j}_0) = -\frac{\omega}{\pi}\int_0^{2\pi/\omega} \frac{4\pi\lambda^2}{c} \times \overleftrightarrow{G}\left(-\frac{4\pi}{c}\lambda^2 \mathbf{j}_0 \sin\omega t\right) \sin\omega t \sin 3\omega t dt.$$

Substituting this expression for $\mathbf{j}_{e3\omega}$ into relation (13) and considering that $\mathbf{j}_0$ and $\mathbf{E}_{\text{inc}}$ are connected via relation (7), we obtain the following expression for the nonlinear response amplitude in terms of the amplitude of the incident wave:

$$\mathbf{E}_{3\omega}^{\text{rad}} = -6i\overleftrightarrow{h}(\lambda^2, \mathbf{j}_0)\frac{\omega\lambda^2}{cd}\mathbf{E}_{\text{inc}}. \quad (15)$$

Let us analyze the results. First of all, it is clear that these results can easily be generalized to the case when quantity $\overleftrightarrow{1/\lambda^2}$ cannot be represented in the form $(1/\lambda^2)\delta_{ik}$. For this purpose, we must assume that λ^2 in formulas (13)–(15) is a tensor reciprocal to $\overleftrightarrow{1/\lambda^2}$ and multiply it by the vector standing on the right of it. Further, we recall that the nonlinearity is regarded as weak; this means that $\|\overleftrightarrow{h}\| \ll 1$, which immediately implies that the nonlinear response is much smaller than the linear one. However, the attenuation associated with almost total reflection of the incident wave also exists. Indeed, the result contains a small factor $\omega\lambda^2/cd$, which is due to the fact that the incident wave experiences almost total reflection. Since the value of $\omega\lambda^2/cd$ is on the order of 10^{-5} according to our estimates, while $\|\overleftrightarrow{h}\|$ is not greater than 10^{-1}, the nonlinear response is suppressed by at least six orders of magnitude as compared to the linear response.

Let us now consider this result from the standpoint of subsequent considerations. Formulas (13)–(15) make it possible to express the nonlinear response amplitude in terms of the amplitude of the incident wave if we know the constitutive relation for the film. In other words, knowing $\overleftrightarrow{1/\lambda^2}$ and $\overleftrightarrow{G}(\mathbf{A})$, we can immediately find the nonlinear response. Obviously, the obtained relations are valid for any constitutive equation; the only requirement is that the constitutive equation should be decomposable into the sum of parts, which are linear and nonlinear in $\mathbf{A}$. Consequently, we must first find the constitutive equation in the framework of our model of the superconductor. Then, we must find extraneous current $\mathbf{j}_e$ and its Fourier transform $\mathbf{j}_{e3\omega}$ in terms of the linear current. After this, it only remains for us to use formula (15) for the nonlinear response.

3. NONLINEAR RESPONSE OF A SUPERCONDUCTOR WITH A COMPLEX TWO-COMPONENT ORDER PARAMETER

In this section, we will find the constitutive relation for a thin superconducting film with a two-component order parameter. We consider a thin HTSC film with cuprate layers parallel to its surface and will find the constituent relation using the Ginzburg–Landau theory. We assume that the film has a complex two-component order parameter, whose first component, $\psi_d = f_d\exp(i\theta_d)$, is associated with d pairing and the second component, $\psi_s = f_s\exp(i\theta_s)$, is due to s pairing. It is known [14] that the crystal lattice of an HTSC-type compound may possess various types of symmetry. Depending on the preparation method, these can be symmetry groups D_{2h}, D_{4h}, and others. We will consider the D_{4h} symmetry, which is usually realized in the case of high-temperature annealing of the sample. In this case, we can write the Ginzburg–Landau free energy in the form [11]

$$\begin{aligned} F = \int\{ & a_d|\psi_d|^2 + b_2|\psi_d|^4 + \gamma_d|\mathbf{\Pi}\psi_d|^2 \\ & + a_s|\psi_s|^2 + b_1|\psi_s|^4 + \gamma_s|\mathbf{\Pi}\psi_s|^2 + b_3|\psi_s|^2|\psi_d|^2 \\ & + b_4(\psi_s^{*2}\psi_d^2 + \psi_d^{*2}\psi_s^2) + \gamma_v((\Pi_y\psi_s)^*(\Pi_y\psi_d) \\ & - (\Pi_x\psi_s)^*(\Pi_z\psi_d) + \text{c.c.}) \\ & + (8\pi)^{-1}(\text{curl}\mathbf{A})^2\}d^2\mathbf{r}_\perp dz, \end{aligned} \quad (16)$$

where $\Pi = -i\hbar\nabla - 2ec^{-1}\mathbf{A}$; $a_{s,d} = \alpha_{s,d}(T - T_{s,d})$, and the x and y axes are directed along the a and b crystallographic axes, respectively. Since a generally accepted microscopic theory of HTSC has not been worked out as yet, we are not aware of parameters appearing in this functional. Thus, we must make certain assumptions. We assume that $T_d > T_s$ due to the fact that the d component of the order parameter is predominant. In addition, we assume that $T_d - T_s \sim T_s$ to separate by temperature the effects associated with closeness to T_d from those associated with closeness to T_s. Note that in this case there are no temperatures for which the expressions for $a_d = \alpha_d(T - T_d)$ and $a_s = \alpha_s(T - T_s)$, which can be obtained from expansion in temperature in the vicinity of T_d and T_s, are valid simultaneously. For this reason, our assumption generally contradicts the condition of the applicability of the Ginzburg–Landau theory. However, it will be shown below that it is only important for our analysis to establish the temperature at which $a_s = 0$; consequently, we can use the simplest expression for a_s even when $T_d - T_s \sim T_s$. In addition, we introduce only a certain quantitative error using a functional of type (16) in the entire temperature range. As regards the remaining parameters of the theory, we can estimate them using the BCS theory [15], according to which parameters b_1, b_2, b_3, and b_4 have the same order of magnitude, as do parameters γ_s, γ_d, γ_v, and also α_s and α_d.

Let us now find the constitutive relation. For this purpose, we derive the expression for current density by varying functional (16) in **A**. We recall that parameters f_s and f_d are also uniform since the current density is uniform in the coordinate. Taking into account this circumstance in the relation obtained as a result of variation, we arrive at the following expression for current density [12]:

$$\mathbf{j} = 2en_d\mathbf{v}_d + 2en_s\mathbf{v}_s = 2en_{sd}\bar{\mathbf{v}}_{sd}\cos(\theta_d - \theta_s); \quad (17)$$

here,

$$n_s = f_s^2, \quad \mathbf{v}_s = 2\gamma_s\left(\hbar\nabla\theta_s - \frac{2e}{c}\mathbf{A}\right), \quad n_d = f_d^2,$$

$$\mathbf{v}_d = 2\gamma_d\left(\hbar\nabla\theta_d - \frac{2e}{c}\mathbf{A}\right), \quad n_{sd} = \sqrt{n_s n_d},$$

$$\bar{\mathbf{v}}_{sd} = 2\gamma_v\left(\hbar\bar{\nabla}\theta_s + \hbar\bar{\nabla}\theta_d - \frac{4e}{c}\bar{\mathbf{A}}\right)$$

(the bar over a vector $\mathbf{C} = C_y\mathbf{y}_0 + C_x\mathbf{x}_0$ means that $\bar{\mathbf{C}} = C_y\mathbf{y}_0 - C_x\mathbf{x}_0$). It can be seen that the current is the sum of the current of the *d* component, the current of the *s* component, and the interference part. It can easily be proved that quantities $\mathbf{v}_s$, $\mathbf{v}_d$, $\bar{\mathbf{v}}_{sd}$, and $\theta_d - \theta_s$ are gauge invariant. We have already fixed a calibration for **A** (div**A** = 0) and now choose a calibration for phases θ_d and θ_s. Let us assume that $\theta_s = 0$ and $\theta_d = \theta_0$. In this case, the order parameter components ψ_d and ψ_s are determined by three real numbers f_d, f_s, and θ_0, from which f_d and f_s are responsible for the moduli of the order parameter components and θ_0 is the angle between ψ_d and ψ_s on the complex plane. To get rid of the cosine, we write the expression for current density in the form

$$\mathbf{j} = -\frac{8e^2}{c}\{\gamma_d(U^2 + V^2)\mathbf{A} + \gamma_s f_s^2\mathbf{A} + 2\gamma_v U f_s\bar{\mathbf{A}}\}, \quad (18)$$

where $U = f_d\cos\theta_0$ and $V = f_d\sin\theta_0$. In this expression, the quantity f_s, U, and V determining the order parameters must be found from the condition that these quantities are responsible for the free energy minimum. To derive the equations for these quantities, we vary the free energy in ψ_s^* and ψ_d^*. Taking into account the uniformity of f_s, U, and V in the resulting relations, we obtain the following system of equations for the order parameters:

$$\left(a_s + \frac{4e^2}{c^2}\gamma_s\mathbf{A}^2\right)f_s + 2b_1 f_s^3 + b_3 f_3(U^2 + V^2) + 2b_4 f_s(U^2 - V^2) + \frac{4e^2}{c^2}\gamma_v(A_y^2 - A_x^2)U = 0, \quad (19)$$

$$\left(a_d + \frac{4e^2}{c^2}\gamma_d\mathbf{A}^2\right)U + 2b_2(U^2 + V^2)U + (b_3 + 2b_4)f_s^2 U + \frac{4e^2}{c^2}\gamma_v(A_y^2 - A_x^2)f_s = 0, \quad (20)$$

$$\left(a_d + \frac{4e^2}{c^2}\gamma_d\mathbf{A}^2\right)V + 2b_2(U^2 + V^2)V + (b_3 - 2b_4)f_s^2 V = 0. \quad (21)$$

We must find all solutions to this system and choose from these solutions the one associated with the deepest minimum of the free energy. We will refer to such a solution as a minimizing solution. The form of system (19)–(21) implies that the minimizing solution is a function of **A**, which leads precisely, as mentioned above, to a nonlinear constitutive relation. It should also be noted that our task is to analyze the relation between the nonlinear response and phase transitions. Consequently, we must find the nonlinear response as a function of temperature so that, by varying temperature, to be able to carry out our analysis in the vicinity of phase transitions and far away from them and thus determine the effect of phase transitions on the nonlinear response. In addition, we must also determine its dependence on other parameters of the Ginzburg–Landau theory. For this reason, we will seek the minimizing solution for the entire temperature range as well as for a certain range of other parameters of the theory.

Unfortunately, it is impossible to find exact solutions to system (19)–(21) in most cases. Hence, we will use an approximate method based on perturbation theory using the fact that the magnetic field and the vector potential are small. It should be recalled that the effects associated with the suppression of the order parameter by the magnetic field are assumed to be weak. Then, in solving the system of equations, we can assume that

$$\frac{4e^2}{c^2}\gamma_{s,d,v}\mathbf{A}^2 \ll a_{s,d} \quad (22)$$

and solve the system using perturbation theory. In the zeroth order of perturbation theory, we set **A** = 0 and solve the system to find the order parameters minimizing the free energy in zero electromagnetic field. Substituting this solution into relation (18), we obtain the

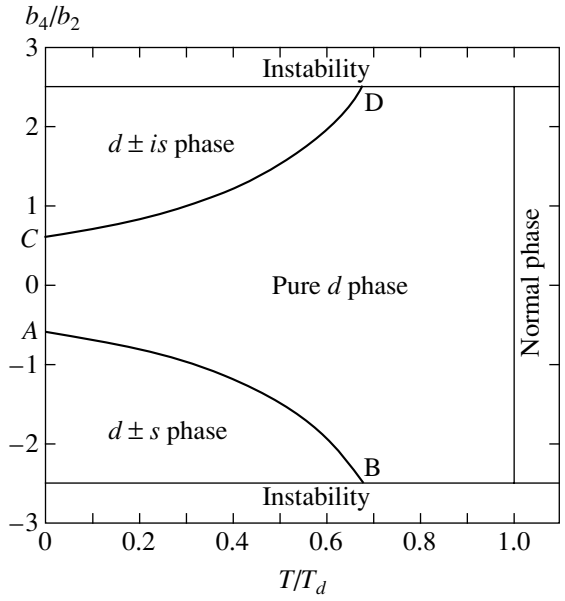

Fig. 3. Phase diagram of a superconductor with a two-component order parameter in the absence of current for the following values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $T_s/T_d = 0.5$, and $b_3/b_2 = 3$.

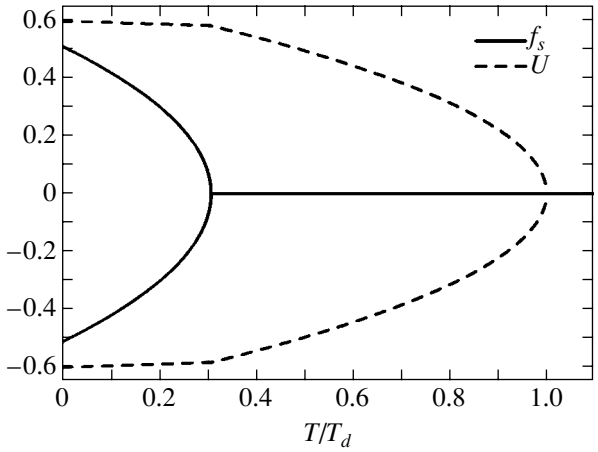

Fig. 4. Dependence of the order parameters on temperature $\tau = T/T_d$ in the absence of current. The order parameters are measured in units of $\sqrt{\alpha_d T_d/b_2}$. The solid curve corresponds to $f_s(\tau)$ and the dashed curve, to $\sqrt{U^2 + V^2}(\tau)$. In the temperature range, where $f_s \neq 0$, we have $U = 0$. The values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $b_3/b_2 = 3$, $b_4/b_2 = 1$, and $T_s/T_d = 0.5$.

component of $\overleftrightarrow{f}(\psi_d, \psi_s)$ in formula (2), which is independent of **A** (i.e., tensor $\overleftrightarrow{1/\lambda^2}$). Then, in the first order of perturbation theory, we obtain corrections to this solution, which determine the component of $\overleftrightarrow{f}(\psi_d, \psi_s)$ that depends on **A** (an analogous method for finding order parameters was used in [16]). The exact mathematical condition for applicability of such an approach will be formulated below. We will first consider system (19)–(21) in the zeroth approximation (i.e., for **A** = 0). In this case, this system can easily be solved exactly. We have the following solutions:

$$f_s = 0, \quad U = 0, \quad V = 0, \tag{i}$$

$$f_s = 0, \quad U^2 + V^2 = -\frac{a_d}{2b_2}, \tag{ii}$$

$$f_s^2 = \frac{-2b_2a_s + (b_3 + 2b_4)a_d}{4b_1b_2 - (b_3 + 2b_4)^2}, \quad V = 0,$$
$$U^2 = \frac{-2b_1a_d + (b_3 + 2b_4)a_s}{4b_1b_2 - (b_3 + 2b_4)^2}, \tag{iii}$$

$$f_s^2 = \frac{-2b_2a_s + (b_3 - 2b_4)a_d}{4b_1b_2 - (b_3 - 2b_4)^2}, \quad U = 0,$$
$$V^2 = \frac{-2b_1a_d + (b_3 - 2b_4)a_s}{4b_1b_2 - (b_3 - 2b_4)^2}, \tag{iv}$$

$$f_s^2 = -\frac{a_s}{2b_1}, \quad U = 0, \quad V = 0. \tag{v}$$

Solution (i) corresponds to the normal phase, solution (ii) to the pure d phase, solution (iii) to the $d \pm s$ phase, solution (iv) to the $d \pm is$ phase, and, finally, solution (v) to the pure s phase. Analyzing second derivatives of the free energy with respect to the order parameter components, we can easily choose the minimizing solution for various parameters of the Ginzburg–Landau theory. It turns out that the results can be conveniently depicted on the (b_4, T) phase diagram (Fig. 3). It can be seen that if $|b_4|$ is larger than a certain quantity, two second-order phase transitions may occur upon a variation of temperature from a value higher than the critical temperature to zero. The first of these transitions is the transition from the normal to the superconducting state, while the second is the transition from the pure d phase to a mixed d and s phase. If b_4 is smaller than zero, the second phase transition occurs to the $d \pm s$ phase, and for $b_4 > 0$, to the $d \pm is$ phase. We denote the first transition as AB and the second as CD. We will also consider the temperature dependence of the order parameters (Fig. 4). It should be noted that in the range of parameters in which the pure d phase is realized, θ_0 can assume any value since solution (ii) corresponding to the pure d phase contains only the combination $U^2 + V^2$. Such a situation is due to the fact that $f_s = 0$, but $f_d \neq 0$; for this reason, the angle between ψ_d and $\psi_s = 0$ can assume any value. This means that we have degeneracy in U and V in the case of the pure d phase. This occurs only when one of the components of

the order parameter is zero and the other differs from zero (i.e., for the pure *d* phase).

Let us now consider the first order of perturbation theory, i.e., the case $\mathbf{A} \neq 0$. System (19)–(21) in this case has no solutions for which $f_s = 0$, but $U^2 + V^2$ differs from zero. Consequently, the degeneracy in *U* and *V* existing in the region of the pure *d* phase in the absence of current is removed. In this connection, the problem of appropriate choice of the zeroth approximation arises. To solve this problem, we simultaneously analyze system (19)–(21) and the free energy generating this system. Since the left-hand side of Eq. (21) is the product of *V* and a certain polynomial, two types of solutions exist for system (19)–(21): the first (I) type for $V = 0$ and the second (II) type for $V \neq 0$. Type I corresponds to the $s \pm d$ type of the order parameter ($\theta_0 = 0$, π), while type II corresponds to the $s \pm e^{i\theta_0}$ type of the order parameter ($\theta_0 \neq 0, \pi$). Let us find which of the solutions is minimizing depending on the parameters of the Ginzburg–Landau theory. This will give the answer to the problem of the correct zeroth approximation in the solution of the system. For this purpose, we write the part of the free energy containing the order parameters. Formula (16) leads to

$$\begin{aligned} F_\psi = \int\Bigg\{&\left(a_d + \frac{4e^2}{c^2}\gamma_d \mathbf{A}^2\right)(U^2 + V^2) \\ &+ b_2(U^2 + V^2)^2 + \left(a_s + \frac{4e^2}{c^2}\gamma_s \mathbf{A}^2\right) f_s^2 \\ &+ b_1 f_s^4 + b_3 f_s^2 (U^2 + V^2) + 2b_4 f_s^2 (U^2 - V^2) \\ &+ \frac{8e^2}{c^2}\gamma_v (A_y^2 - A_x^2) f_s U\Bigg\} d\mathbf{r}dz. \end{aligned} \tag{23}$$

It can be seen that the mutual distribution as well as the signs of *U* and *V* are determined only by terms $2b_4 f_s^2 (U^2 - V^2)$ and $8e^2c^{-2}\gamma_v(A_y^2 - A_x^2)f_s U$ since *U* and *V* appear in the remaining terms only in combination $U^2 + V^2$. The fact that minimizing parameters f_s, *U*, and *V* correspond to the deepest minimum of *F* leads to the following conclusions.

1. The sign of *U* is determined only by the term $8e^2c^{-2}\gamma_v(A_y^2 - A_x^2)f_s U$. Consequently, the solution with $U < 0$ corresponds to a lower energy as compared to the solution with $U > 0$ for $A_y^2 > A_x^2$, while the solution with $U > 0$ (for $\gamma_v > 0$) brings about a deeper energy minimum for $A_y^2 < A_x^2$. Thus, the sign of minimizing *U* is chosen depending on the direction of polarization of the incident wave relative to the crystallographic axes.

2. Quantity *V* appears in the expression for free energy only as V^2; i.e., if a minimum exists for a certain *V*, a minimum of the same depth exists for $-V$ as well.

3. The mutual distribution of *U* and *V* depends on both terms; it can be observed that

(i) for $b_4 < 0$, the distribution $V = 0$, $U \neq 0$ is the one sought since this distribution minimizes both the terms considered here irrespective of the value of f_s; taking into account the consideration formulated in item 1, we can state that the $d - s$ phase is minimizing if the polarization of the electromagnetic wave is such that $A_y^2 > A_x^2$, while the $d + s$ phase is minimizing for $A_y^2 < A_x^2$;

(ii) for $b_4 > 0$, quantities *U* and *V* compete with each other: on the one hand, an increase in *U* leads to a decrease in the term $8e^2c^{-2}\gamma_v(A_y^2 - A_x^2)f_s U$; on the other hand, an increase in *V* leads to a decrease in the term $2b_4 f_s^2 (U^2 - V^2)$. Consequently, in this case we must take into account the factors in front of $U^2 - V^2$ and *U* in the terms we are interested in.

Let us recall that $\mathbf{A}$ is assumed to be small in the sense of relation (22). It should be noted that the factor in front of *U* contains the small quantity $(A_y^2 - A_x^2)f_s$, while the factor in front of $U^2 - V^2$ contains f_s^2. Consequently, if the value of f_s is large (which is observed in the range of parameters in which the $d \pm is$ phase is realized in Fig. 3), type II solutions exist. If, however, f_s is on the order of $\mathbf{A}$ (which takes place in the range of parameters corresponding to the pure *d* phase in Fig. 3), both the terms considered here are of the same order of smallness. In this case, a more detailed analysis is required to determine the type of solutions.

Since an analytic solution cannot be obtained for the given range of parameters, we carried out numerical simulation to calculate the position of local minima of F_ψ depending on temperature for various values of the parameters in the Ginzburg–Landau theory. Figure 5a shows the typical temperature dependence of the order parameters bringing about the minima of F_ψ for $b_4 > 0$. It can be seen that two second-order phase transitions take place (one between the normal phase and the $d - s$ phase and the other between the $d - s$ phase and the $d \pm e^{i\theta}s$ phase). Thus, the passage of current leads to a transition from the pure *d* phase to the $d \pm s$ phase and not to the $d \pm is$ phase, as could be expected, even for $b_4 > 0$. It should be noted that, as expected, the second phase transition occurs in the vicinity of the CD interface separating the pure *d* and the $d \pm s$ phases in Fig. 3. The simulation was also carried out for $b_4 < 0$ and confirmed our conclusions. The typical temperature dependences of the order parameter are shown in Fig. 5b. It can be seen that, as expected, the $d - s$ phase is realized at all temperatures.

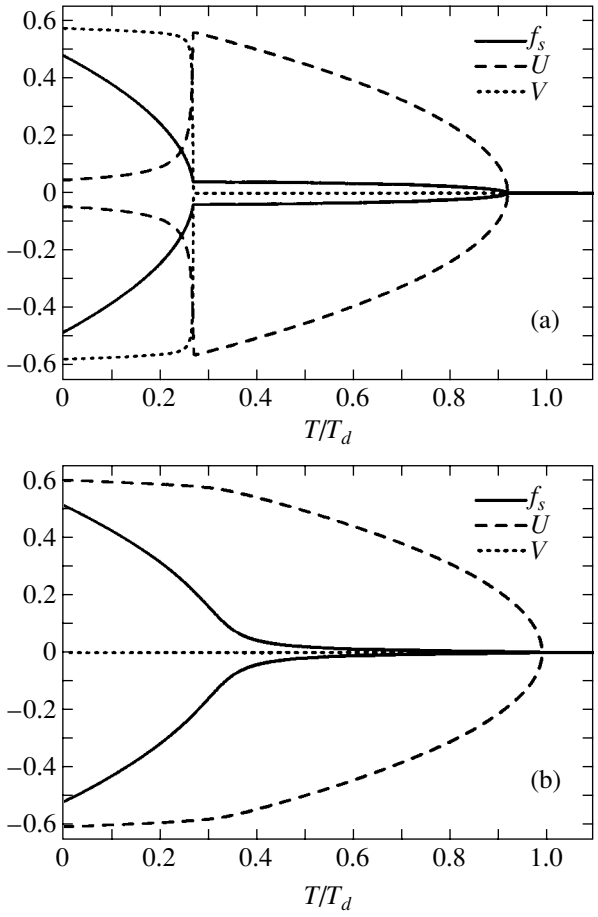

Fig. 5. Dependence of the order parameters on temperature $\tau = T/T_d$ for (a) $b_4 > 0$ and (b) $b_4 < 0$. The order parameters are measured in units of $\sqrt{\alpha_d T_d/b_2}$. The values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $b_3/b_2 = 3$, $b_4/b_2 = 1$ (a), –1 (b); $\gamma_s/\gamma_d = 1$, $\gamma_v/\gamma_d = 1$, $A\sqrt{4e^2\gamma_d/c^2\alpha_d T_d} = 0.3$ (a) and 0.1 (b); $T_s/T_d = 0.5$; the incident wave is polarized along the y axis.

In conclusion, we can state that, in the first order of perturbation theory, the phase transition between the pure d phase and the mixed d and s phase, which existed in the absence of current, vanishes for $b_4 < 0$ and is transformed into a phase transition between the $d - s$ phase and the $d \pm e^{i\theta_0} s$ phase for $b_4 > 0$. Thus, in the presence of current the following transformations take place:

(1) the $d \pm s$ phase is formed in the entire range corresponding to the pure d phase in the absence of current, the sign depending on the polarization of the incident wave; (2) in the region corresponding to the $d \pm s$ phase, this phase is preserved, but the plus or minus sign is chosen depending on the polarization of the incident wave; (3) in the region of the $d \pm is$ phase, the $d \pm e^{i\theta_0} s$ phase is formed.

For this reason, as the zeroth approximation, we must take

(1) solution (ii) with $V = 0$ in the region of the pure d phase in Fig. 3; (2) solution (iii) in the region of the $d \pm s$ phase in Fig. 3; (3) solution (iv) in the region of the $d \pm is$ phase in Fig. 3.

To avoid confusion, we will apply the term "phase transition" to phase transitions occurring in the absence of current, i.e., in the zeroth approximation.

Let us now find a quantitative solution to system (19)–(21) in the first order of perturbation theory and the constitutive relation for all regions on the phase diagram. We have proved that the phase diagram contains three different regions corresponding to different zeroth approximations used for solving the system. For this reason, analytic expressions for the nonlinear response in different regions on the phase diagram are different. We will carry out the subsequent analysis for each region separately and find the corresponding analytic formulas for the nonlinear response in each region. Then we will plot the temperature dependences of the nonlinear response for various values of b_4 to find out how the nonlinear response behaves when we cross the phase-transition lines. It should be noted that the choice of temperature as an independent variable and b_4 as a parameter is dictated only by the fact that temperature is a parameter that can easily be measured and varied in experiments.

3.1. Nonlinear Response in the Region of the Pure d Phase

In this case, solution (ii) with $V = 0$ (Fig. 3) is the zeroth approximation in solving the system. Since Eq. (21) is satisfied identically in this case, the system is transformed to

$$\left(a_s + \frac{4e^2}{c^2}\gamma_s \mathbf{A}^2\right) f_s + 2b_1 f_s^3 + (b_3 + 2b_4) f_s U^2 + \frac{4e^2}{c^2}\gamma_v (A_y^2 - A_x^2) U = 0, \tag{24}$$

$$\left(a_d + \frac{4e^2}{c^2}\gamma_d \mathbf{A}^2\right) U + 2b_2 U^3 + (b_3 + 2b_4) f_s^2 U + \frac{4e^2}{c^2}\gamma_v (A_y^2 - A_x^2) f_s = 0, \tag{25}$$

the zeroth approximation being $f_s = 0$, $U^2 = f_{d0}^2$, where $f_{d0} = \sqrt{-a_d/2b_2}$. Then, in the first order, Eq. (25) gives, to within the second-order infinitesimal,

$$U = -\mathrm{sgn}(A_y^2 - A_x^2)\sqrt{-\frac{a_d}{2b_2}}, \tag{26}$$

while Eq. (24) for f_s assumes the form

$$a_{s+}^* f_s + 2b_1 f_s^3 - \frac{4e^2}{c^2}\gamma_v\left|(A_y^2 - A_x^2)U\right| = 0, \qquad (27)$$

where

$$a_{s+}^* \equiv a_s - \frac{(b_3 + 2b_4)a_d}{2b_2} = 0.$$

This equation contains two (linear and cubic) terms with unknown f_s. If the coefficient of the linear term is large enough, the cubic term can be ignored in solving this equation. Indeed, in this case, we find that $f_s \propto \mathbf{A}^2$ is a small quantity; this means that $a_{s+}^* f_s \gg 2b_1 f_s^3$. However, such a solution is valid only when a_{s+}^* is not small. It can easily be proved that the value of a_{s+}^* in the given range of parameters can be large as well as small. For this purpose, we represent a_{s+}^* in the form

$$a_{s+}^* = \alpha_{s+}^*(T - T_{s+}^*),$$

where

$$\alpha_{s+}^* = \alpha_s - \alpha_d(b_3 + 2b_4);$$

in this case, $a_{s+}^* = 0$ at $T = T_{s+}^*$. Calculations show that temperature T_{s+}^* corresponds to the temperature of the AB phase transition on the phase diagram (see Fig. 3). Thus, in the range of parameters considered here, the temperature changes from T_{s+}^* to T_d, whereas a_{s+}^* varies from 0 to a certain large value. Consequently, we can ignore the cubic terms everywhere except in a narrow region in the vicinity of the AB transition curve. Since we are interested in the nonlinear response for any values of the parameters, we will solve Eq. (27) exactly.

It can easily be seen that Eq. (27) has only one real root; consequently, this equation can be effectively solved using the Cardan formula. In this case, the solution has the form

$$f_s = \sqrt[3]{\sqrt{\frac{f_{s1}^6}{4} + \frac{(a_{s+}^*)^3}{(6b_1)^3}} + \frac{f_{s1}^3}{2}} - \sqrt[3]{\sqrt{\frac{f_{s1}^6}{4} + \frac{(a_{s+}^*)^3}{(6b_1)^3}} - \frac{f_{s1}^3}{2}}, \qquad (28)$$

where

$$f_{s1} = \{2e^2c^{-2}\gamma_v b_1^{-1}\left|(A_y^2 - A_x^2)\right| f_{d0}\}^{1/3}.$$

Formulas (26) and (28), as well as the relation $V = 0$, determine the solution to system (19)–(21), i.e., the sought dependence of the order parameters on the vector potential. We must now substitute these formulas into expression (18) for current and thus find the constitutive relation. It should be noted that, since $f_s \ll |U|$, in Eq. (18) we have $|\gamma_s f_s^2 \mathbf{A}| \ll |2\gamma_v f_s U\bar{\mathbf{A}}|$ and can ignore the current of the *s* component as compared to the interference term. In this case, the constitutive relation can be written in the form

$$\mathbf{j} = -\frac{4e^2\gamma_d}{cb_2}|a_d|\mathbf{A} + \frac{16e^4}{c^3 b_2}\gamma_d^2\mathbf{A}^2\mathbf{A} + \frac{16e^2\gamma_v}{c}\mathrm{sgn}(A_y^2 - A_x^2) \times f_{d0}\left\{\left[\sqrt{\frac{(A_y^2 - A_x^2)^2 f_{d0}^2}{4e^{-2}\gamma_v^{-1}c^2 b_1^2} + \frac{(a_{s+}^*)^3}{(6b_1)^3}} + \frac{\left|A_y^2 - A_x^2\right| f_{d0}}{e^{-2}\gamma_v^{-1}c^2 b_1}\right]^{1/3} - \left[\sqrt{\frac{(A_y^2 - A_x^2)^2 f_{d0}^2}{4c^2 b_1^2 e^{-2}\gamma_v^{-1}} + \frac{(a_{s+}^*)^3}{(6b_1)^3}} - \frac{\left|A_y^2 - A_x^2\right| f_{d0}}{c^2 b_1 e^{-2}\gamma_v^{-1}}\right]^{1/3}\right\}\bar{\mathbf{A}}. \qquad (29)$$

Thus, we have derived the constitutive relation. It can be seen that it is the sum of three terms. The first two terms are associated with the current of the *d* component and could be obtained using the Ginzburg–Landau theory for the one-component *d* order parameter. The first term is linear in $\mathbf{A}$ and defines tensor $\overleftrightarrow{1/\lambda^2}$, which is equal to $(1/\lambda_d^2)\delta_{ik}$ in the given case, where

$$\lambda_d = \sqrt{\frac{c^2 b_2}{32\pi\gamma_d e^2|a_d|}}$$

is the London depth of penetration to a *d* superconductor. The second term is nonlinear in $\mathbf{A}$ and would determine $\overleftrightarrow{G}(\mathbf{A})$ in the case of a one-component *d* order parameter. The third term is determined by the interference term in Eq. (18) and is directly associated with the two-component nature of the order parameter.

Let us find the nonlinear response. In accordance with the technique described in Section 2, we must first of all find $\mathbf{j}_e$, viz., the source in Eq. (9), expressing $\mathbf{j}_e$ in terms of the linear current $\mathbf{j}_l$ in the film. For this purpose, we must express $\mathbf{A}$ in the formula of the constitutive relation in terms of $\mathbf{j}$ in a linear approximation and then substitute the obtained expression for $\mathbf{A}$ into the

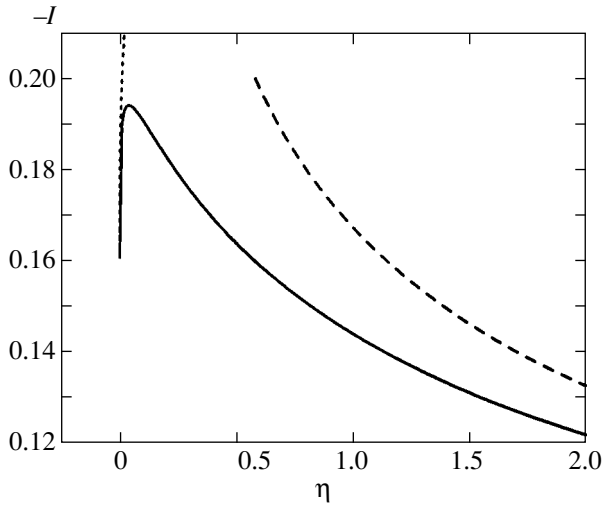

Fig. 6. Dependence $I(\eta)$. The solid curve is the result of numerical calculation, the dotted line is the asymptotic form for $\eta \longrightarrow 0$, and the dashed curve is the asymptotic form for $\eta \longrightarrow \infty$.

nonlinear part of the constitutive relation. This can be done as follows. In the linear approximation, from expression (29) for the current density we obtain $\mathbf{j}_l = -(c/4\pi\lambda_d^2)\mathbf{A}_l$; this current should be treated as preset (see Eq. (7)). Then $\mathbf{A}_l = -4\pi c^{-1}\lambda_d^2\mathbf{j}_l$. Substituting this expression into the nonlinear part of formula (29), we obtain

$$\mathbf{j}_e = -\frac{j_l^2}{j_{c1}^2(T)}\mathbf{j}_l - 2\,\mathrm{sgn}(A_y^2 - A_x^2) \times \frac{\gamma_v}{\gamma_d}\left\{\sqrt[3]{\sqrt{\mu^2+\nu}+\mu} - \sqrt[3]{\sqrt{\mu^2+\nu}-\mu}\right\}\bar{\mathbf{j}}_l, \tag{30}$$

where

$$\mu = \frac{1}{2}\frac{b_2}{b_1}\frac{\gamma_v}{\gamma_d}\frac{\left|j_{ly}^2 - j_{lx}^2\right|}{j_{c1}^2(T)}, \quad \nu = \frac{1}{27}\frac{b_2^3}{b_1^3}\frac{a_{s+}^{*3}}{|a_d|^3},$$

and $j_{c1}(T) = \sqrt{4e^2|a_d|^3\gamma_d/b_2^2}$ is the characteristic current density, which coincides in order of magnitude with the pair-breaking current in a superconductor with a one-component d order parameter. It is appropriate to recall here that the condition for the applicability of our perturbation theory is the weakness of nonlinear effects in the sense that the order parameter changes insignificantly in the presence of currents in comparison with its value in the absence of currents. In fact, this means that $\mathbf{j}_e$ must be much smaller than $\mathbf{j}_l$ since $\mathbf{j}_l$ is proportional to the order parameter in the absence of current, while $\mathbf{j}_e$ is proportional to the deviation of the order parameter from this value. Then it becomes clear from formula (30) that our analysis is valid when the linear current in the film is much smaller than the pair-breaking current for a superconductor with a one-component d order parameter. Let us first estimate the characteristic linear current emerging during reflection of an electromagnetic wave from the film. By way of example, we consider the experiment described in [17]. The characteristic power of the incident wave in these experiments was 0.1 W and the area on which this wave was incident was about 0.02 mm^2. Then the characteristic current density in a 1000-Å-thick film is 10^6 A/cm^2. Let us now estimate the critical pair-breaking current for HTSC materials. The typical London penetration depth for an HTSCs is 2000 Å, while the characteristic scale of variation of order parameter ξ is about 20 Å. We can easily calculate the characteristic pair-breaking current, which is found to be 10^8 A/cm^2. Thus, our basic assumption concerning the weakness of nonlinear effects is satisfactory.

In accordance with Section 2, we can now find $\mathbf{j}_{e3\omega}$ using formula (12). This gives

$$\mathbf{j}_{e3\omega} = \frac{1}{4}\frac{1}{(1-\tau)^3}\frac{j_0^2}{j_{c1}^2(0)}\mathbf{j}_0 - 2\,\mathrm{sgn}(A_y^2 - A_x^2)\frac{\gamma_v}{\gamma_d}\sqrt[3]{\mu}I(\eta)\bar{\mathbf{j}}_0, \tag{31}$$

where $\tau = T/T_d$ is the dimensionless temperature and

$$I(\eta) = \frac{1}{\pi}\int_0^{2\pi}\left\{\sqrt[3]{\sqrt{\sin^4 z + \eta} + \sin^2 z} - \sqrt[3]{\sqrt{\sin^4 z + \eta} - \sin^2 z}\right\}\sin z \sin 3z\, dz, \tag{32}$$

$$\eta = \frac{4}{27}\frac{b_2}{b_1}\frac{\gamma_d^2}{\gamma_v^2}\frac{\alpha_{s+}^{*3}}{\alpha_d^3}\frac{j_{c1}^4(0)}{(j_{0y}^2 - j_{0x}^2)^2}(\tau - \tau_{s+}^*)^3(1-\tau)^3. \tag{33}$$

Integral $I(\eta)$ cannot be evaluated analytically; however, for any η, the value of this integral can be determined numerically. We carried out numerical simulation and plotted the $I(\eta)$ dependence (Fig. 6). It can be seen that $-I(\eta)$ has a peak at $\eta \approx 0.04$. Consequently, the nonlinear response, which repeats the shape of the $-I(\eta)$ curve, has a peak in the vicinity of the phase-transition curve. It should be noted that such a behavior of the nonlinear response is associated with the interference term in the expression for current since integral $I(\eta)$ is associated precisely with this term. Thus, we can find the value of $\mathbf{j}_{e3\omega}$ for any η. Knowing the $\mathbf{j}_{e3\omega}(\mathbf{j}_0)$ dependence, we can find how the nonlinear response amplitude depends on the amplitude of the incident wave. We will not write here the formula for $\mathbf{E}_{3\omega}^{\mathrm{rad}}(\mathbf{E}_{\mathrm{inc}})$ since in

fact it repeats formula (31) and can be obtained by multiplying relation (31) by $-i12\pi\omega\lambda_d^2/c^2$ (see Eq. (13)) and substituting for $\mathbf{j}_0$ its expression in terms of $\mathbf{E}_{\text{inc}}$ (see Eq. (7)). We will give below the curves describing the temperature dependence of $\mathbf{E}_{3\omega}^{\text{rad}}$, plotted using formula (15), but we will now consider the $\mathbf{j}_{e3\omega}(\mathbf{j}_0)$ dependence obtained here since it is this dependence that contains most interesting information for our analysis.

Let us analyze relations (31)–(33) derived above. The control parameter in these equations is quantity $\eta \propto j_{c1}^4(0)/j_0^4$. Our analysis based on perturbation theory presumes that $j_0 \ll j_{c1}(T)$; consequently, $\eta \gg 1$ except in the range of parameters close to the phase-transition curve AB, where the value of η can be on the order of unity and even much smaller than unity due to factor $(\tau - \tau_{s+}^*)^3$. In a region on the phase diagram far away from AB, we can try to simplify the result using the fact that the value of η is large. We can prove that the value of $I(\eta)$ tends asymptotically to $-1/6\sqrt[3]{\eta}$ as $\eta \longrightarrow \infty$. Substituting this expression into Eq. (31), we obtain

$$\mathbf{j}_{e3\omega} = \frac{1}{4(1-\tau)^3}\frac{j_0^2}{j_{c1}^2(0)}\mathbf{j}_0 + \frac{1}{2}\frac{\alpha_d}{\alpha_{s+}^*}\frac{\gamma_\nu^2}{\gamma_d^2}\frac{1}{(\tau-\tau_{s+}^*)(1-\tau)^2}\frac{j_{0y}^2 - j_{0x}^2}{j_{c1}^2(0)}\bar{\mathbf{j}}_0; \tag{34}$$

this formula is valid in the entire range of parameters under investigation except the close neighborhood of the phase-transition curve AB. It should be noted that this relation could be derived without using the Cardan formula by approximately solving Eq. (27) if we ignore the term $2b_1 f_s^3$ in this equation. It can be seen that the increase in the linear response is associated with the presence of quantity $\tau - \tau_{s+}^*$ in the denominator of the second term in Eq. (34). This formula becomes invalid when the value of η becomes on the order of unity. In this case, we should use the numerical result. However, in the range of parameters close to the AB phase-transition curve, where $\eta \ll 1$, integral $I(\eta)$ can again be evaluated asymptotically. We can prove that the asymptotic value of this integral for $\eta \longrightarrow 0$ is

$$I(\eta) = \sqrt[3]{2}M_1 - M_2\sqrt[3]{\frac{\eta}{2}},$$

where

$$M_1 = \frac{1}{\pi}\int_0^{2\pi}\sin^{5/3}z\sin 3z\,dz \approx -0.127,$$

$$M_2 = \frac{1}{\pi}\int_0^{2\pi}\sin^{1/3}z\sin 3z\,dz \approx 0.231.$$

Substituting this relation into Eq. (31), we can write

$$\mathbf{j}_{e3\omega} = \frac{1}{4}\frac{j_0^2}{j_{c1}^2(T_{s+}^*)}\mathbf{j}_0 - 2M_1\frac{\gamma_\nu^{4/3}}{\gamma_d^{4/3}}\frac{b_2^{1/3}}{b_1^{1/3}}\frac{(j_{0y}^2 - j_{0x}^2)^{1/3}}{j_{c1}^{2/3}(T_{s+}^*)}\bar{\mathbf{j}}_0 + \frac{2}{3}M_2\frac{\gamma_\nu^{2/3}}{\gamma_d^{2/3}}\frac{b_2^{2/3}}{b_1^{2/3}}\frac{\alpha_{s+}^*(\tau-\tau_{s+}^*)}{\alpha_d(1-\tau_{s+}^*)}\frac{j_{c1}^{2/3}(T_{s+}^*)}{(j_{0y}^2 - j_{0x}^2)^{1/3}}\bar{\mathbf{j}}_0. \tag{35}$$

It should be noted that this formula could be derived by solving approximately Eq. (27) in perturbation theory for $a_{s+}^* f_s \ll 2b_1 f_s^3$. The second term in Eq. (35) was obtained in the zeroth order of such a perturbation theory (i.e., disregarding the term $a_{s+}^* f_s$), while the third term (linear in a_{s+}^*) was obtained in the first order, i.e., taking into account the term $a_{s+}^* f_s$ as a small correction. This relation will be helpful in subsequent analysis for constructing the temperature dependence of the nonlinear response; it is also interesting as such since it shows that the larger value of the nonlinear response in the vicinity of the phase transition AB is due to the nonanalytic nature of the $\mathbf{j}_e(\mathbf{j}_0)$ dependence at the point of this phase transition.

Let us now consider the relation between the polarizations of the incident, reflected, and re-emitted waves at the third harmonic. In our case, in the linear approximation, we have $\mathbf{j} \parallel \mathbf{A}$ and, hence, $\mathbf{j} \parallel \mathbf{E}$ also; consequently, the reflected wave is polarized in the same way as the incident wave. However, the nonlinear response has a different polarization due to the presence of the interference term in the expression for current. Expressions (31), (34), and (35) have a structure of the form

$$\mathbf{j}_{e3\omega} = H_1\mathbf{j}_0 + H_2\bar{\mathbf{j}}_0; \tag{36}$$

term $H_2\bar{\mathbf{j}}_0$ responsible for nonparallelism of $\mathbf{j}_{e3\omega}$ and $\mathbf{j}_0$ can be obtained from the interference term in Eq. (18). Let the incident wave be polarized at angle α to the x axis (it should be recalled that the x and y axes are chosen along the crystallographic axes of the film). Then we can easily find from Eq. (36) that the nonlinear response is polarized at angle ϕ such that

$$\tan\phi = \frac{1 + H_2/H_1}{1 - H_2/H_1}\tan\alpha. \tag{37}$$

Let us first consider the range of parameters far from the phase-transition curve AB. In this case, it is convenient to use formula (34) for calculations. Substituting the expressions for H_1 and H_2 corresponding to formula (34) into relation (37), we obtain

$$\tan\phi = \frac{1 - p(\tau)\cos 2\alpha}{1 + p(\tau)\cos 2\alpha}\tan\alpha, \tag{38}$$

where

$$p(\tau) = \frac{2\alpha_d\gamma_v^2(1-\tau)}{\alpha_{s+}^*\gamma_d^2(\tau-\tau_{s+}^*)}$$

is a temperature-dependent parameter on the order of unity. Thus, we have analytic formulas for the amplitude and polarization of the nonlinear response. In accordance with our plan, we will analyze these formulas later and will now find the nonlinear response in the next region on the phase diagram.

3.2. Nonlinear Response in the Region of the s ± d Phase

In this case, solution (iii) serves as the zeroth approximation in solving system (19)–(21) for the order parameters. We have $V = 0$; hence, analogously to the previous section, the system is simplified and transformed into system (24), (25). Bearing in mind that solution (iii) is the zeroth approximation and taking into account the above considerations concerning the sign of U, we must choose as the zeroth approximation the solution

$$f_s = \sqrt{\frac{2|a_{s+}^*|}{b_2\Delta}} \equiv f_{s0}, \tag{39}$$

$$U = -\mathrm{sgn}(A_y^2 - A_x^2)\sqrt{\frac{2b_1|a_{d+}^*|}{b_2^2\Delta}} \equiv U_0, \tag{40}$$

where

$$\Delta = \frac{4b_1}{b_2} - \frac{(b_3 - 2|b_4|)^2}{b_2^2}, \quad a_{d+}^* = a_d - \frac{(b_3+2b_4)a_s}{2b_1}.$$

In the first approximation, we seek the solution to system (24), (25) in the form $f_s = f_{s0} + f_s''$, $U = U_0 + U''$, where f_s'' and U'' are small corrections to the solutions in the zeroth approximation. Substituting these solutions into the system, we obtain, generally speaking, a certain nonlinear system in f_s'' and U'', viz., an analog of Eq. (27) for f_s' in the case of the pure d phase. It can easily be shown that, like in the previous section, the terms cubic in f_s'' and U'' must be taken into account only in the vicinity of the phase-transition curve AB. For this reason, we can confine our analysis to linear terms alone; in this case, the obtained solutions will not be valid in the vicinity of the AB phase transition. It will be shown below that a more exact solution in this case would improve the joining of the results on the right and on the left of the phase-transition curve only slightly and would not lead to any qualitative effects. Thus, taking into account only the terms linear in f_s'' and U'', we obtain the following linear system of equations for the corrections:

$$\begin{aligned} &4b_1 f_{s0}^2 f_s'' + 2(b_3+2b_4)U_0 f_{s0} U'' \\ &= -\frac{4e^2}{c^2}\gamma_s \mathbf{A}^2 f_{s0} - \frac{4e^2}{c^2}\gamma_v(A_y^2 - A_x^2)U_0, \end{aligned} \tag{41}$$

$$\begin{aligned} &4b_2 U_0^2 f_s'' + 2(b_3+2b_4)U_0 f_{s0} f_s'' \\ &= -\frac{4e^2}{c^2}\gamma_d \mathbf{A}^2 U_0 - \frac{4e^2}{c^2}\gamma_v(A_y^2 - A_x^2) f_{s0}. \end{aligned} \tag{42}$$

Solving this system, we obtain

$$\begin{aligned} f_s'' &= 2\frac{(b_3+2b_4)\gamma_d - 2b_2\gamma_s}{e^{-2}c^2 b_2^2 f_{s0}\Delta}\mathbf{A}^2 \\ &+ 2\gamma_v\frac{(b_3+2b_4)f_{s0}^2 - 2b_2U_0^2}{e^{-2}c^2 b_2^2 f_{s0}^2 U_0\Delta}(A_y^2 - A_x^2), \end{aligned} \tag{43}$$

$$\begin{aligned} U'' &= 2\frac{(b_3+2b_4)\gamma_s - 2b_1\gamma_d}{e^{-2}c^2 b_2^2 U_0\Delta}\mathbf{A}^2 \\ &+ 2\gamma_v\frac{(b_3+2b_4)U_0^2 - 2b_1 f_{s0}^2}{e^{-2}c^2 b_2^2 f_{s0} U_0^2\Delta}(A_y^2 - A_x^2). \end{aligned} \tag{44}$$

We must now substitute these relations for the order parameters into expression (18) for current. It should be noted that we cannot disregard in this case the current of the s component as it was done in Section 3.1 since f_s is not small. Substituting $f_s = f_{s0} + f_s''$ and $U = U_0 + U''$ into Eq. (18) and ignoring the terms quadratic in f_s'' and U'', we obtain the constitutive relation in the form

$$\begin{aligned} \mathbf{j} = -\frac{8e^2}{c}\{&\gamma_d U_0^2\mathbf{A} + \gamma_s f_{s0}^2\mathbf{A} + 2\gamma_v U_0 f_{s0}\bar{\mathbf{A}} \\ &+ 2\gamma_d U_0 U''\mathbf{A} + 2\gamma_s f_{s0} f_s''\mathbf{A} + 2\gamma_v U_0 f_s''\bar{\mathbf{A}} \\ &+ 2\gamma_v f_{s0} U''\bar{\mathbf{A}}\}. \end{aligned} \tag{45}$$

The first three terms describe the part linear in **A**, while the remaining term describe the nonlinear part. It can be seen that vector **j** is not parallel to **E** even in the linear approximation due to the fact that the interference term in this approximation differs from zero. It follows from Eq. (45) that tensor $\overleftrightarrow{1/\lambda^2}$ in this case is equal to

$(1/\lambda_d^2)\overleftrightarrow{T}$, where $\overleftrightarrow{T}$ is a diagonal tensor with the components

$$T_{xx} = \frac{4}{\Delta}\frac{\gamma_s\alpha_{s+}^*}{\gamma_d\alpha_d}\frac{\tau_{s+}^*-\tau}{1-\tau} + \frac{4}{\Delta}\frac{b_1\alpha_{d+}^*}{b_2\alpha_d}\frac{\tau_{d+}^*-\tau}{1-\tau}$$
$$- \operatorname{sgn}(A_y^2-A_x^2)\frac{8}{\Delta}\frac{\gamma_v\sqrt{b_1\alpha_{s+}^*\alpha_{d+}^*}}{\gamma_d\sqrt{b_2\alpha_d^2}}\sqrt{\frac{(\tau_{s+}^*-\tau)(\tau_{d+}^*-\tau)}{(1-\tau)^2}},$$

$$T_{yy} = \frac{4}{\Delta}\frac{\gamma_s\alpha_{s+}^*}{\gamma_d\alpha_d}\frac{\tau_{s+}^*-\tau}{1-\tau} + \frac{4}{\Delta}\frac{b_1\alpha_{d+}^*}{b_2\alpha_d}\frac{\tau_{d+}^*-\tau}{1-\tau}$$
$$+ \operatorname{sgn}(A_y^2-A_x^2)\frac{8}{\Delta}\frac{\gamma_v\sqrt{b_1\alpha_{s+}^*\alpha_{d+}^*}}{\gamma_d\sqrt{b_2\alpha_d^2}}\sqrt{\frac{(\tau_{s+}^*-\tau)(\tau_{d+}^*-\tau)}{(1-\tau)^2}},$$

where

$$\alpha_{d+}^* = \alpha_d - \frac{(b_3+2b_4)\alpha_s}{2b_1}.$$

The problem of nonlinear response in such a case was not considered in Section 2 and will be solved now. The general approach naturally remains the same. First of all, we must express the linear current in the film in terms of the incident wave amplitude. Then we must express **A** in terms of **j** in the linear approximation and substitute the result for **A** into the nonlinear part of the constitutive relation, thus determining the extraneous current $\mathbf{j}_e$. After this, we must solve the problem of emission of electromagnetic waves by this current. Implementing this plan, we will encounter some singularities, which will be considered when necessary. Let us express $\mathbf{j}_l$ in terms of $\mathbf{E}_{\text{inc}}$. We assumed before that the linear current in the film and the electric field in the incident wave are parallel and connected via relation (7). In the case of an anisotropic relation between **j** and **E**, this is not obvious. We have solved the problem of the reflection of a wave from the film for such a constitutive relation and found that relation (7) holds in all cases when the reflection coefficient is modulo close to unity irrespective of tensor $\overleftrightarrow{1/\lambda^2}$. Consequently, we can, as before, operate with linear current $\mathbf{j}_l$ as with a preset quantity, which repeats the behavior of $\mathbf{E}_{\text{inc}}$ to within a scalar factor.

Thus, we will find the nonlinear response. We must know $\mathbf{j}_e$, viz., the source in Eq. (9). For this purpose, we must express **A** in the constitutive relation in terms of **j** in the linear approximation and then substitute the obtained expression for **A** into the nonlinear part of the constitutive relation. We find that relation $\mathbf{j}_l = -(c/4\pi\lambda_d^2)\overleftrightarrow{T}\mathbf{A}_l$ is valid in the linear approximation. It will be convenient to express $\mathbf{A}_l$ in terms of $\mathbf{j}_l$ in the form

$$\mathbf{A}_l = -\frac{4\pi\lambda_d^2}{c}\sqrt{\frac{\sin^2\alpha}{T_{yy}^2}+\frac{\cos^2\alpha}{T_{xx}^2}}\,j_l\mathbf{m}, \tag{46}$$

where α is the angle between the direction of polarization of the incident wave and the *x* axis and **m** is a unit vector with coordinates

$$m_x = \frac{1}{T_{xx}\sqrt{1/T_{xx}^2+\tan^2\alpha/T_{yy}^2}},$$
$$m_y = \frac{\tan\alpha}{T_{yy}\sqrt{1/T_{xx}^2+\tan^2\alpha/T_{yy}^2}}. \tag{47}$$

Substituting this expression for $\mathbf{A}_l$ into the nonlinear part of constitutive relation (45), we obtain the following expression for extraneous current $\mathbf{j}_e$:

$$\mathbf{j}_e = B_1\left(\frac{\sin^2\alpha}{T_{yy}^2}+\frac{\cos^2\alpha}{T_{xx}^2}\right)^{3/2}\frac{j_l^3}{j_{c1}^2(T)}\mathbf{m}$$
$$+ B_2\left(\frac{\sin^2\alpha}{T_{yy}^2}-\frac{\cos^2\alpha}{T_{xx}^2}\right)\left(\frac{\sin^2\alpha}{T_{yy}^2}+\frac{\cos^2\alpha}{T_{xx}^2}\right)^{1/2}$$
$$\times\frac{j_l^3}{j_{c1}^2(T)}\mathbf{m} + B_3\left(\frac{\sin^2\alpha}{T_{yy}^2}+\frac{\cos^2\alpha}{T_{xx}^2}\right)^{3/2}\frac{j_l^3}{j_{c1}^2(T)}\overline{\mathbf{m}}$$
$$+ B_4\left(\frac{\sin^2\alpha}{T_{yy}^2}-\frac{\cos^2\alpha}{T_{xx}^2}\right)\left(\frac{\sin^2\alpha}{T_{yy}^2}+\frac{\cos^2\alpha}{T_{xx}^2}\right)^{1/2}\frac{j_l^3}{j_{c1}^2(T)}\overline{\mathbf{m}}, \tag{48}$$

where B_1, B_2, B_3, and B_4 are the following functions of the parameters of the Ginzburg–Landau theory:

$$B_1 = \frac{1}{\Delta}\left(2\frac{b_3+2b_4}{b_2}\frac{\gamma_s}{\gamma_d}-4\frac{b_1}{b_2}\right)$$
$$+\frac{1}{\Delta}\frac{\gamma_s}{\gamma_d}\left(2\frac{b_3+2b_4}{b_2}-4\frac{\gamma_s}{\gamma_d}\right),$$

$$B_2 = \frac{\gamma_v}{\gamma_d}\frac{1}{\Delta}\left\{\left(2\frac{b_3+2b_4}{b_2}-4\frac{\gamma_s}{\gamma_d}\right)\sqrt{\frac{b_1\alpha_{d+}^*}{b_2\alpha_{s+}^*}\frac{\tau_{d+}^*-\tau}{\tau_{s+}^*-\tau}}\right.$$
$$\left.+\left(2\frac{b_3+2b_4}{b_2}\frac{\gamma_s}{\gamma_d}-4\frac{b_1}{b_2}\right)\sqrt{\frac{b_2\alpha_{s+}^*}{b_1\alpha_{d+}^*}\frac{\tau_{s+}^*-\tau}{\tau_{d+}^*-\tau}}\right\},$$

$$B_3 = B_2,$$

$$B_4 = \frac{\gamma_v^2}{\gamma_d^2}\frac{1}{\Delta}\left\{4\frac{b_3+2b_4}{b_2}-2\frac{\alpha_{s+}^*}{\alpha_{d+}^*}\frac{\tau_{s+}^*-\tau}{\tau_{d+}^*-\tau}\right.$$
$$\left.-2\frac{b_1}{b_2}\frac{\alpha_{d+}^*}{\alpha_{s+}^*}\frac{\tau_{d+}^*-\tau}{\tau_{s+}^*-\tau}\right\}.$$

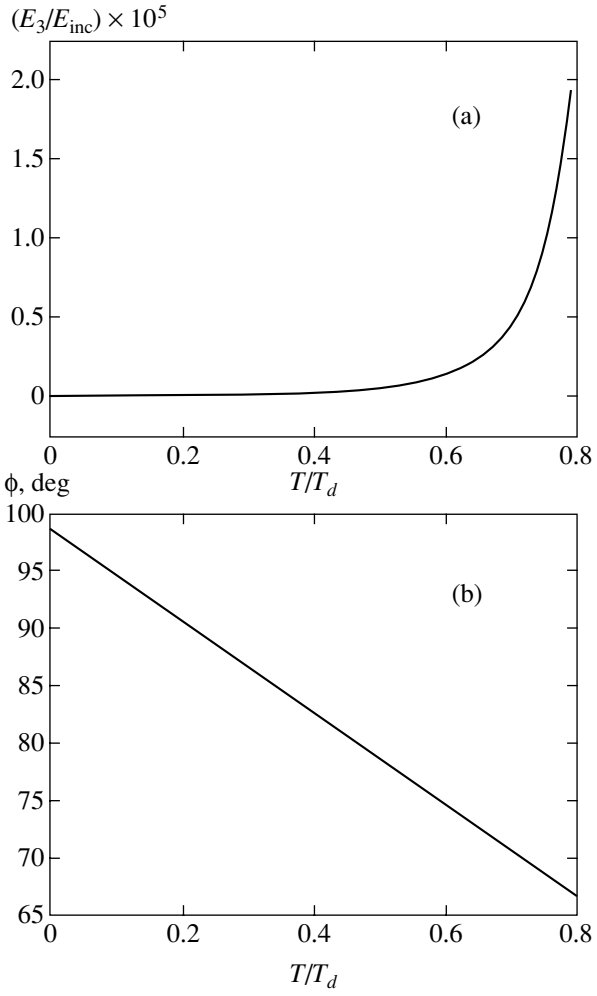

Fig. 7. Temperature dependence of (a) the amplitude and (b) the polarization of the nonlinear response in the case when the line along which the curve is plotted does not intersect the phase-transition curve. The nonlinear response amplitude $E_{3\omega}^{rad}$ is measured in amplitudes E_{inc} of the incident wave. The values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $b_3/b_2 = 3$, $b_4/b_2 = -0.25$, $\gamma_s/\gamma_d = 1$,$\gamma_v/\gamma_d = 0.6$ (a) and 1 (b), $T_s/T_d = 0.5$, $\alpha = 60°$; (a) $j_0/j_{c1}(0) = 0.05$, $\omega\lambda_d^2(0)/cd = 10^{-5}$.

In accordance with Section 2, we must find $\mathbf{j}_{e3\omega}$ using formula (12). It should be noted that in the given case $\mathbf{j}_e$ depends on time as $\sin^3\omega t$. Since

$$\frac{\omega}{\pi}\int_0^{2\pi/\omega} \sin^3\omega t \sin 3\omega t dt = -1/4,$$

to avoid cumbersome expressions, we can simply write

$$\mathbf{j}_{e3\omega} = -\frac{1}{4}\frac{\mathbf{j}_e}{\sin^3\omega t}. \quad (49)$$

Thus, we know the extraneous current emitting an electromagnetic wave at the third harmonic of the frequency of the incident wave. We must now find the amplitude of this wave. We solved this problem in Section 2 for $\overleftrightarrow{1/\lambda^2} = (1/\lambda^2)\delta_{ik}$ and derived formula (13). It can easily be shown that a generalization of this formula to the case of an anisotropic constitutive relation has the form

$$\mathbf{E}_{3\omega}^{rad} = -(12\pi i\omega/c^2)(\overleftrightarrow{1/\lambda^2})^{-1}\mathbf{j}_{e3\omega},$$

where $(\overleftrightarrow{1/\lambda^2})^{-1}$ is a tensor reciprocal to $\overleftrightarrow{1/\lambda^2}$. In fact, we must invert tensor $\overleftrightarrow{1/\lambda^2}$. We have already solved a similar problem in inverting dependence $\mathbf{j}_l(\mathbf{A}_l)$ and obtained formula (46). Using this formula, we can immediately write

$$\mathbf{E}_{3\omega}^{rad} = -i\frac{12\pi\omega\lambda_d^2}{c^2}\sqrt{\frac{\sin^2\phi_{e3\omega}}{T_{yy}^2} + \frac{\cos^2\phi_{e3\omega}}{T_{xx}^2}}\, j_{e3\omega}\mathbf{l}, \quad (50)$$

where $\mathbf{l}$ is a unit vector with the coordinates

$$l_x = \frac{1}{T_{xx}\sqrt{1/T_{xx}^2 + \tan^2\phi_{e3\omega}/T_{yy}^2}},$$
$$l_y = \frac{\tan\phi_{e3\omega}}{T_{yy}\sqrt{1/T_{xx}^2 + \tan^2\phi_{e3\omega}/T_{yy}^2}}, \quad (51)$$

and $\phi_{e3\omega}$ is the angle between vector $\mathbf{j}_{e3\omega}$ and the x axis. Substituting the obtained expression for $\mathbf{j}_{e3\omega}$ into Eq. (50), we can write the relation for the nonlinear response. We will not write these cumbersome relations here.

Let us analyze the relations derived above. It should be noted first of all that these relations become inapplicable in the vicinity of the AB phase transition since coefficients B_2 and B_4 contain quantities proportional to $1/(\tau_{s+}^* - \tau)$ and $\tau = \tau_{s+}^*$ on the phase-transition curve. Generally speaking, to derive correct formulas, we must take into account cubic terms in the system for f_s'' and U''. However, such a procedure cannot be carried out analytically since it leads to a sixth-order system that has no analytic solution. We will now try to prove that the inclusion of cubic terms would not give qualitatively new information. Let us consider a neighborhood of the AB phase transition. On the right of the transition curve, formula (35) holds, which implies an increasing temperature dependence for the nonlinear response with a large but finite slope (see Fig. 8a below). On the left of the phase-transition curve, the above relations for the nonlinear response are valid, which also give an increasing dependence for the nonlinear response in the vicinity of the phase transition, but become invalid exactly at the phase transition point.

Comparing these two approximations, we can say that, taking into account the cubic terms on the left on the transition exactly, we will only obtain a more exact coincidence of the temperature dependences on the left and on the right of the transition curve and a more exact coincidence of the slopes of these curves.

Let us now consider the relation between the polarizations of the incident, reflected, and re-emitted (at third harmonic) waves. In the given case, vector **j** is not collinear to vector **E** even in the linear approximation. However, calculations show that we can assume, to a high degree of accuracy, that the incident and reflected waves are polarized identically in view of almost total reflection of the incident wave. However, the nonlinear response is polarized differently. This is due to the fact that, first, vector $\mathbf{j}_e$ is not parallel to $\mathbf{j}_l$ and, second, vector $\mathbf{E}_{3\omega}^{\text{rad}}$ is not parallel to $\mathbf{j}_e$. Since $\mathbf{j}_e(\mathbf{j}_l)$ has the form (36), angle $\phi_{e3\omega}$ can be determined by formula (37). Using formula (51), we can prove that the slope ϕ of the electric field in the nonlinear response can be determined by the formula

$$\tan\phi = \frac{T_{xx}}{T_{yy}}\tan\phi_{e3\omega}. \tag{52}$$

Let us now find the nonlinear response in the next region on the phase diagram.

3.3. Nonlinear Response in the Region of the s ± id Phase

In this case, solution (iv) could be the zeroth approximation in solving system (19)–(21) for the order parameters. We have $V \neq 0$. However, we will not carry out analysis using perturbation theory since the system can be solved exactly for $V \neq 0$. Let us prove this. Equation (21) in this case can be divided by V, leading to the following relation between V and f_s:

$$V^2 = -\frac{b_3 - 2b_4}{2b_2} f_s^2 - \frac{a_d + 4e^2c^{-2}\gamma_d \mathbf{A}^2}{2b_2} - \frac{e^4\gamma_v^2(A_y^2 - A_x^2)^2}{c^4 b_4^2 f_s^2}. \tag{53}$$

Substituting this expression into Eq. (20), we can write the relation between U and f_s in the form

$$U = -\frac{e^2\gamma_v(A_y^2 - A_x^2)}{c^2 b_4 f_s}. \tag{54}$$

Substituting expressions (53) for V and (54) for U into the first equation (19) of the system, we obtain a simple equation for f_s, which gives

$$f_s^2 = \frac{2|a_{s-}^*|}{\Delta b_2} + \frac{8e^2}{c^2\Delta b_2}\left(\frac{b_3 - 2b_4}{2b_2} - \gamma_s\right)\mathbf{A}^2, \tag{55}$$

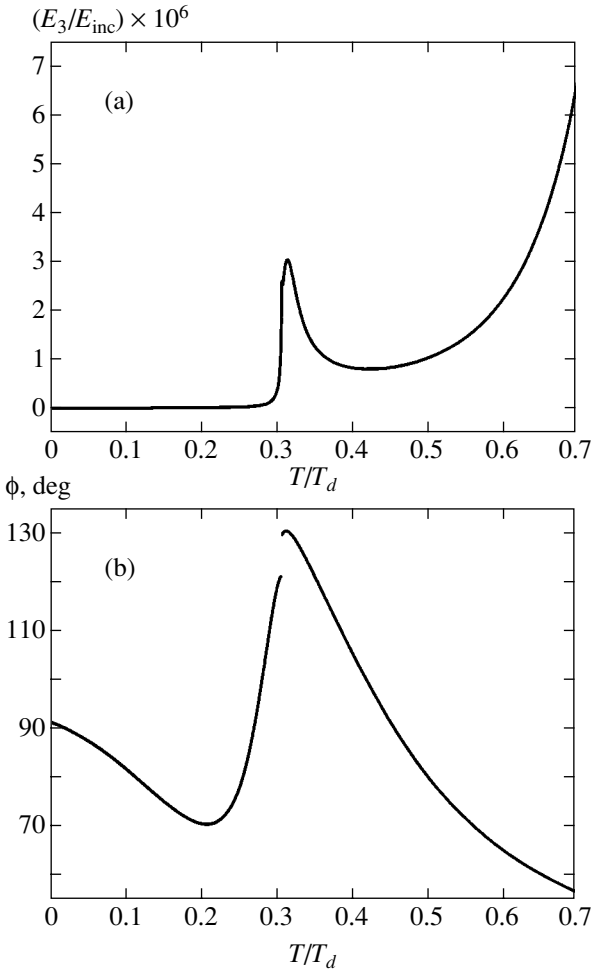

Fig. 8. Temperature dependence of (a) the amplitude and (b) the polarization of the nonlinear response in the case of intersection of the line along which the curve is plotted with the AB phase-transition curve. The nonlinear response amplitude $E_{3\omega}^{\text{rad}}$ is measured in amplitudes E_{inc} of the incident wave. The values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $b_3/b_2 = 3$, $b_4/b_2 = -1$ (a) and -0.25 (b); $\gamma_s/\gamma_d = 1$, $\gamma_v/\gamma_d = 0.6$, $T_s/T_d = 0.5$, $\alpha = 60°$; (a) $j_0/j_{c1}(0) = 0.05$, $\omega\lambda_d^2(0)/cd = 10^{-5}$.

where

$$a_{s-}^* = a_s - \frac{(b_3 - 2b_4)a_d}{2b_2}.$$

Substituting f_s back into Eqs. (53) and (54), we obtain the solution for V and U,

$$V^2 = \frac{2b_1}{\Delta b_2}\frac{|a_{d-}^*|}{b_2} + \frac{8e^2 b_1}{c^2\Delta b_2^2}\left(\frac{b_3 - 2b_4}{2b_2}\gamma_s - \gamma_d\right)\mathbf{A}^2, \tag{56}$$

$$U = -\frac{e^2\gamma_v(A_y^2 - A_x^2)}{c^2 b_4 f_s}, \tag{57}$$

where

$$a_{d-}^* = a_d - \frac{(b_3 - 2b_4)a_s}{2b_1}.$$

Thus, we managed to solve the system of equations for the order parameters exactly. To derive the constitutive relation, we substitute the results into expression (18) for current, which gives

$$\begin{aligned}\mathbf{j} = &-\frac{8e^2}{c}\left(\frac{2b_1}{\Delta b_2^2}\gamma_d|a_{d-}^*| + \frac{2}{\Delta b_2}\gamma_s|a_{s-}^*|\right)\mathbf{A}\\ &-\frac{64e^4}{c^4}\frac{b_1\gamma_d}{\Delta b_2^2}\left(\frac{b_3 - 2b_4}{2b_1}\gamma_s - \gamma_d\right)\mathbf{A}^2\mathbf{A}\\ &-\frac{64e^4}{c^4}\frac{\gamma_s}{\Delta b_2}\left(\frac{b_3 - 2b_4}{2b_2}\gamma_d - \gamma_s\right)\mathbf{A}^2\mathbf{A}\\ &+\frac{16e^4\gamma_v^2}{c^4 b_4}(A_y^2 - A_x^2)\bar{\mathbf{A}}.\end{aligned} \tag{58}$$

Thus, we have found the constitutive relation, which is the sum of four terms. The first term defines the linear relation between $\mathbf{j}$ and $\mathbf{A}$. It can be seen that the interference term in this case makes zero contribution to the linear part of the constitutive relation since $U = 0$ in the zeroth approximation. Then the current in the linear approximation is just the sum of the currents of the d and s components. Consequently, $\mathbf{j}_l$ is parallel to $\mathbf{A}_l$ and tensor $\overleftrightarrow{1/\lambda^2}$ is equal to $(1/\lambda_{sd}^2)\delta_{ik}$, where

$$\frac{1}{\lambda_{sd}^2} = \frac{32\pi e^2}{c^2}\left(\frac{2b_1}{\Delta b_2^2}\gamma_d|a_{d-}^*| + \frac{2}{\Delta b_2}\gamma_s|a_{s-}^*|\right).$$

The remaining terms determine the nonlinear part of the constitutive relation.

Let us now find the nonlinear response. First of all, we must determine $\mathbf{j}_e$. To do this, we will use the recommendations given in Section 2. We express $\mathbf{A}$ in terms of $\mathbf{j}$ in the linear approximation. Formula (58) yields

$$\mathbf{A}_l = -\frac{4\pi\lambda_{sd}^2}{c}\mathbf{j}_l.$$

Substituting this relation into the nonlinear part of Eq. (58), we can write the extraneous nonlinear current in the form

$$\begin{aligned}\mathbf{j}_e = &\left(\frac{4b_1\alpha_{d-}^*}{\Delta b_2\alpha_d}(\tau_{d-}^* - \tau) + \frac{4\gamma_s\alpha_{s-}^*}{\Delta\gamma_d\alpha_d}(\tau_{s-}^* - \tau)\right)^{-3}\\ &\times\left\{\left[\frac{4b_1}{b_2\Delta}\left(\frac{\gamma_s}{\gamma_d}\frac{b_3 - 2b_4}{2b_2} - 1\right)\right.\right.\\ &\left.+\frac{4\gamma_s}{\gamma_d\Delta}\left(\frac{b_3 - 2b_4}{2b_1} - \frac{\gamma_s}{\gamma_d}\right)\right]\frac{j_l^2}{j_{c1}^2(0)}\mathbf{j}_l\\ &\left.-\frac{b_2\gamma_v^2}{2b_4\gamma_d^2}\frac{j_{ly}^2 - j_{lx}^2}{j_{c1}^2(0)}\bar{\mathbf{j}}_l\right\}.\end{aligned} \tag{59}$$

We must now find the Fourier transform of $\mathbf{j}_e$. It should be noted that $\mathbf{j}_e$ depends on time as $\sin^3\omega t$. Consequently, to find $\mathbf{j}_{e3\omega}$, we must simply use formula (49), which implies that $\mathbf{j}_{e3\omega}$ repeats the amplitude of $\mathbf{j}_e$ to within a constant factor. Knowing dependence $\mathbf{j}_{e3\omega}(\mathbf{j}_0)$, we can find the dependence of the nonlinear response amplitude on the amplitude of the incident wave. We will not write here the formula for $\mathbf{E}_{3\omega}^{\mathrm{rad}}(\mathbf{E}_{\mathrm{inc}})$ since it in fact repeats formula (59) and can be derived by multiplying Eq. (59) by $-i12\pi\omega\lambda_{sd}^2/c^2$ (see Eq. (13)) and substituting for $\mathbf{j}_0$ its expression in terms of $\mathbf{E}_{\mathrm{inc}}$ (see Eq. (7)).

In the given case, the reflected wave is polarized in the same way as the incident wave, but the nonlinear response has a different polarization. Since the nonlinear response is polarized in the same way as $\mathbf{j}_{e3\omega}$, it is sufficient to determine the direction of current $\mathbf{j}_{e3\omega}$ to find the nonlinear response polarization. Formula (59) for $\mathbf{j}_{e3\omega}$ has the form (36); consequently, to find the slope of the electric field in the nonlinear response to the x axis, we can use formula (37). This gives

$$\tan\phi = \frac{1 - p_2\cos 2\alpha}{1 + p_2\cos 2\alpha}\tan\alpha, \tag{60}$$

where

$$\begin{aligned}p_2 = &-\frac{b_2\gamma_v^2}{2b_4\gamma_d^2}\left\{\frac{4b_1}{b_2\Delta}\left(\frac{\gamma_s}{\gamma_d}\frac{b_3 - 2b_4}{2b_2} - 1\right)\right.\\ &\left.+\frac{4\gamma_s}{\gamma_d\Delta}\left(\frac{b_3 - 2b_4}{2b_1} - \frac{\gamma_s}{\gamma_d}\right)\right\}^{-1}\end{aligned} \tag{61}$$

is a temperature-independent parameter on the order of unity.

Thus, we have found the analytic expression for the nonlinear response on the entire phase diagram. The

expressions derived above are rather cumbersome and can hardly be used to determine the behavior of the nonlinear response at various points on the phase diagram. For this reason, we will analyze the temperature dependences of the nonlinear response for various values of b_4 since these dependences can be plotted as graphs. In this way, we will be able to find the nonlinear response in all regions and in the vicinity of phase-transition curves on the phase diagram.

3.4. Temperature Dependence of the Nonlinear Response

In this section, we consider the temperature dependences of the nonlinear response for various values of b_4. In fact, we analyze the variation of the nonlinear response on the constant-b_4 line. Since the constant-b_4 line can intersect various regions on the phase diagram, we must use the results obtained for the nonlinear response in these regions. To cover all regions on the phase diagram, we must generally consider the temperature variation from temperatures higher than T_d to 0. At $T > T_d$, the nonlinear response is zero since the order parameter is zero in this case. In the range of temperatures smaller than but quite close to T_d, our theory is invalid since the main approximation of the theory is that the magnetic field of the incident wave changes the order parameter only slightly as compared to its value in zero magnetic field. This condition is obviously violated in the vicinity of T_d since the order parameter in this region is small. Consequently, our results are valid beginning from temperatures not very close to and lower than T_d. For this reason, we will study the temperature dependences of the nonlinear response in a temperature range from zero to temperatures lower than but close to T_d. It follows from the phase diagram that there are three different cases. The first case corresponds to values of b_4 such that the constant-b_4 line in Fig. 3 does not intersect any phase-transition curve; in the second case, the constant-b_4 line intersects the AB phase-transition curve; in the third case, the constant-b_4 line intersects the CD phase-transition curve.

Let us consider the first case. Formula (34) can be successfully used for determining the nonlinear response since there are no regions close to phase transitions on the entire span of the line b_4 = const. Let us plot the temperature dependence of the electric field amplitude in the nonlinear response using formula (15) for this purpose. It should be noted that quantity $\omega\lambda^2/cd$ appearing in this formula can be represented in this case as

$$\frac{\omega\lambda^2}{cd} = \frac{\omega\lambda_d^2(0)}{cd}\frac{1}{1-\tau}, \qquad (62)$$

where $\lambda_d^2(0)$ is the London penetration depth for a *d*-type superconductor at $T = 0$. To derive the temperature dependence of the nonlinear response, we must multiply formula (34) by $(1-\tau)^{-1}$ and by a certain constant factor. Assuming that $\omega\lambda_d^2(0)/cd = 10^{-5}$, we plot the temperature dependence of $\mathbf{E}_{3\omega}^{\mathrm{rad}}$. We assume that the amplitude of the incident wave is such that $j_0/j_{c1}(0) = 0.05$; i.e., $E_{\mathrm{inc}} = 0.05 j_{c1}(0) \cdot 2\pi d/c$. In addition, we assume that the incident is polarized at an angle of 60°. The temperature dependence is found to be a smoothly descending curve (Fig. 7a) without kinks. Using formula (38), we plot the temperature dependence of the nonlinear response polarization (Fig. 7b). At temperatures close to T_d, the polarization virtually coincides with polarization α of the incident wave; as the temperature decreases, angle ϕ increases smoothly, passes through 90°, and goes to the second quadrant. Thus, at a certain temperature, the nonlinear response is polarized along the crystal axis in spite of the fact that the incident wave is polarized at a certain angle to this axis.

Let us consider the second case, when the line b_4 = const (along which we plot the temperature dependences) intersects the AB phase-transition curve. In this case, we must use relation (31) for $\mathbf{j}_{e3\omega}$ on the right of the phase-transition curve and then formula (13) for the nonlinear response analogously to the previous case. On the left of the phase-transition curve, we must use formulas (48) and (49) for $\mathbf{j}_{e3\omega}$ and then formula (50) for the nonlinear response. Thus, we obtain a curve consisting of three pieces (Fig. 8a). It can be seen that the temperature dependence of the nonlinear response in this case has a peak very close to the phase transition. From the point of view of mathematics, this peak is associated with divergence of the solution to Eq. (27) in the linear approximation in small quantity f_s in the vicinity of the phase transition. If we take into account the cubic term, the divergence disappears and we obtain just a sharp peak near $a_{s+}^* = 0$. The fact that the peak is very sharp is associated, from the standpoint of mathematics, with nonanalyticity of dependence $\mathbf{j}_e(\mathbf{j}_0)$ in the vicinity of the AB phase transition. The situation described above corresponds to resonance in a certain oscillatory system. Temperature variation corresponds to a change in the natural frequency of the oscillatory system and the inclusion of nonlinear terms corresponds to the limitation imposed by nonlinearity on the resonance amplitude. From the standpoint of physics, the peak is associated with the smallness of the restoring force in the system at the phase-transition point, since the expansion of the free energy in the order parameter begins with fourth-order terms in contrast to points that do not correspond to the phase transition. Figure 8b shows the temperature dependence of the polarization angle of the nonlinear response. It can be seen that this dependence has a clearly pronounced peak in the vicinity of the AB phase transition. Thus, both the amplitude and the polarization of the nonlinear

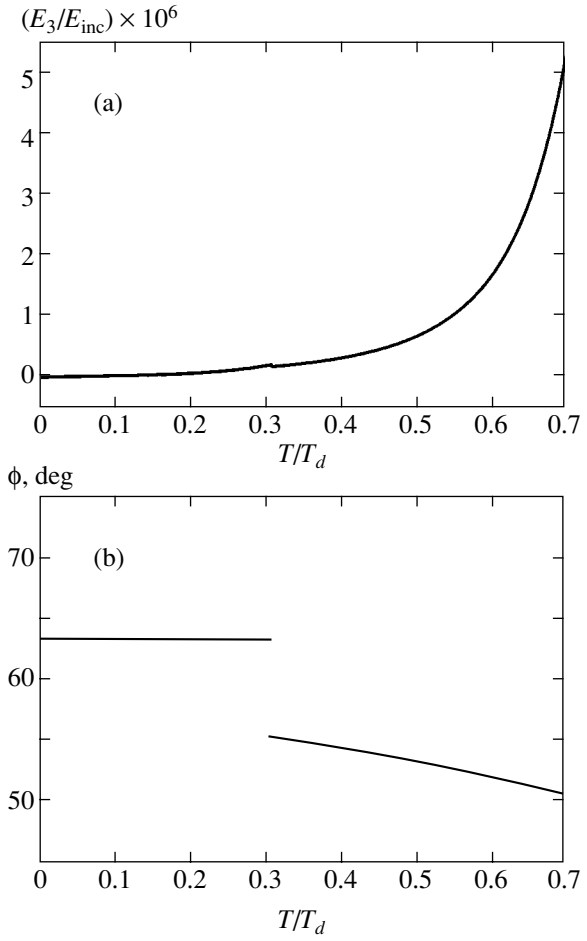

Fig. 9. Temperature dependences of (a) the amplitude and (b) the polarization of the nonlinear response in the case when the line along which the curve is plotted intersects the CD phase-transition curve. The nonlinear response amplitude $E_{3\omega}^{rad}$ is measured in amplitudes E_{inc} of the incident wave. The values of parameters: $\alpha_s/\alpha_d = 1.8$, $b_1/b_2 = 1$, $b_3/b_2 = 3$, $b_4/b_2 = 1$, $\gamma_s/\gamma_d = 1$, $\gamma_v/\gamma_d = 0.6$, $T_s/T_d = 0.5$, $\alpha = 60°$; (a) $j_0/j_{c1}(0) = 0.05$, $\omega\lambda_d^2(0)/cd = 10^{-5}$.

response are sensitive to the $d \longleftrightarrow d \pm s$ phase transition.

Let us consider the third case, when the line $b_4 =$ const intersects the CD phase-transition curve. Here, we must use formula (34) for the nonlinear response and formula (15) for the nonlinear response amplitude on the right of the phase transition. For the region on the left of the phase transition, we must use relation (59) for $\mathbf{j}_e$ and then formulas (49) for $\mathbf{j}_{e3\omega}$ and (15) for the nonlinear response. Thus, we obtain the temperature dependence of the nonlinear response, consisting of two pieces (Fig. 9a). It can be seen that the nonlinear response amplitude smoothly decreases with temperature and is virtually unaffected by the phase transition. It should be noted that, from the standpoint of mathematics, this is due to the following circumstance. In contrast to the situation with the AB phase transition, formula (34) becomes inapplicable in the vicinity of the CD phase transition not because of the requirement that the cubic term must be taken into account in expression (27), but due to the transition of the system to a state with $V \neq 0$. For this reason, the temperature dependence on the right of the CD phase transition decreases monotonically with temperature down to a very close neighborhood of the transition. Analogously, the temperature dependence on the left of the phase-transition curve also increases monotonically with temperature up to a very close neighborhood of the transition. We cannot say what happens in the vicinity of the phase transition, but the above considerations and close matching of the curves in Fig. 9a suggest that the curve has no spikes or other singularities at the phase-transition line. Thus, the nonlinear response amplitude in the given case is insensitive to the phase transition. However, the polarization angle may be sensitive to the CD phase transition. Let us prove this. In the present case, we can use formula (37) for the polarization angle, where H_2 and H_1 must be determined from relation (59). It should be noted that formula (59) implies that H_2 and H_1 are independent of temperature. Consequently, the polarization angle is also independent of temperature for $T < T_{s-}^*$. At the same time, the polarization depends on temperature for $T > T_{s-}$ in accordance with formula (38). Figure 9b shows the temperature dependence of the polarization angle for the same values of parameters, which we used for plotting the nonlinear response amplitude. It should be noted that the discontinuity on this curve is due to inapplicability of our theory in the vicinity of the phase transition. Apparently, the actual temperature dependence of the nonlinear response polarization should not have any discontinuities. Consequently, we can expect that, in the vicinity of the phase transition, the sloping curve continuously transforms into a horizontal line.

Thus, we have studied the temperature dependence of the amplitude of the nonlinear response as well as its polarization in the cases when various phase transitions take place. It turns out that both the amplitude of the nonlinear response and its polarization are very sensitive to the $d \longleftrightarrow d \pm s$ phase transition. The nonlinear response amplitude is found to be insensitive to the $d \longleftrightarrow d \pm is$ phase transition, while the polarization is weakly sensitive to this transition. Let us now consider our results from the standpoint of their experimental observation.

4. DISCUSSION

We have determined the nonlinear response of a superconductor with a two-component *s* and *d* order parameter in the framework of the Ginzburg–Landau theory. Our main result is the presence of a sharp peak in the nonlinear response in the vicinity of the $d \longleftrightarrow d \pm s$ phase transition. We have proved that this peak, as

well as the sign reversal of the derivative of the polarization angle with respect to temperature, is associated with the interference term in the expression for current. We have also demonstrated that the nonlinear response amplitude is weakly sensitive to the $d \longleftrightarrow d \pm is$ phase transition and behaves in the vicinity of this transition in the same way as the nonlinear response of a simple *d* superconductor. However, the nonlinear response polarization is quite sensitive to this phase transition also since the polarization is independent of temperature on the left of this phase transition, while the temperature dependence of polarization on the right of it is quite strong.

Let us consider the results from the standpoint of their experimental observation. First of all, we consider the absolute values of the nonlinear response amplitude. In plotting the graphs for the nonlinear response amplitude, we used the values of the quantities from experimental study [17], in which the nonlinear response amplitude was measured as a function of temperature, constant magnetic field, and the incident wave amplitude. Consequently, we can compare our theoretical results with the experimental results obtained in [17]. We found that the amplitude of the nonlinear response is smaller than the amplitude of the linear response by 5–6 orders of magnitude. Thus, the characteristic value of the nonlinear response power is 10–12 orders of magnitude smaller than the power of the incident wave. The nonlinear response power determined in [17] at temperatures not very close to the superconducting transition temperature was 10^{-11} W for an incident wave power of 0.1 W. Thus, the coincidence is quite satisfactory. In addition, it was found in the experiments that the temperature dependence of the nonlinear response has a peak at temperatures close to 40 K. Our result also predicts the presence of a peak on the nonlinear response curve in the vicinity of the $d \longleftrightarrow d \pm s$ phase transition; however, the shape of the theoretical peak is sharper than observed in the experiments. It should be noted that the sharp peak obtained under ideal conditions could be blurred, for example, due to sample inhomogeneities or for some other reasons. Consequently, we conclude that, first, the nonlinear response at moderate temperatures may be associated with the Ginzburg–Landau mechanism of nonlinearity and, second, the sample may have a two-component *d* and *s* order parameter.

Let us now consider the problems associated with observation of nonlinear response polarization. It is well known [14] that YBaCuO can be a polycrystal or a single crystal. In the case of a perfect single crystal, all our results are valid without any changes. In the case of a polycrystal, when the entire sample consists of a large number of crystallites, for which the *a* and *b* axes are disoriented relative to each other, while the *c* axis is the same, our results are also applicable, but must be averaged over all possible orientations of the *a* and *b* axes. In averaging our results, we must assume that vector $\mathbf{E}_{\text{inc}}$ remains unchanged and that the coordinate system rotates about the *z* axis. Let us consider such an averaging procedure using formula (34) as an example. This formula has the form

$$\mathbf{j}_{e3\omega} = s_1\mathbf{j}_0 + s_2\bar{\mathbf{j}}_0\cos 2\alpha.$$

Averaging this formula as indicated above, we find that $\langle \mathbf{j}_{e3\omega} \rangle$ has a component in the $\mathbf{j}_0$ direction and has no component orthogonal to $\mathbf{j}_0$. The component of $\langle \mathbf{j}_{e3\omega} \rangle$ in the direction of $\mathbf{j}_0$ has the form

$$\langle \mathbf{j}_{e3\omega} \rangle_{\text{inc}} = \left(s_1 - \frac{1}{2}s_2\right)j_0. \tag{63}$$

Thus, averaged formula (34) has the form

$$\mathbf{j}_{e3\omega} = \left\{\frac{1}{4(1-\tau)^3} + \frac{1}{4}\frac{\alpha_d}{\alpha^*_{s+}}\frac{\gamma_v^2}{\gamma_d^2}\frac{1}{(\tau-\tau^*_{s+})(1-\tau)^2}\right\}\frac{j_0^2}{j_{c1}^2(0)}\mathbf{j}_0. \tag{64}$$

It can be seen that the temperature dependence of the nonlinear response amplitude does not differ substantially from that for a single crystal. At the same time, the polarization of the nonlinear response simply coincides with the incident wave polarization in contrast to the case of a single crystal. The physical meaning of this phenomenon is that the only preferred direction in a polycrystal is the direction of polarization in the incident wave. If the nonlinear response differs from zero, it can be polarized only along the electric field in the incident wave. Consequently, all effects associated with the difference in the polarizations of the incident wave and the nonlinear response disappear upon averaging. As a result, it turns out that the $d \longleftrightarrow d \pm is$ transition in a polycrystal cannot be observed from analysis of the nonlinear response. At the same time, at the $d \longleftrightarrow d \pm s$ transition point, a peak of the nonlinear response amplitude is observed for single crystals as well as for polycrystals; i.e., this phase transition can be observed from analysis of the nonlinear response in the case of a polycrystal as well.

It would also be interesting to consider the effect of various inhomogeneities. We used the model of a homogeneous superconductor, assuming that the crystal lattice is perfect. However, real HTSC samples may be strongly inhomogeneous due to the fact that YBaCuO can form several different metastable phases. This leads to the emergence of additional nonlinearity mechanisms associated with Josephson coupling between grains as well as with fast flowing of current to the grains during their transition to the superconducting state. These mechanisms can also lead to the emergence of a low-temperature peak in the nonlinear response,

which is analogous to the peak emerging in our theory. Consequently, the problem of experimental method for separating the nonlinear responses associated with these different mechanisms arises. Since our theory is based on the crystal structure itself, the results should not depend qualitatively on the presence or absence of large-scale inhomogeneities, nor on their positions. For this reason, we can single out a part of the nonlinear response associated with the structure of the order parameter by preparing several samples using the same technique. The part of the nonlinear response associated with inhomogeneities will change since it is impossible to prepare two samples with the same distribution of inhomogeneities, while the part associated with the order parameter will remain unchanged.

In conclusion, let us compare the potentialities of the linear and nonlinear diagnostics of phase transitions. We have established that the nonlinear response is sensitive to the superconducting phase transition as well as to other phase transitions that may exist in a superconductor with a complex order parameter. We can compare these results with the results for the linear response. It is well known that the linear response is also very sensitive to the superconducting phase transition. Let us see whether it is sensitive to additional phase transitions. The formulas derived by us in the linear approximation readily show that the amplitude of the reflected wave does not respond to additional phase transitions. The polarization of the reflected wave is sensitive in principle to the $d \longleftrightarrow d \pm s$ phase transition; however, first, this effect is very weak due to almost total reflection of the incident weave and, second, it will disappear in all probability after averaging in the case of a polycrystal. Summarizing the result, we can conclude that nonlinear diagnostics may reveal additional phase transitions, while linear diagnostics cannot be used for this purpose.

ACKNOWLEDGMENTS

We are grateful to Yu.N. Nozdrin and A.S. Mel'nikov for fruitful discussions and to E.E. Pestov for detailed explanation of the experimental technique.

This study was financed by the Russian Foundation for Basic Research (project no. 03-02-16533).

REFERENCES

1. C. C. Tsuei and J. R. Kirtley, Rev. Mod. Phys. **72**, 969 (2000).
2. D. J. Van Harlingen, Rev. Mod. Phys. **67**, 515 (1995).
3. T. Strohm and M. Cardona, Solid State Commun. **104**, 233 (1997).
4. H. Srikanth, B. A. Willemsen, T. Jacobs, *et al.*, Phys. Rev. B **55**, 14733 (1997).
5. A. G. Sun, D. A. Gajewski, M. B. Maple, and R. C. Dynes, Phys. Rev. Lett. **72**, 2267 (1994).
6. J. Ma, C. Quitmann, R. J. Kelley, *et al.*, Science **267**, 862 (1995).
7. A. E. Ruckenstein, P. J. Hirschfeld, and J. Appel, Phys. Rev. B **36**, 857 (1987).
8. G. Kotliar, Phys. Rev. B **37**, 3664 (1988).
9. Q. P. Li, B. E. C. Koltenbah, and R. Joynt, Phys. Rev. B **48**, 437 (1993).
10. J. H. Xu, L. Shen, J. H. Miller, and C. S. Ting, Phys. Rev. Lett. **73**, 2492 (1994).
11. J. F. Annett, Adv. Phys. **39**, 83 (1990).
12. R. Heeb, A. van Otterlo, M. Sigrist, and G. Blatter, Phys. Rev. B **54**, 9385 (1996).
13. A. S. Mel'nikov, I. M. Nefedov, D. A. Ryzhov, *et al.*, Phys. Rev. B **62**, 11820 (2000).
14. N. M. Plakida, *High-Temperature Superconductivity* (Springer, Berlin, 1995; Mezhdunarodnaya Programma Obrazovaniya, Moscow, 1996).
15. Yong Ren, Ji-Hai Xu, and C. S. Ting, Phys. Rev. Lett. **74**, 3680 (1995).
16. M. Zapotovsky, D. L. Maslov, and P. M. Goldbart, Phys. Rev. B **55**, 6599 (1997).
17. A. Yu. Aladyshkin, A. A. Andronov, E. E. Pestov, *et al.*, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **46**, 123 (2003).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 597–607.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 674–685.

ELECTRONIC PROPERTIES OF SOLIDS

Thermal Correction to Resistivity in Dilute Si-MOSFET Two-Dimensional Systems¶

M. V. Cheremisin

Ioffe Institute for Physics and Technology, St. Petersburg, 194021 Russia
e-mail: maksim.vip1@pop.ioffe.rssi.ru

Received April 15, 2003; in final form, April 24, 2004

Abstract—Neglecting electron–electron interactions and quantum interference effects, we calculate the classical resistivity of a two-dimensional electron (hole) gas, taking into account the degeneracy and the thermal correction due to the combined Peltier and Seebeck effects. The resistivity is found to be a universal function of the temperature, expressed in units of $(h/e^2)(k_F l)^{-1}$. Analysis of the compressibility and thermopower points to the thermodynamic nature of the metal–insulator transition in two-dimensional systems. We reproduce the beating pattern of Shubnikov–de Haas oscillations in both the crossed field configuration and Si-MOSFET valley splitting cases. The consequences of the integer quantum Hall effect in a dilute Si-MOSFET two-dimensional electron gas are discussed. The giant parallel magnetoresistivity is argued to result from the magnetic-field-driven disorder. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Recently, much interest has been focused on the anomalous transport behavior of a variety of low-density two-dimensional (2D) systems [1–5]. It has been found that below some critical density, cooling causes an increase in resistivity, whereas in the opposite, high-density case, the resistivity decreases. Another property of dilute 2D systems is their unusual response to the parallel magnetic field. At low temperatures, the magnetic field was found to suppress the metallic behavior and result in increasing the resistivity upon enhancement of the spin polarization degree [6, 7]. A strong perpendicular magnetic field, if applied simultaneously with the parallel one, results in suppression of the parallel magnetoresistivity [8]. Although numerous theories have been put forward to account for these effects, the origin of the above behavior is still the subject of a heated debate.

The ohmic measurements are known to be carried out at a low current ($I \longrightarrow 0$) in order to prevent Joule heating. In contrast to the Joule heat, the Peltier and Thomson effects are linear in the current. As shown in [9–11], the Peltier effect results in a correction to the measured resistance. When the current is running, one of the sample contacts is heated, and the other is cooled because of the Peltier effect. The contact temperatures are different. The voltage drop across the circuit includes thermoelectromotive force, which is linear in the current. There exists a thermal correction $\Delta\rho$ to the ohmic resistivity ρ of the sample. For low-density 2D electron gas (2DEG), the correction may be comparable to the resistivity because $\Delta\rho/\rho \sim (kT/\mu)^2$, where μ is the Fermi energy. In the present paper, we report on a study of low-T transport in 2D systems, taking both the carrier degeneracy and the Peltier-effect-induced correction to resistivity into account.

2. GENERAL FORMALISM

For clarity, we consider the (100) Si-MOSFET 2DEG system. Within the strong inversion regime, we further neglect a depletion layer charge in the semiconductor bulk. At a fixed gate voltage, the quasi-Fermi level μ in the semiconductor is shifted with respect to that in the metal gate. The number of occupied states below the quasi-Fermi level determines the density of electrons assumed to occupy the first quantum-well subband with the isotropic energy spectrum $\varepsilon(\mathbf{k}) = \hbar^2 k^2/2m$.

We consider a sample connected to the current source by means of two identical leads (Fig. 1). Both contacts are ohmic. The voltage is measured between the open ends c and d kept at the temperature of the external thermal reservoir. The sample is placed in a chamber with the mean temperature T_0. According to our basic assumption, the contacts a and b may have different respective temperatures, T_a and T_b. Including

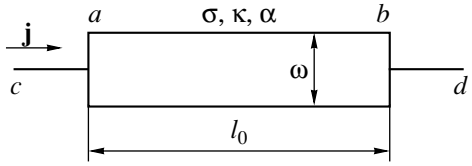

Fig. 1. The experimental setup.

¶ This article was submitted by author in English.

the temperature gradient term, the current density **j** and the energy flux density **q** are given by

$$\mathbf{j} = \sigma(\mathbf{E} - \alpha\nabla T), \quad \mathbf{q} = (\alpha T - \zeta/e)\mathbf{j} - \kappa\nabla T, \quad (1)$$

where $\mathbf{E} = \nabla\zeta/e$ is the electric field, $\zeta = \mu - e\varphi$ is the electrochemical potential, α is the thermopower, $\sigma = Ne\mu_0$ is the conductivity, $\mu_0 = e\tau/m$ is the mobility, τ is the momentum relaxation time, $\kappa = LT\sigma$ is the thermal conductivity, and $L = \pi^2k^2/3e^2$ is the Lorentz number.

In general, one can uniquely solve Eq. (1) and then find the difference of contact temperatures, $\Delta T = T_a - T_b$, for an arbitrary circuit cooling. But below approximately 1 K, the electron–phonon coupling is known to be weak [12]. In the actual case where $I \longrightarrow 0$, we can then omit the Joule heating. We therefore consider a simple case of adiabatic cooling, with the 2DEG thermally insulated from the environment. We emphasize that under the above conditions, the sample is not heated. Indeed, at small currents, we have $T_a \approx T_b \approx T_0$. Hence, the amount of the Peltier heat $Q_a = I\Delta\alpha T_0$ evolved at contact *a* and that absorbed at contact *b* are equal. Here, $\Delta\alpha$ is the difference of the 2DEG and metal conductor thermopowers. We recall that the energy flux is continuous at each contact,

$$-\kappa\nabla T|_{a,b} = \mathbf{j}\Delta\alpha T_{a,b},$$

and therefore the temperature gradient is constant downstream the current. The difference of the contact temperatures is then given by [9, 10]

$$\Delta T = \frac{\Delta\alpha l_0}{L\sigma w}I,$$

where l_0 and w are respectively the sample length and width. For example, for a 2×2-mm sample, a typical current of $I = 1$ nA, a 2D resistivity on the order of h/e^2, and $\alpha \sim k^2T/e\mu \sim 0.01k/e$, the contact-temperature difference is $\Delta T = 10$ mK $\ll T_0$, and therefore our approach is well justified. From Eq. (1), the voltage drop between ends *c* and *d* is given by

$$U = RI + \Delta\alpha\Delta T,$$

where R is the ohmic resistance of the circuit. The second term is the conventional Seebeck thermoelectromotive force. Because $\Delta T \propto I$, we finally obtain the total 2DEG resistivity as

$$\rho^{\text{tot}} = \rho(1 + \alpha^2/L), \quad (2)$$

where we assume that $\Delta\alpha \approx -\alpha$. We note that within the adiabatic approach, Eq. (2) can also be applied for the 2D hole gas and in the case of four-point contact measurements. In the Appendix, we discuss in more detail the case of 2DEG realistic cooling.

3. RESULTS

3.1. 2D Density and Thermopower

Using Gibbs statistics, we find that the 2DEG density $N = -(\partial\Omega/\partial\mu)_T$ is given by

$$N = N_0\xi F_0(1/\xi), \quad (3)$$

where

$$\Omega = -kT\sum_{\mathbf{k}}\ln\left[1 + \exp\left(\frac{\mu - \varepsilon(\mathbf{k})}{kT}\right)\right]$$

is the thermodynamic potential, $\xi = kT/\mu = T/T_F$ is the dimensionless temperature, T_F is the Fermi temperature, and $F_n(z)$ is the Fermi integral. At the moment, we disregard the valley splitting, reported to be of the order of 1 K in the (100) Si-MOSFET 2DEG system [13]. Below we discuss the importance of a nonzero valley splitting in the context of low-field Shubnikov–de Haas oscillations. Next, we introduce the density of states $D = 2m/\pi\hbar^2$, where m is the effective mass. For the density of strongly degenerate 2DEG, we have $N_0 = D\mu$. In what follows, we consider both the classical Boltzman ($\mu < 0$) and Fermi ($\mu > 0$) cases, and therefore use the dimensionless concentration $n = N/|N_0|$ (Fig. 2a). In the classical Boltzman limit ($\mu < 0$, $|\xi| \ll 1$), the 2D electron density is thermally activated:

$$n = |\xi|\exp(-1/|\xi|).$$

For strongly degenerate electrons ($\xi \ll 1$), we obtain

$$n = 1 + \xi\exp(-1/\xi).$$

Then, at elevated temperatures ($\xi \geq 1$), the density

$$n = 1/2 + \xi\ln 2$$

becomes linear in the temperature. We note that at a fixed temperature, the 2DEG density always exceeds the zero-temperature value, i.e., $N > N_0$ (see Fig. 3, inset). Experimentally, the concentration extracted from the period of the Shubnikov–de Haas quantum oscillations [14, 15] determines the density of strongly degenerate 2DEG, i.e., $N_{\text{SdH}} = N_0$. In contrast, the classical low-field Hall measurements [4, 14] make it possible the total carrier density $N_{\text{Hall}} = N$, which coincides with the density of strongly degenerate electrons for $\xi \ll 1$. We argue that in dilute 2D systems, the accuracy provided by both methods becomes questionable, which seems to be the reason for the sample and temperature-dependent deviation $N_{\text{Hall}} - N_{\text{SdH}}$ observed in Si-MOSFETs [14].

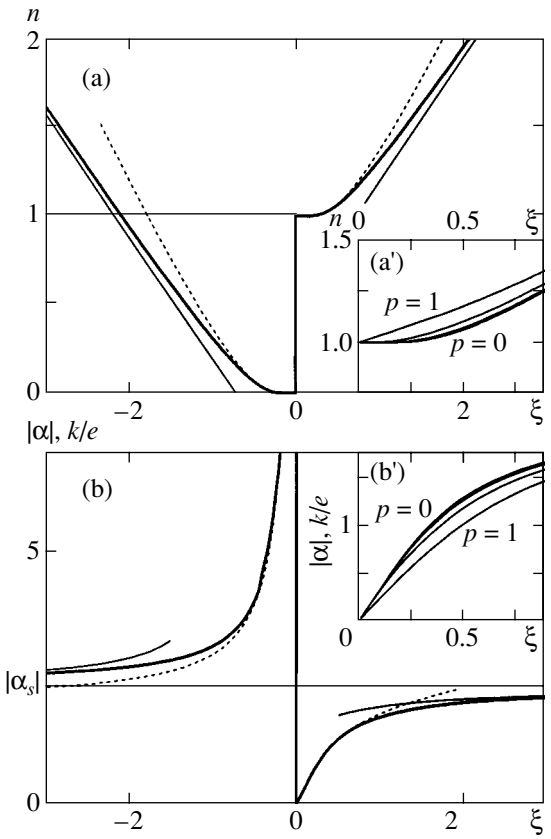

Fig. 2. The zero-field 2DEG density (a) and thermopower (b) given by Eqs. (3) and (4) respectively vs. the dimensionless temperature ξ. Asymptotes shown by dotted fines correspond to $|\xi| \ll 1$ and those shown by thin lines, to $|\xi| \gg 1$. Insets: 2DEG density (a') and thermopower (b') for the spin polarization degree $p = 0$, 0.3, 0.6, 1.

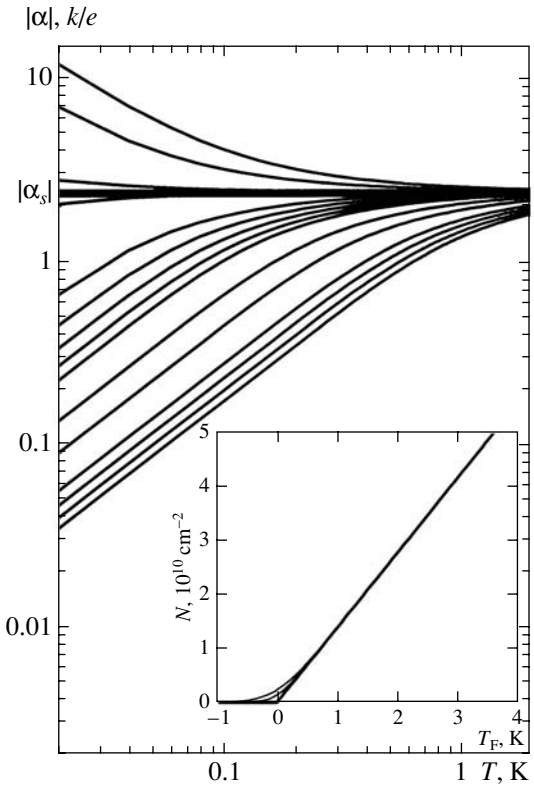

Fig. 3. Temperature dependence of the thermopower given by Eq. (4) for T_F [K] = 2 – 0.25 (step 0.25), 0.2 – 0.05 (the step 0.05), 0.01, 0 (bold line), –0.1, –0.2. Inset: density vs. Fermi energy at the fixed temperature T [K] = 0 (piecewise bold line), 0.15, 0.25 in Si-MOSFET system.

Following the conventional Boltzman equation formalism, the explicit formula for the 2DEG thermopower can be written as

$$\alpha = -\frac{k}{e}\left[\frac{2F_1(1/\xi)}{F_0(1/\xi)} - \frac{1}{\xi}\right]. \quad (4)$$

For simplicity, we assume that the electron scattering is characterized by the energy-independent momentum relaxation time. In the classical limit ($\mu < 0$, $|\xi| \ll 1$), the thermopower is given by the conventional formula

$$\alpha = -\frac{k}{e}\left(2 - \frac{1}{\xi}\right).$$

For strongly degenerate 2DEG ($\xi \ll 1$), we obtain the temperature dependence of the thermopower (Fig. 2b) as

$$\alpha = -\frac{k}{e}\left[\frac{\pi^2\xi}{3} - (1 + 3\xi)\exp\left(-\frac{1}{\xi}\right)\right].$$

At elevated temperatures ($\xi > 1$), the thermopower first grows with temperature and then approaches the universal value

$$\alpha_s = -\frac{k}{e}\frac{2F_1(0)}{F_0(0)} = -\frac{k}{e}\frac{\pi^2}{6\ln 2}.$$

The above behavior is confirmed by low-temperature thermopower measurement data [16], found to diverge at a certain value near $0.6k/e$, which is on the order of α_s (see the bold line in Fig. 3).

3.2. Zero-Field Resistivity

We now calculate the total 2DEG resistivity given by Eq. (2). Figure 4 represents the T-dependent resistivity at fixed Fermi temperatures that correspond to a certain 2DEG density range (see Fig. 3, inset). For a fixed disorder strength, we represent the data found at different densities (or T_F) in a single plot (Fig. 4). In real units, increasing the disorder results in the upshift of resistivity curves. The temperature dependence of the resistivity (see, e.g., the curve at $T_F = 0.25$ K in Fig. 4) exhibits the metallic behavior (i.e. $d\rho/dT > 0$) for $T \le T_F$ and then becomes insulating ($d\rho/dT < 0$) at $T \ge T_F$.

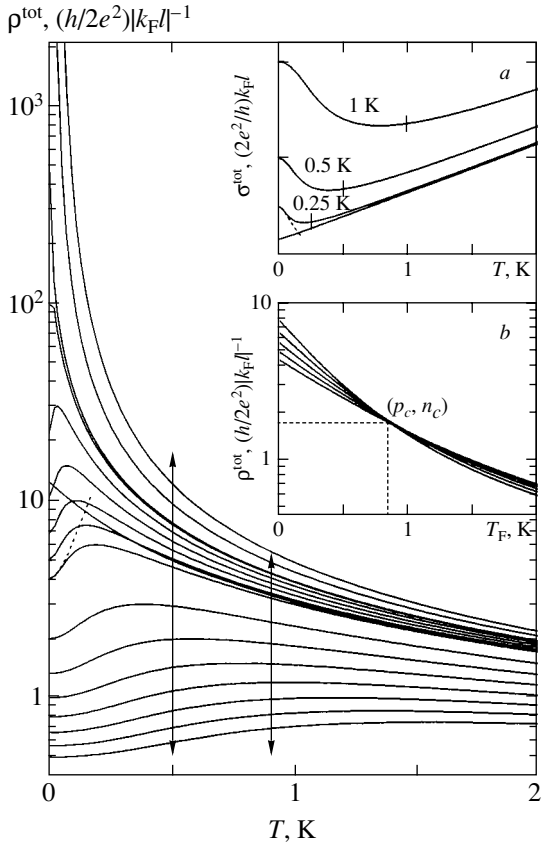

Fig. 4. Zero-field temperature dependence of resistivity given by Eq. (2) for Fermi temperatures depicted in the main panel of Fig. 3. The dotted line corresponds to the asymptote $\xi \ll 1$, the thin line, to $\xi > 1$ for fixed T_F = 0.25 K. The arrows depict the region of temperatures explored in inset b. Inset a: temperature dependence of the inverse resistivity σ^{tot} at T_F [K] = 0.25, 0.5, 1.0 (marked by vertical bars). Inset b: density dependence of the resistivity depicted in the main panel in the temperature range T = 0.5–0.9 K.

Within the low-temperature metallic region, the 2DEG resistivity can be approximated (see the dotted line in Fig. 4) by

$$\rho^{tot} = \rho_0(1 + \pi^2\xi^2/3),$$

where

$$\rho_0 = \frac{1}{N_0 e\mu_0} = \frac{h}{2e^2}\frac{1}{k_F l}$$

is the resistivity at $T \longrightarrow 0$, $k_F = \sqrt{2m\mu}/\hbar$ is the Fermi vector, and $l = \hbar k_F\tau/m$ is the mean free path. For the high-temperature insulating region, we then obtain the asymptote

$$\rho^{tot} = \rho_0\frac{1 + \alpha_s^2/L}{\xi\ln 2} \propto \frac{1}{T}$$

depicted in Fig. 4 by the thin line. The metallic-to-insulating behavior crossover occurs at $T \approx 0.8T_F$. In fact, the low-temperature metallic resistivity is explained within our model in terms of the thermal correction given by Eq. (2), while the high-temperature insulating behavior results from a decrease of the 2D degeneracy. It should be specially noted that in the Boltzman limit (see curves at $\mu < 0$ in Fig. 4), the resistivity can be scaled in units of the disorder parameter k_Fl, where the substitution $\mu \longrightarrow |\mu|$ must be made.

The resistivity data (Fig. 4, inset *b*), represented as a function of the 2D density (or T_F), exhibit a well-pronounced transition point. The critical resistivity ρ_c is roughly inversely proportional to the critical density n_c. We note that the same experimental range $(0.1–10)h/e^2$ expected to eliminate the metal–insulator transition in Si-MOSFETs with different mobilities provides a higher temperature range, critical density n_c, and hence lower ρ_c for more disordered samples. This result is confirmed by experimental observations [17].

Recent experiments [4, 18–20] confirm our predictions and demonstrate that the metallic-region data obey a scaling where the disorder parameter k_Fl (not the ratio of the Coulomb interaction energy to the Fermi energy [20]) and the dimensionless temperature T/T_F appear explicitly. These experimental findings were argued to rule out the electron–electron interactions [4], the shape of the potential well [18], spin–orbit effects, and quantum interference effects [19, 20] as possible origins of the metallic behavior mechanism. In addition, our concept of the high-temperature insulating behavior is qualitatively confirmed by the experimental data [21, 22] within the insulating side of the metal–insulator transition exhibiting the nonhopping $1/T$ dependence. As an example, for *p*-GaAs/AlGaAs 2D hole gas [22] with the peak mobility $\mu_0 = 2 \times 10^5$ cm^2/(V s), we obtain the linear dependence (see the thin asymptote in Fig. 4, inset a) of the inverse resistivity, $\sigma^{tot}[e^2/h] = 1/\rho^{tot} = 1.4T$ [K], which is consistent with the experimental value $3.3T$ [K]. It is to be noted that the conventional theory [23, 24] used to explain the 2D metallic behavior [7, 21, 25] fails to account for both $T \longrightarrow 0$ and $T \geq T_F$ cases.

We emphasize that Eq. (2) provides the actually measured effective 2D mobility and yields

$$\mu_{eff} = \frac{\mu_0}{1 + \alpha^2/L}.$$

Experimentally, at a fixed temperature, the mobility data can be unambiguously extracted using independent measurements of the 2D resistivity and the low-field Hall density $N_{Hall} \approx N$. With the help of Eq. (3), we plot the density dependence of μ_{eff} in Fig. 5. Upon depletion of 2DEG, the dependence $\mu_{eff}(N)$ falls down at low densities near 10^9 cm^{-2} as $\xi \longrightarrow 0$. In the high-

density case, we predict $\mu_{eff} \approx \mu_0$. We argue that in real experiment, the above behavior can be masked by impurity-assisted ($Si–SiO_2$ roughness-associated) suppression of momentum scattering time in the respective cases of low (high) densities [26].

3.3. 2DEG Magnetoresistivity

In contrast to the conventional Shubnikov–de Haas formalism extensively used to reproduce low-field data, we use an alternative approach [27] that appears aimed at resolving the magnetotransport problem within both the Shubnikov–de Haas and integer-quantum-Hall-effect regimes.

The Si-MOSFET energy spectrum modified with respect to valley and spin splitting is given by

$$\varepsilon_n = \hbar\omega_c\left(n_L + \frac{1}{2}\right) \pm \frac{\Delta_s}{2} \pm \frac{\Delta_v}{2}, \quad (5)$$

where $n_L = 0, 1, \ldots$ is the Landau level number, $\omega_c = eB_\perp/mc$ is the cyclotron frequency, $\Delta_s = g^*\mu_B B$ is the Zeeman splitting, g^* is the effective g-factor, and $B = \sqrt{B_\perp^2 + B_\parallel^2}$ is the total magnetic field. Next, Δ_v [K] $= \Delta_v^0 + 0.6B_\perp$ [T] is the density-independent [28] valley splitting. In contrast to the valley splitting, the spin susceptibility $\chi = g^*m/2m_0$ (where m_0 is the free electron mass) is known to exhibit strong enhancement upon 2D carrier depletion. This result is confirmed independently by magnetotransport measurements in a tilted magnetic field [6, 29], the perpendicular field [30], and by the beating pattern of the Shubnikov–de Haas oscillations [31] in crossed fields.

We recall that in strong magnetic fields ($\hbar\omega_c \gg kT, \hbar/\tau$), the electrons can be considered dissipationless, and therefore $\sigma_{xx}, \rho_{xx} \approx 0$. Under current carrying conditions, the only reason for a finite longitudinal resistivity seems to be the thermal correction mechanism discussed earlier [27]. Following [27], we obtain

$$\rho = \rho_{yx}\alpha^2/L, \quad (6)$$

where α is the thermopower, $\rho_{yx}^{-1} = Nec/B_\perp$ is the Hall resistivity, $N = -(\partial\Omega/\partial\mu)_T$ is the 2D density,

$$\Omega = -kT\Gamma\sum_n \ln\left[1 + \exp\left(\frac{\mu - \varepsilon_n}{kT}\right)\right]$$

is the thermodynamic potential modified with respect to the energy spectrum mentioned above, and $\Gamma = eB_\perp/hc$ is the zero width of the Landau-level density of states. In strong magnetic fields, the 2*D* thermopower is a universal quantity [32], proportional to the entropy per electron, $\alpha = -S/eN$, where $S = -(\partial\Omega/\partial T)_\mu$ is the entropy. Both S and N and, hence, α and ρ are universal functions of ξ and the dimensionless magnetic field $\hbar\omega_c/\mu = 4/\nu$, where $\nu = N_0/\Gamma$ is the conventional filling factor.

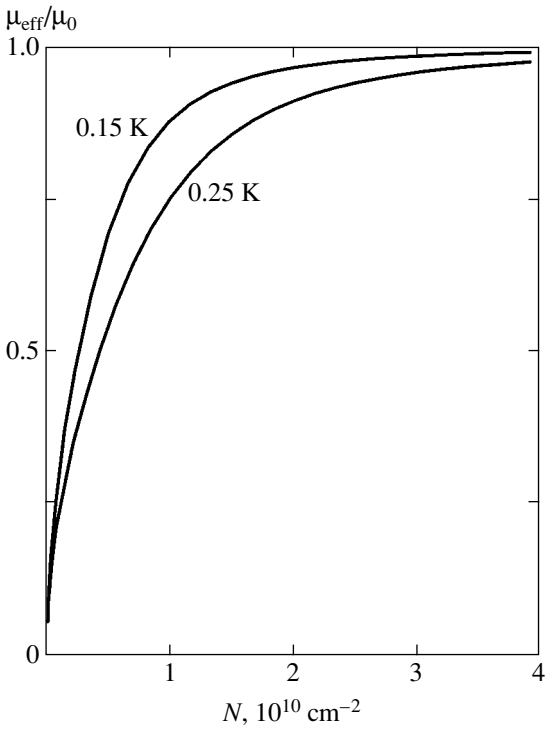

Fig. 5. The dimensionless effective mobility specified in the text vs. the 2D density for the Si-MOSFET system at T = 0.15 and 0.25 K.

Using the Lifshitz–Kosevich formalism, we can easily derive asymptotic formulas for N and S, and hence for ρ_{yx} and ρ, valid at low temperatures $\xi < 1$ and weak magnetic fields $\nu^{-1} < 1$:

$$N = N_0\xi F_0\left(\frac{1}{\xi}\right) + 2\pi\xi N_0\sum_{b=1}^{\infty}\frac{(-1)^b\sin(\pi b\nu/2)}{\sinh r_b}R(\nu),$$
$$S = S_0 - 2\pi^2\xi kN_0\sum_{b=1}^{\infty}(-1)^b\Phi(r_b)\cos\left(\frac{\pi b\nu}{2}\right)R(\nu), \quad (7)$$

where

$$S_0 = kN_0\frac{d}{d\xi}\left[\xi^2 F_1\left(\frac{1}{\xi}\right)\right]$$

is the entropy at $B_\perp = 0$, $F_n(z)$ is the Fermi integral,

$$\Phi(z) = \frac{1 - z\coth z}{z\sinh z},$$

and $r_b = \pi^2\xi\nu b/2$ is the dimensionless parameter. Then $R(\nu) = \cos(\pi bs)\cos(\pi b\nu)$ is the form factor, $s = \Delta_s/\hbar\omega_c = \chi_B/B_\perp$ is the dimensionless Zeeman spin split-

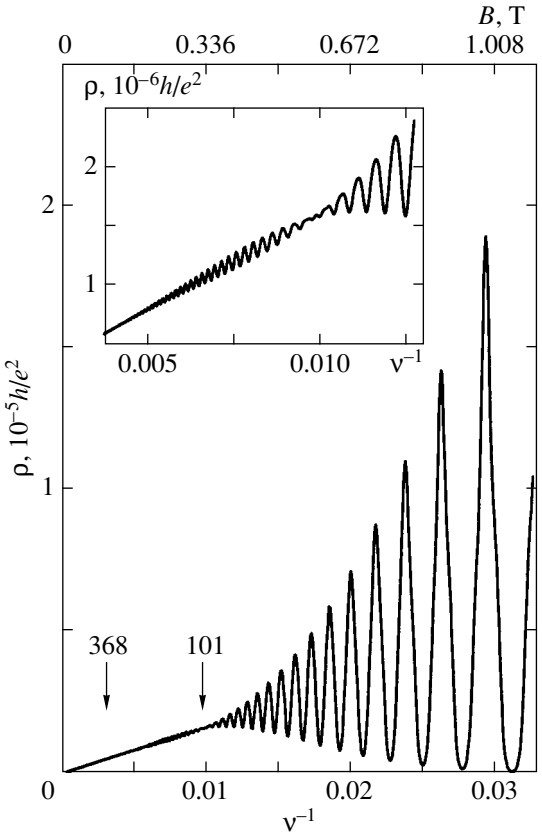

Fig. 6. The Shubnikov–de Haas oscillations at $T = 0.3$ K for a Si-MOSFET sample [13]: $N_0 = 8.39 \times 10^{11}$ cm^{-2}, the spin susceptibility $\chi = 0.305$, and the valley splitting Δ_ν [K] = $\Delta_\nu^0 + 0.6B_\perp$ [T]. The zero-field valley splitting $\Delta_\nu^0 = 0.92$ K is a fitting parameter. The arrows depict the beating nodes at $i = 1, 3$. Inset: an enlarged plot of the beating node from the main panel.

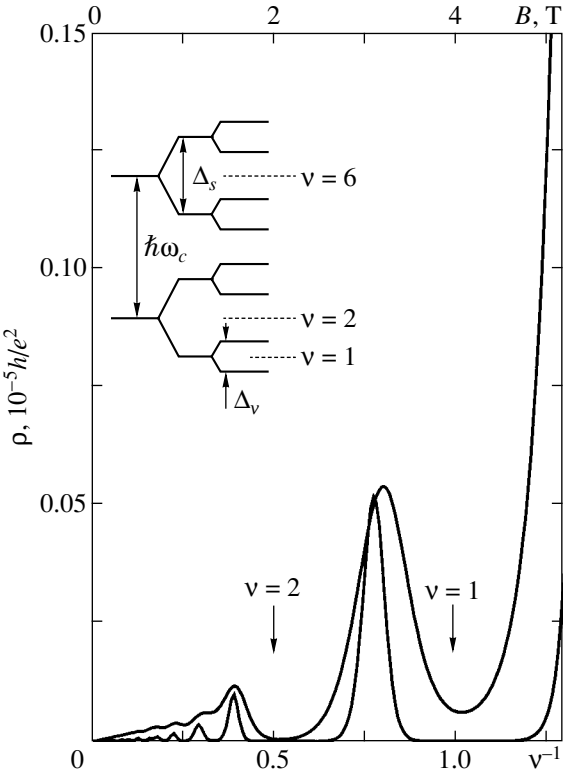

Fig. 7. Magnetoresistivity at $T = 0.36$ K (upper curve) and $T = 0.18$ K for dilute 2DEG Si-MOSFET [33]: $N_0 = 10^{11}$ cm^{-2}, the spin susceptibility $\chi = 0.5$, and the valley splitting specified in the capture Fig. 6. Inset: Energy spectrum given by Eq. (5) for two lowest Landau levels.

ting, and $\nu = \Delta_\nu/\hbar\omega_c = \Delta_\nu^0 \nu/4\mu + 0.12$ is the dimensionless valley splitting.

We first consider the zero-$B_\parallel$ case, where the Zeeman spin splitting is reduced to a field-independent constant, i.e., $s = \chi$. In the low-T, $B_\perp$ limit, the valley splitting Δ_ν^0 is then known to be resolved [13], and therefore leads to the beating of Shubnikov–de Haas oscillations. In the actual first-harmonic case (i.e., $b = 1$), the beating nodes can be observed when $\cos(\pi\nu) = 0$, or

$$\nu_i^\nu = \frac{4\mu(i/2 - 0.12)}{\Delta_\nu^0},$$

where $i = 1, 3, \ldots$ is the beating node index. For 2DEG parameters reported in [13] (Fig. 6), we estimate $\nu_1^\nu = 101$ and therefore $\Delta_\nu^0 = 0.92$ K. The second node is expected to appear at $\nu_3^\nu = 368$. However, the Shubnikov–de Haas oscillations are in fact resolved when $\nu \le 1/\xi = 203$, and therefore the second beating node was not observed in experiment [13]. Moreover, the disappearance of the first beating node upon 2D carrier depletion $N < 3 \times 10^{11}$ cm^{-2} is governed by the same condition because in this case, $\nu \le 1/\xi = 73$ is of the order of the first beating node. We note, however, that suppression of the beating nodes at higher densities ($N > 9 \times 10^{11}$ cm^{-2}) reported in [13] is unexpected within our simple scenario.

We now analyze the case of a low-density 2D system in a strong magnetic field with only the lowest Landau levels occupied. For extremely dilute 2DEG ($N \approx 10^{11}$ cm^{-2}), the energy spectrum (see inset to Fig. 7) is known to be strongly affected by enhanced spin susceptibility. In contrast to the high-density case with cyclotron minima occurring at $\nu = 4, 8, 12, \ldots$, only the spin minima ($\nu = 2, 6, 10, \ldots$) are observed in dilute 2DEG [33]. As expected, the spin (cyclotron) minimum fillings are proportional to odd (even) numbers multiplied by a factor of two due to the valley degeneracy. In stronger fields, the magnetoresistivity data exhibit a $\nu = 1$ minimum associated with valley splitting. With the energy spectrum implied by Eq. (5), we can easily find that the last minimum occurs when the Fermi level lies between the lowest valley-split Landau levels, i.e., at

$\mu = \hbar\omega_c(1 - \chi)/2$. The sequence of minima at $B = 4$, 2, 0.66 T reported in [33] provides an independent test for spin susceptibility in the high-$B_\perp$ limit. In Fig. 7, we represent the magnetoresistivity specified by Eq. (6) and then use $\chi = 0.5$ in order to fit the observed sequence of minima. Surprisingly, the value of spin susceptibility is lower than $\chi = 0.86$ extracted from the crossed-field low-field Shubnikov–de Haas beating pattern analysis [31]. We attribute this discrepancy, for example, to the possible magnetic-field dependence of spin susceptibility.

Finally, we focus on the magnetotransport problem in the crossed magnetic field configuration. Following experiments [13], we further neglect the zero-field valley splitting in the actual high-density case ($N > 9 \times 10^{11}$ cm^{-2}). At a fixed parallel magnetic field, the dimensionless Zeeman splitting is given by

$$s = \chi\sqrt{1 + \nu^2/\nu_\parallel^2},$$

where we introduce an auxiliary "filling factor" $\nu_\parallel = hcN_0/eB_\parallel$ associated with the parallel field. In the low-$B_\perp$ limit, the spin splitting induced by the parallel field also results in the beating of the Shubnikov–de Haas oscillations. We can easily derive the condition for the Shubnikov–de Haas beating nodes as $\cos(\pi s) = 0$ or

$$\nu_j^s = \nu_\parallel\sqrt{(j/2\chi)^2 - 1}, \quad j = 1, 3, \ldots.$$

The sequence of the beating nodes observed in [31] allowed the authors to deduce the density dependence of the spin susceptibility. As an example, in Fig. 8, we reproduce the magnetoresistivity implied by Eqs. (6) and (7) for 2DEG parameters [31]. The phase of the Shubnikov–de Haas oscillations remains the same between the adjacent beating nodes, and changes by π through the node are consistent with experiments.

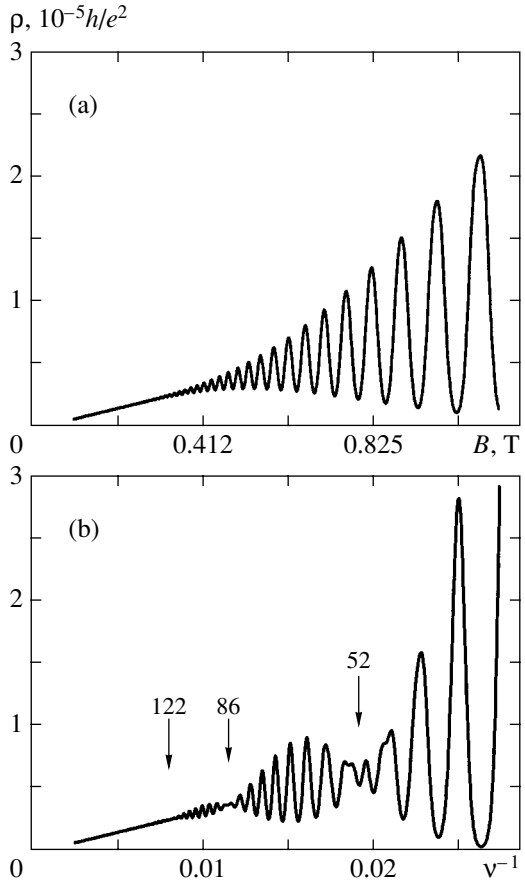

Fig. 8. The Shubnikov–de Haas beating pattern oscillations at $T = 0.35$ K for a Si-MOSFET sample [31]: $N_0 = 10.6 \times 10^{11}$ cm^{-2}, the spin susceptibility $\chi = 0.27$, $\Delta_\nu^0 = 0$ and $B_\parallel = 0$ (a), $B_\parallel = 4.5$ T (b), $\nu_\parallel = 9.25$. The arrows depict the beating nodes at $j = 3$, 5, 7.

We now consider the 2DEG magnetotransport in a tilted configuration case with the sample rotated in a constant magnetic field [6, 7, 29]. In this case, the Shubnikov–de Haas beating pattern is known to depend on the spin polarization degree $p = \Delta_s/2\mu = 2\chi/\nu_{tot}$, where we introduce the auxiliary "filling factor" $\nu_{tot} = hcN_0/eB$ associated with the total magnetic field. Conventionally, the degree of spin polarization is related to the parallel field B_c required for complete spin polarization, and therefore, $p = B/B_c$. Performing a minor modification in Eq. (6), namely, that $s = \chi\nu/\nu_{tot}$, in Fig. 9 we represent the magnetoresistivity as a function of the filling factor for a 2DEG plane rotated with respect to the constant magnetic field $B = 18$ T [7]. For simplicity, we omit zero-field valley splitting. Then, assuming that the broadening of Landau levels is neglected within our simple approach, we use a somewhat higher temperature compared to that in experiment [7]. For a spin-polarized system, the Shubnikov–de Haas oscillations ($p = 1.01$ in Fig. 9a) are caused by the only lowest valley-degenerate spin-up subband. At low temperatures, the valley splitting at $\nu = 3$ is resolved. With the energy spectrum specified by Eq. (5), the high-filling maxima occur at $4(N + 1/2)/(1 + p) \sim 2N + 1$ and therefore have a period of $\Delta\nu = 2$. In contrast, the partially polarized high-density 2DEG case ($p = 0.29$) depicted in Fig. 9b demonstrates a rather complicated beating pattern caused by both spin-up and spin-down subbands. It can be easily demonstrated that high-filling maxima occur at $4(N + 1/2)/(1 \pm p)$ (dots in Fig. 9b) and hence depend on the spin polarization degree. The ratio of oscillation frequencies for the two spin subbands is

$$\frac{f_\downarrow}{f_\uparrow} = \frac{1 - p}{1 + p},$$

consistently with experiment [7]. At the moment, however, we cannot explain the puzzling behavior of low-filling magnetoresistivity data known to be insensitive to the parallel field component [6, 33].

We emphasize that the data represented in Figs. 6–9 differ from those provided by the conventional formalism in the following aspects: (i) the low-field ($\omega_c\tau \le 1$)

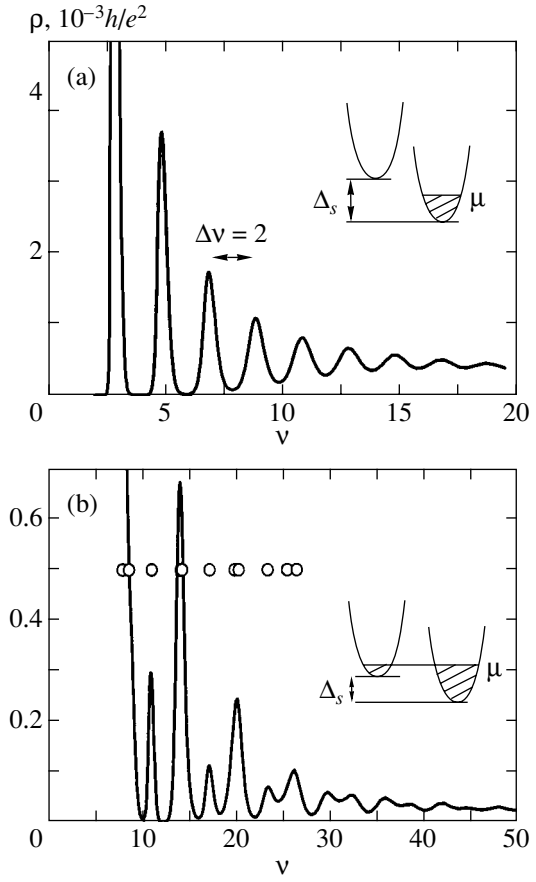

Fig. 9. The small-angle Shubnikov–de Haas oscillations at $T = 1.35$ K for a Si-MOSFET system [7]: (a) spin polarized electrons ($p = 1.01$) at $N_0 = 3.72 \times 10^{11}$ cm^{-2}, the spin susceptibility $\chi = 0.42$ [31], the "effective filling factor" $\nu_{tot} = 0.83$, and (b) the partially polarized case ($p = 0.29$) at $N_0 = 9.28 \times 10^{11}$ cm^{-2}, the spin susceptibility $\chi = 0.30$ [31], and $\nu_{tot} = 2.06$. The positions of maxima are represented by empty dots. Insets: schematic band diagrams at $B = B_{||}$.

quantum interference and classical negative magnetoresistivity background is excluded within our approach and (ii) in contrast to the conventional Shubnikov–de Haas analysis, our approach determines (at $\omega_c\tau \gg 1$) the absolute value of magnetoresistivity and, moreover, provides a continuous transition from the Shubnikov–de Haas regime to the quantum Hall effect ($\hbar\omega_c \gg kT$). A minor point is that our approach predicts a somewhat lower Shubnikov–de Haas oscillation amplitude compared to that in experiment. However, in the integer-quantum-Hall-effect regime, the magnetoresistivity magnitude is very comparable to experimental values [27].

3.4. Parallel-Field Magnetoresistivity

One of the most intriguing features of the Si-MOSFET 2D system is its enormous response to the magnetic field applied in the plane of the electrons. At a fixed temperature, the parallel-field resistivity is known to exhibit a dramatic increase at both sides of the zero-field metal–insulator transition. On the metallic side, the resistivity increases by more than an order of magnitude and then saturates above a certain value of the parallel magnetic field. The saturation field corresponds to the complete spin polarization [6], when $p = 1$. On the insulating metal–insulator transition side, the saturation of the magnetoresistivity is not observed [6]. We now give a qualitative argument in favor of the magnetic-field-driven disorder origin of the observed magnetoresistivity data.

At a fixed parallel magnetic field, the behavior of the T-dependent resistivity is reported [17] to be similar to that in the zero-field case (see Fig. 4). Moreover, the same data plotted as a function of density also exhibit a well-pronounced transition point as in the case of the zero-field metal–insulator transition (see Fig. 4, inset b). Both the critical resistivity ρ_c^B and the density n_c^B depend on the magnetic field strength. Surprisingly, the critical diagram ρ_c^B vs. n_c^B was found [17] to coincide with that obtained in the case of the zero-field metal–insulator transition for different mobility Si-MOSFET samples. Assuming that the thermal correction mechanism is also valid in the presence of the parallel field, we attribute the observed magnetoresistivity behavior to the field-driven disorder enhancement, i.e., $\tau(p) < \tau(0)$. Indeed, with the energy spectrum specified by Eq. (5), the explicit formulas for the 2DEG density and thermopower are

$$N = \frac{N_0\xi}{2}\sum_i F_0\left(\frac{1-\varepsilon_i}{\xi}\right),$$
$$\alpha = -\frac{k}{e}\left[\frac{\sum_i\left[2F_1\left(\frac{1-\varepsilon_i}{\xi}\right) + \frac{\varepsilon_i}{\xi}F_0\left(\frac{1-\varepsilon_i}{\xi}\right)\right]}{\sum_i F_0\left(\frac{1-\varepsilon_i}{\xi}\right)} - \frac{1}{\xi}\right], \quad (8)$$

where $\varepsilon_i = \pm p$ is the dimensionless energy deficit between the bottom of spin subbands and that of the ground state. For simplicity, we here neglect the zero-field valley splitting. Both the 2D density and thermopower exhibit (see Fig. 2, insets a' and b') only a minor perturbation upon parallel field enhancement within $0 < p < 1$. We therefore conclude that the field-driven disorder enhancement can be responsible for the observed magnetoresistivity behavior. A detailed analysis of the prevailing $\tau(p)$ mechanism (see, e.g., [34]) is beyond the scope of the present paper.

3.5. 2D Compressibility

Herein, we refer the reader to experimental data for the most part obtained for the n-GaAs/AlGaAs 2DEG system, and therefore we should substitute $D \longrightarrow D/2$

in what follows. In general, of particular interest is the 2DEG compressibility

$$K = \frac{dN}{d\mu} = -\frac{d^2\Omega}{d\mu^2},$$

known to be a fundamental quantity generally more amenable to theoretical and experimental analysis [15, 35]. For noninteracting electrons, Eq. (3) yields

$$K(\xi) = DF_0'(1/\xi),$$

where $F_n'(z) = dF_n(z)/dz$ is the derivative of the Fermi integral. Figure 10 represents the dependence of the actually measured inverse compressibility $d(\mu) = \varepsilon/Ke^2$. For strongly degenerate electrons ($\xi \ll 1$), we obtain a constant value $d_0 = \varepsilon/De^2$, consistent with the conventional capacitance measurements [36]. But as the 2DEG degeneracy decreases, the penetration length of the AC electric field [15, 35] diminishes and, furthermore, the negative inverse compressibility also decreases compared to d_0.

Conventionally, this behavior is explained [15] in terms of a Hartree–Fock exchange, which is omitted in our simple approach. In contrast, for an extremely depleted 2DEG, the inverse compressibility data always exhibit an abrupt upturn, which cannot be explained within the Hartree–Fock scenario [35]. We assume that the above feature has a natural explanation within our model (see the dotted line in Fig. 10), because $d = d_0\exp(-1/|\xi|)$ at $\mu < 0$, $|\xi| \ll 1$ and hence the inverse compressibility exhibits the *T*-activated behavior. For example, upon depletion, the inverse compressibility [15] strongly increases at $T_F = 0.63$ K ($N = 2 \times 10^9$ cm^{-2}), being of the order of the bath temperature $T = 0.3$ K.

In the general case of 2DEG placed in the perpendicular magnetic field, the compressibility is

$$K = -\left(\frac{\partial\Omega}{\partial\mu}\right)_T = \frac{D}{\xi\nu}\sum_n \cosh^{-2}\left(\frac{\varepsilon_n - \mu}{2kT}\right),$$

or

$$K \approx D\left[F_0'\left(\frac{1}{\xi}\right) + \pi^2\xi\nu\sum_b \frac{(-1)^b b\cos(\pi b\nu/2)}{\sinh r_b}\right], \quad (9)$$

where we use the thermodynamic potential modified with respect to the single-valley spin-unresolved Landau level energy spectrum. According to Eq. (9), at a fixed magnetic field and temperature, the dependence $d(\mu)$ can be viewed (see Fig. 10) as a superposition of the zero-field dependence and the Landau-level-related oscillations. For a typical GaAs/AlGaAs system [15], we represent the data for $B = 0.5$ T ($\hbar\omega_c = 1$ K > *T*) in

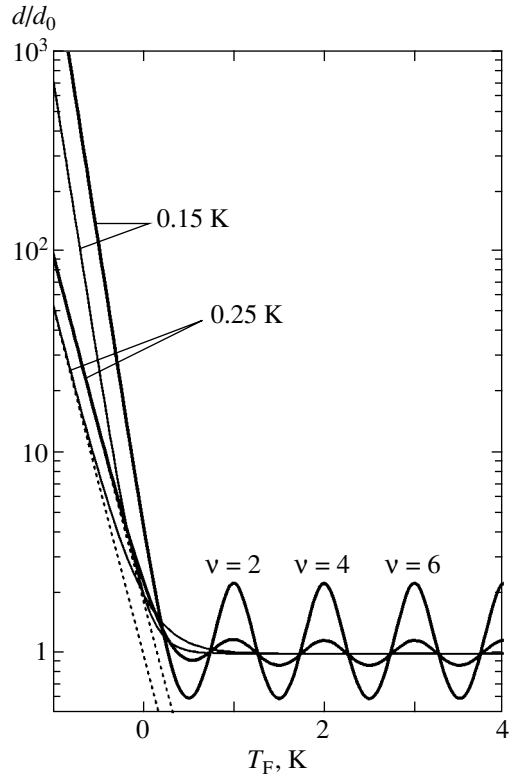

Fig. 10. The dimensionless inverse compressibility vs. the Fermi temperature at zero magnetic field (thin lines) and $\hbar\omega_c = 1$ K (bold lines) at fixed temperatures $T = 0.15$ K and $T = 0.25$ K. Dotted lines depict asymptotes at $\xi < 0$, $|\xi| \ll 1$.

Fig. 10. The Landau-level-assisted oscillations at $\nu = 2$, 4, 6, … are well resolved. We note that for a dilute 2DEG system at $\xi < 0$, $|\xi| \ll 1$ in the presence of a strong magnetic field, $\hbar\omega_c \gg kT$ (i.e., $\xi\nu \ll 1$), we obtain the *T*-activated behavior as

$$d = \frac{d_0\xi\nu}{2}\exp\left(\frac{1}{\xi\nu} - \frac{1}{|\xi|}\right),$$

similarly to the zero-field case (see dotted lines in Fig. 10).

4. CONCLUSIONS

In conclusion, the total resistivity of a dilute 2D system in Si-MOSFET with the thermal correction included is found to be a universal function of the temperature, expressed in units $(h/e^2)(k_F l)^{-1}$. We have demonstrated the relevance of the approach suggested in [27] to the low-field beating pattern of the Shubnikov–de Haas oscillations in both crossed and tilted magnetic field configurations. The features concerning the integer quantum Hall effect in dilute Si-MOSFET systems are discussed. The strong increase of the parallel magnetoresistivity was argued to result from the spin-dependent disorder.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 03-02-17588) and the LSF (HPRI-CT-2001-00114, Weizmann Institute).

APPENDIX

Real Cooling of the 2D System

We consider the more realistic situation of electron cooling caused by a finite strength of the electron–phonon coupling. Cooling of the phonon-to-mixing-chamber could then predominately occur over the sample surface. The power balance equations linearized with respect to small temperature perturbations are

$$\begin{aligned} \operatorname{div}(\kappa\nabla T) + \frac{j^2}{\sigma} - jT\nabla\alpha - \beta(T - T_p) &= 0, \\ \operatorname{div}(\kappa_p\nabla T_p) - \gamma(T_p - T) - \beta(T_p - T_0) &= 0, \end{aligned} \tag{10}$$

where κ_p is the phonon thermal conductivity, T_p is the local phonon temperature, and β and γ are the respective electron–phonon and sample-to-mixing-chamber cooling strengths. With the phonon diffusion assumed weak in the sample bulk, the phonon temperature

$$T_p = \frac{\gamma T + \beta T_0}{\gamma + \beta}$$

coincides with the electron (bath) temperature upon predominant cooling. In general, $T_0 < T_p < T$. The electron–phonon coupling term in Eq. (10), rewritten in terms of the bath temperature, is

$$\beta^*(T - T_0) = \frac{\gamma\beta}{\gamma + \beta}(T - T_0),$$

which depends on both coupling constants. As expected, a weak heat path channel provides thermal cooling of the 2DEG system.

With $T_0^* = T_0 + j^2/\sigma\beta^*$ being the Joule heat enhanced temperature, Eq. (10) yields

$$\Delta\Theta^2 - u\nabla\Theta^2 - 2(\Theta - 1) = 0, \tag{11}$$

where $\theta = T/T_0^*$ is the dimensionless electron temperature, $\eta = x/\lambda$ is the dimensionless coordinate, $\lambda = (L\sigma T_0^*/\beta^*)^{1/2}$ is the thermal diffusion length scale, and $u(\Theta) = (\lambda_j/L\sigma)d\alpha/dT$ is the dimensionless parameter. Because $T \approx T_0$, the 2D thermopower can be considered constant, and we therefore omit the second term in Eq. (11). Then the energy flux continuity at both ends of the sample provides symmetric boundary conditions with the temperature gradients $\nabla\Theta|_{a,b} = -j\lambda\Delta\alpha/L\sigma T_0^*$. Under these conditions, solving Eq. (11) is straightforward [10]. The temperature profile downstream the sample is governed by the sample-to-thermal diffusion length ratio l_0/λ. Our approach of the adiabatic cooling is justified when $l_0 \ll \lambda$. In the opposite case of strong cooling ($l_0 \gg \lambda$), the electron temperature exhibits sharp deviation with respect to T_0^* near the contacts and then coincides with T_0^* in the sample bulk.

Considering that the use of interior potential probes gives uniform resistivity data, Prus *et al.* [37] suggested these data as a precursor of strong cooling in real Si-MOSFETs. With the *T*-dependent resistivity of 2DEG used as a thermometer, the electron–phonon coupling constant was extracted [37] from the simplified energy balance condition $T = T_0^*$ valid in the sample bulk when $l_0 \gg \lambda$. We stress that the above procedure is only justified when the electron–phonon coupling, and hence the thermal diffusion length are known a priori. Indeed, the weak coupling, if present, provides a constant temperature gradient and hence a uniform resistivity as well. Nevertheless, the simple balance condition used in [37] becomes useless because $T \neq T_0^*$ throughout the sample. It turns out that the weak electron–phonon coupling constant cannot be extracted in the conventional manner. One can estimate the critical electron–phonon coupling at which our adiabatic approach is valid, i.e., $l_0 < \lambda$ or $\beta^* < L\sigma T_0^*/l_0^2$. For $l_0 = 3$ mm, $\sigma \sim e^2/h$, and $T_0 = 100$ mK, we obtain $\beta^* < 1.1 \times 10^{-10}$ W/(K cm^2).

It should be noted that the Peltier effect correction to resistivity becomes strongly damped at higher frequencies because of the thermal inertial effects [9]. Our dc approach is valid below some critical frequency $f_{cr} \approx \hbar/ml_0^2 = 0.3$ kHz, and therefore the spectral dependence of the 2D resistivity can be used to estimate the thermal correction.

REFERENCES

1. S. V. Kravchenko, G. V. Kravchenko, J. E. Furneaux, *et al.*, Phys. Rev. B **50**, 8039 (1994).
2. D. Popovic, A. B. Fowler, and S. Washburn, Phys. Rev. Lett. **79**, 1543 (1997).
3. P. T. Coleridge, R. L. Williams, Y. Feng, and P. Zawadzki, Phys. Rev. B **56**, R12764 (1997).
4. M. Y. Simmons, A. R. Hamilton, M. Pepper, *et al.*, Phys. Rev. Lett. **80**, 1292 (1998); Physica E (Amsterdam) **11**, 161 (2001).
5. Y. Hanein, U. Meirav, D. Shahar, *et al.*, Phys. Rev. Lett. **80**, 1288 (1998).
6. T. Okamoto, K. Hosoya, S. Kawaji, and A. Yagi, Phys. Rev. Lett. **82**, 3875 (1999).
7. S. A. Vitkalov, H. Zheng, K. M. Mertes, *et al.*, Phys. Rev. Lett. **85**, 2164 (2000); cond-mat/0101196.
8. D. Simonian, S. V. Kravchenko, M. P. Sarachik, and V. M. Pudalov, Phys. Rev. Lett. **79**, 2304 (1997).

9. C. G. M. Kirby and M. J. Laubitz, Metrologia **9**, 103 (1973).
10. M. V. Cheremisin, Zh. Éksp. Teor. Fiz. **119**, 409 (2001) [JETP **92**, 357 (2001)].
11. M. V. Cheremisin, in *Proceedings on ICPS'26* (Edinburgh, UK, 2002), D113.
12. A. Mittal, M. W. Keller, R. G. Wheeler, and D. E. Prober, Physica B (Amsterdam) **194–196**, 167 (1994).
13. V. M. Pudalov, A. Punnoose, G. Brunthaler, *et al.*, cond-mat/0104347.
14. V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 48 (1999) [JETP Lett. **70**, 48 (1999)].
15. J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **68**, 674 (1992); Phys. Rev. B **50**, 1760 (1994).
16. R. Fletcher, V. M. Pudalov, A. D. B. Radcliffe, and C. Possanzini, Semicond. Sci. Technol. **16**, 386 (2001).
17. V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, cond-mat/0103087.
18. A. R. Hamilton, M. Y. Simmons, M. Pepper, *et al.*, Phys. Rev. Lett. **87**, 126802 (2001).
19. G. Brunthaler, A. Prinz, G. Bauer, and V. M. Pudalov, Phys. Rev. Lett. **87**, 096802 (2001).
20. A. Lewalle, M. Pepper, C. J. B. Ford, and E. H. Hwang, Phys. Rev. B **66**, 075324 (2002).
21. H. Noh, M. P. Lilly, D. C. Tsui, and J. A. Simmons, Phys. Rev. B **68**, 165308 (2003); cond-mat/0301301.
22. Xuan P. A. Gao, A. P. Mills, Jr., A. P. Ramirez, and L. N. Pfeiffer, Phys. Rev. Lett. **88**, 166 803 (2002); cond-mat/0308003.
23. A. Gold and V. T. Dolgopolov, Phys. Rev. B **33**, 1076 (1986).
24. G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **65**, 020201 (2002).
25. V. M. Pudalov, M. E. Gershenson, H. Kojima, and G. Brunthaler, Phys. Rev. Lett. **91**, 126403 (2003).
26. A. K. Savchenko, private communication.
27. M. V. Cheremisin, in *Proceedings on NATO ASI* (Windsor, UK, 2001); cond-mat/0102153.
28. V. M. Pudalov, S. G. Semenchinskii, and V. S. Edelman, Zh. Éksp. Teor. Fiz. **89**, 1870 (1985) [Sov. Phys. JETP **62**, 1079 (1985)].
29. F. F. Fang and P. J. Stiles, Phys. Rev. **174**, 823 (1968).
30. S. V. Kravchenko, A. A. Shashkin, D. A. Bloore, and T. M. Klapwijk, Solid State Commun. **116**, 495 (2000).
31. V. M. Pudalov, M. E. Gershenson, H. Kojima, and N. Butch, Phys. Rev. Lett. **88**, 196404 (2002).
32. S. M. Girvin and M. Jonson, J. Phys. C **15**, L1147 (1982).
33. S. V. Kravchenko, D. Simonian, M. P. Sarachik, and A. D. Kent, Phys. Rev. B **58**, 3553 (1998).
34. V. T. Dolgopolov and A. Gold, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 42 (2000) [JETP Lett. **71**, 27 (2000)].
35. S. C. Dultz and H. W. Jiang, Phys. Rev. Lett. **84**, 4689 (2000).
36. S. I. Dorozhkin, J. H. Smet, and K. von Klitzing, Phys. Rev. B **63**, 121301 (2001).
37. O. Prus, M. Reznikov, U. Sivan, and V. M. Pudalov, Phys. Rev. Lett. **88**, 016801 (2002).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 608–616.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 686–695.

ELECTRONIC PROPERTIES OF SOLIDS

The Electron Energy Spectrum and Superconducting Transition Temperature of Strongly Correlated Fermions with Three-Center Interactions

V. V. Val'kov[a,b,c,*] and D. M. Dzebisashvili[a,c]

[a] *Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*
[b] *Krasnoyarsk State Technical University, Krasnoyarsk, 660074 Russia*
[c] *Krasnoyarsk State University, Krasnoyarsk, 660075 Russia*
**e-mail: vvv@iph.krasn.ru*

Received August 2, 2002; in final form, July 14, 2004

Abstract—The renormalizations of the fermionic spectrum are considered within the framework of the t–J^* model taking into account three-center interactions ($H_{(3)}$) and magnetic fluctuations. Self-consistent spin dynamics equations for strongly correlated fermions with three-center interactions were obtained to calculate quasi-spin correlators. A numerical self-consistent solution to a system of ten equations was obtained to show that, in the nearest-neighbor approximation, simultaneously including $H_{(3)}$ and magnetic fluctuations at $n > n_1$ ($n_1 \approx 0.72$ for $2t/U = 0.25$) caused qualitative changes in the structure of the energy spectrum. A new Van Hove singularity is then induced in the density of states, and an additional maximum appears in the $T_c(n)$ concentration dependence of the temperature of the transition to the superconducting phase with order parameter symmetry of the $d_{x^2-y^2}$ type. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is commonly assumed [1–5] that strong electron correlation plays an important role in the mechanism of high-temperature superconductivity. One of the key models that includes strong electron correlation is the Hubbard model [6], which, in the simplest case, contains two energy parameters: the electron hopping integral between the nearest sites t and the Coulomb repulsion energy U between two electrons at one site with opposite spin moment projections (double occupancy). The inequality $U \gg |t|$ holds under strong electron correlation conditions.

At low electron concentrations, $n \ll 1$, this system is well described by Fermi liquid theory [7]. In the other limiting case of $n \longrightarrow 1$, we have the scenario of a Heisenberg antiferromagnet. At intermediate concentrations, $n < 1$, the problem of the ground state can only be described approximately.

At $U \gg t$ and $n < 1$, the Hubbard model is often considered in a truncated Hilbert space without double occupancy. The corresponding effective Hamiltonian contains not only terms that describe antiferromagnetic correlations between the spin moments of charge carriers but also three-center terms [8, 9]. Three-center interactions insignificantly influence the dispersion dependence of the spectrum of fermionic excitations [10, 11]. At the same time, their contribution becomes substantial in the superconducting phase. For the first time, this was shown in [12, 13]. In particular, the renormalization of the coupling constant when three-center terms are included [13] decreases the superconducting transition temperature by more than an order of magnitude [14].

The role played by magnetic fluctuations in superconducting pairing was studied in many works [3, 4, 15–17]. In recent years, magnetopolaron bound states have been introduced into the theory of the electronic structure of the CuO_2 plane [18].

We show in this work that magnetic fluctuations are capable of qualitatively renormalizing the electron energy spectrum and the density of states provided three-center interactions are taken into account. Such modifications substantially influence the concentration dependence of the superconducting transition temperature T_c. In this context, recent work [19] is noteworthy. In [19], the ideology of a substantial influence of three-center interactions and magnetic correlations was applied to describe the properties of n-type cuprates.

The present paper is organized as follows. The meaning of effective interactions in the t–J^* model is considered in Section 2. Section 3 contains self-consistency equations. Spin correlator calculations in the t–J^* model are described in Section 4. In what follows, the combined influence of three-center interactions and magnetic fluctuations is considered.

2. THE DOUBLE OCCUPANCY OPERATOR AND EFFECTIVE INTERACTIONS

The passage to the effective Hamiltonian in the atomic representation can be performed at $U \gg t$ and $n < 1$ either by canonical transformation [4] or using the operator form of perturbation theory [14, 20]. With an accuracy to terms proportional to t^2/U, we have

$$H_{\text{eff}} \equiv H_{t-J^*} = H_{t-J} + H_{(3)}, \tag{1}$$

where

$$H_{t-J} = \sum_{f\sigma}(\epsilon - \mu)X_f^{\sigma\sigma} + \sum_{fm\sigma} t_{fm} X_f^{\sigma 0} X_m^{0\sigma} + \frac{1}{2}\sum_{fm\sigma} J_{fm}(X_f^{\sigma\bar{\sigma}} X_m^{\bar{\sigma}\sigma} - X_f^{\sigma\sigma} X_m^{\bar{\sigma}\bar{\sigma}}) \tag{2}$$

is the Hamiltonian of the t–J model [21] ($J_{fm} = 2t_{fm}^2/U$) and the three-center operator

$$H_{(3)} = \sum_{\substack{fmg\sigma \\ f \neq g}} \left(\frac{t_{fm}t_{mg}}{U}\right) \times \{X_f^{\sigma 0} X_m^{\bar{\sigma}\sigma} X_g^{0\bar{\sigma}} - X_f^{\sigma 0} X_m^{\bar{\sigma}\bar{\sigma}} X_g^{0\sigma}\} \tag{3}$$

takes into account the effect of correlated electron hoppings.

Let us elucidate the physical meaning of the terms proportional to t^2/U in the effective Hamiltonian. For this purpose, consider the operator $\hat{N}_2 = \sum_f X_f^{22}$ of the total number of doubly occupied sites. If $|\Psi_0\rangle$ is the ground state of the system described by the initial Hubbard Hamiltonian, the number of double occupancies is $N_2 = \langle\Psi_0|\hat{N}_2|\Psi_0\rangle$. The canonical transformation

$$H \longrightarrow H_{\text{eff}} = \exp(iS)H\exp(-iS)$$

changes the ground state function by the law

$$|\Psi_0\rangle \longrightarrow |\Psi_0\rangle = \exp(iS)|\Psi_0\rangle.$$

Therefore,

$$N_2 = \langle\Phi_0|\exp(iS)\hat{N}_2\exp(-iS)|\Phi_0\rangle.$$

It follows that the operator of double occupancies in the Hilbert space of the effective Hamiltonian is determined by the equation

$$\tilde{\hat{N}}_2 = \exp(iS)\hat{N}_2\exp(-iS).$$

Clearly, this conclusion remains valid when we pass to finite temperatures. Calculations with an accuracy quadratic in (t/U) give the effective operator of double occupancies in the form

$$\tilde{\hat{N}}_2 = -\sum_{fmg\sigma}\left(\frac{t_{fm}t_{mg}}{U^2}\right) \times \{X_f^{\sigma 0} X_m^{\bar{\sigma}\sigma} X_g^{0\bar{\sigma}} - X_f^{\sigma 0} X_m^{\bar{\sigma}\bar{\sigma}} X_g^{0\sigma}\} = -(H_{\text{exch}} + H_{(3)})/U, \tag{4}$$

where the exchange term H_{exch} is determined by the well-known equation

$$H_{\text{exch}} = \frac{1}{2}\sum_{fm\sigma} J_{fm}\{X_f^{\sigma\bar{\sigma}} X_m^{\bar{\sigma}\sigma} - X_f^{\sigma\sigma} X_m^{\bar{\sigma}\bar{\sigma}}\}. \tag{5}$$

It follows from these results that the Hamiltonian of the t–J^* model can be represented in the form

$$H_{t-J^*} = \sum_{f\sigma}(\epsilon - \mu)X_f^{\sigma\sigma} + \sum_{fm\sigma} t_{fm} X_f^{\sigma 0} X_m^{0\sigma} - U\tilde{\hat{N}}_2. \tag{6}$$

We see that exchange and three-site interactions appear in H_{t-J^*} for the same reason determined by the presence of a finite number of double occupancies in the system.

3. SELF-CONSISTENCY EQUATIONS

The combined effects of magnetic correlations and three-center interactions on the renormalizations of the spectrum of fermionic excitations and the conditions of the existence of $d_{x^2-y^2}$ superconductivity will be studied using the irreducible Green functions constructed on Hubbard operators [22, 23].

When three-center interactions are taken into account, the first exact equation of motion for the anticommutator Green function is written in the form

$$\begin{aligned}(\omega - \varepsilon + \mu)\langle\langle X_f^{0\sigma}|X_g^{\sigma 0}\rangle\rangle_\omega &= \delta_{fg}N_{0\sigma} \\ &+ \sum_{m\sigma_1} t_{fm}\langle\langle(X_f^{\sigma_1\sigma} + \delta_{\sigma_1\sigma}X_f^{00})X_m^{0\sigma_1}|X_g^{\sigma 0}\rangle\rangle_\omega \\ &+ \sum_{m\sigma_1} J_{fm}4\sigma_1\sigma\langle\langle X_f^{0\bar{\sigma}_1}X_m^{\bar{\sigma}\sigma_1}|X_g^{\sigma 0}\rangle\rangle_\omega \\ &+ \sum_{\substack{ml\sigma_1 \\ (m\neq l)}}\left(\frac{t_{fl}t_{mf}}{U}\right)4\sigma_1\sigma\langle\langle X_f^{0\sigma_1}X_l^{0\bar{\sigma}_1}X_m^{\bar{\sigma}0}|X_g^{\sigma 0}\rangle\rangle_\omega \\ &+ \sum_{\substack{ml\sigma_1 \\ (l\neq f)}}\left(\frac{t_{fm}t_{ml}}{U}\right) \\ &\times\langle\langle(X_m^{\sigma_1\bar{\sigma}_1}X_l^{0\sigma_1} - X_m^{\sigma_1\sigma_1}X_l^{0\bar{\sigma}_1})(X_f^{\bar{\sigma}_1\sigma} + \delta_{\sigma_1\bar{\sigma}}X_f^{00})|X_g^{\sigma 0}\rangle\rangle_\omega.\end{aligned} \tag{7}$$

In this equation, the terms that explicitly contain multipliers proportional to t^2/U originate from the inclusion of three-center interactions. The scheme of further calculations is quite usual for the method of irreducible Green functions [22, 23]. Equation (7) is linearized with the introduction of the anomalous function $\langle\langle X_f^{\bar{\sigma}0} | X_g^{\sigma 0}\rangle\rangle$. Next, the equation of motion for $\langle\langle X_f^{\bar{\sigma}0} | X_g^{\sigma 0}\rangle\rangle$ is constructed and linearization is repeated. The $\langle X_f^{\sigma\sigma} X_g^{\sigma'\sigma'}\rangle$ means that appear in this scheme are approximated as [19]

$$\langle X_f^{\sigma\sigma} X_g^{\sigma'\sigma'}\rangle = \frac{1}{4}n^2 + \eta(\sigma)\eta(\sigma')\langle S_f^z S_g^z\rangle, \quad \eta(\sigma) = +1, -1, \quad \sigma = \uparrow, \downarrow. \tag{8}$$

Equation (8) is obtained from the exact equation

$$\langle X_f^{\sigma\sigma} X_g^{\sigma'\sigma'}\rangle = \frac{1}{4}n^2 + \frac{1}{2}\langle(\hat{n}_f - n)(\hat{n}_g - n)\rangle + \eta(\sigma)\eta(\sigma')\langle S_f^z S_g^z\rangle \tag{9}$$

by ignoring the correlator $\langle(\hat{n}_f - n)(\hat{n}_g - n)\rangle$. A similar approximation is used to represent the three-site mean in terms of quasi-spin correlators,

$$\langle(X_f^{00} + X_f^{\sigma\sigma})X_m^{\bar{\sigma}\bar{\sigma}}(X_g^{00} + X_g^{\sigma\sigma})\rangle = \frac{n}{2}\left(1 - \frac{n}{2}\right)^2 - \left(1 - \frac{n}{2}\right)(\langle S_f^z S_m^z\rangle + \langle S_m^z S_g^z\rangle) + \frac{n}{2}\langle S_f^z S_g^z\rangle. \tag{10}$$

The passage to the quasi-momentum representation yields a closed system of equations similar to the Gor'kov equations,

$$(E - \tilde{\varepsilon}_{\mathbf{k}} + \mu)\langle\langle X_{\mathbf{k}\sigma} | X_{\mathbf{k}\sigma}^+\rangle\rangle_E - \Delta_{\mathbf{k}}\langle\langle X_{-\mathbf{k}\bar{\sigma}}^+ | X_{\mathbf{k}\sigma}^+\rangle\rangle_E = 1 - \frac{n}{2}, \tag{11}$$

$$-(\Delta_{\mathbf{k}})^*\langle\langle X_{\mathbf{k}\sigma} | X_{\mathbf{k}\sigma}^+\rangle\rangle_E + (E + \tilde{\varepsilon}_{\mathbf{k}} - \mu)\langle\langle X_{-\mathbf{k}\bar{\sigma}}^+ | X_{\mathbf{k}\sigma}^+\rangle\rangle_E = 0,$$

where the renormalized spectrum of fermionic excitations is determined by the equation

$$\tilde{\varepsilon}_{\mathbf{k}} = \epsilon - \left(\frac{n^2}{16}\right)\frac{(1-n)}{(1-n/2)}J_0 + \left(1 - \frac{n}{2}\right)t_{\mathbf{k}} - \frac{n}{2}\left(1 - \frac{n}{2}\right)\frac{t_{\mathbf{k}}^2}{U} - \frac{1}{N}\sum_{\mathbf{q}}\left\{t_{\mathbf{q}} + \frac{n}{2}J_{\mathbf{k}-\mathbf{q}} + [(2-n)t_{\mathbf{k}} + (1-n)t_{\mathbf{q}}]\frac{t_{\mathbf{q}}}{U}\right\} \times \frac{K_{\mathbf{q}}}{1 - n/2} + \frac{1}{N}\sum_{\mathbf{q}}\left\{t_{\mathbf{k}-\mathbf{q}} + \frac{n}{2}J_{\mathbf{q}} + (2-n)\frac{t_{\mathbf{k}}t_{\mathbf{k}-\mathbf{q}}}{U} - \frac{n}{2}\frac{t_{\mathbf{k}-\mathbf{q}}^2}{U}\right\}\frac{C_{\mathbf{q}}}{1 - n/2}. \tag{12}$$

Here, $t_{\mathbf{q}}$ and $J_{\mathbf{q}}$ are the Fourier transforms of the hopping and exchange integrals, respectively. The renormalizations in spectrum (12) caused by three-center interactions are proportional to the ratio between the square of the hopping integral and the U parameter. The majority of them depend on the kinetic ($K_{\mathbf{q}} = \langle X_{\mathbf{q}\sigma}^+ X_{\mathbf{q}\sigma}\rangle$) and quasi-spin

$$C_{\mathbf{q}} = \sum_f \exp\{-i(\mathbf{R}_f - \mathbf{R}_m)\mathbf{q}\}\{\langle S_f^z S_m^z\rangle + \langle S_f^+ S_m^-\rangle\} \tag{13}$$

correlators. Solving (11) and applying the spectral theorem, we can find the kinetic correlator and the energy spectrum of the system

$$K_{\mathbf{q}} = \left(1 - \frac{n}{2}\right)\left(\frac{E_{\mathbf{k}} - \xi_{\mathbf{k}}}{2E_{\mathbf{k}}} + \frac{\xi_{\mathbf{k}}}{E_{\mathbf{k}}}f\left(\frac{E_{\mathbf{k}}}{T}\right)\right),$$

$$f(x) = (\exp x + 1)^{-1}, \quad E_{\mathbf{k}} = \sqrt{\xi_{\mathbf{k}}^2 + |\Delta_{\mathbf{k}}|^2}, \tag{14}$$

$$\xi_{\mathbf{k}} = \tilde{\varepsilon}_{\mathbf{k}} - \mu,$$

and the self-consistency equation for the superconducting order parameter $\Delta_{\mathbf{k}}$,

$$\Delta_{\mathbf{k}} = \frac{1}{N}\sum_{\mathbf{q}}\left\{2t_{\mathbf{q}} + \frac{n}{2}(J_{\mathbf{k}+\mathbf{q}} + J_{\mathbf{k}-\mathbf{q}}) + 4\left(1 - \frac{n}{2}\right)\frac{t_{\mathbf{k}}t_{\mathbf{q}}}{U} - n\frac{t_{\mathbf{q}}^2}{U}\right\}\frac{\Delta_{\mathbf{q}}}{2E_{\mathbf{q}}}\tanh\left(\frac{E_{\mathbf{q}}}{2T}\right). \tag{15}$$

4. SPHERICALLY SYMMETRICAL CORRELATION FUNCTIONS

Let us find the equation for the quasi-spin correlator to obtain a closure to the self-consistency equations. We will use the ideology of the quantum spin liquid [24–26] and write the equations of motion for Bose Green functions describing the dynamics of the spin degrees of freedom. The first equation of motion has the form

$$\omega\langle\langle X_{j+l}^{\sigma\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega = -\sum_n t_{nl}L_{nl}^{\sigma\bar{\sigma}}(\omega) + 2\sum_n J_{ln}L_{ln}^z(\omega) + \langle\langle[X_{j+l}^{\sigma\bar{\sigma}}, H_{(3)}] | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega, \tag{16}$$

where $L_{nl}^{\sigma\bar{\sigma}}(\omega)$ and $L_{nl}^z(\omega)$ are antisymmetric with

respect to l and n,

$$L_{nl}^{\sigma\bar{\sigma}}(\omega) = \langle\langle X_{j+n}^{\sigma 0} X_{j+l}^{0\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega - [l \longleftrightarrow n],$$
$$L_{ln}^{z}(\omega) = \frac{1}{2}\langle\langle (X_{j+l}^{\sigma\sigma} - X_{j+l}^{\bar{\sigma}\bar{\sigma}}) X_{j+n}^{\sigma\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega - [l \longleftrightarrow n]. \quad (17)$$

The $[l \longleftrightarrow n]$ symbol in the right-hand sides of Eqs. (17) stands for the terms obtained from the preceding terms by the exchange of the indices l and n. The last term in (16) results in the appearance of Green functions containing the product of three Hubbard operators. If two Hubbard operators have equal site indices, then, thanks to the algebra of Hubbard operators, such a higher Green function reduces to a lower one. However, if all site indices are different, the reduction to a lower Green function is performed following the uncoupling scheme

$$\langle\langle X_m^{\bar{\sigma}\bar{\sigma}} X_n^{\sigma 0} X_l^{0\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega \longrightarrow \langle X_m^{\bar{\sigma}\bar{\sigma}}\rangle \langle\langle X_n^{\sigma 0} X_l^{0\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega,$$
$$\langle\langle X_m^{\sigma\bar{\sigma}} X_l^{\sigma 0} X_n^{0\sigma} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega \longrightarrow \langle X_l^{\sigma 0} X_n^{0\sigma}\rangle \langle\langle X_m^{\sigma\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega. \quad (18)$$

After the Fourier transform, (16) takes the form

$$\omega G_q^{\sigma\bar{\sigma}}(\omega) = \frac{1}{N}\sum_k \left(-t_k + \frac{n}{2}\frac{t_k^2}{U}\right) L_{k,q-k}^{\sigma\bar{\sigma}}(\omega) + 2\frac{1}{N}\sum_k J_k L_{q-k,k}^{z}(\omega), \quad (19)$$

where we introduced the Fourier transforms of the corresponding Green functions,

$$\langle\langle X_{j+l}^{\sigma\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega = \frac{1}{N}\sum_q e^{iql} G_q^{\sigma\bar{\sigma}}(\omega),$$
$$L_{nl}^{\alpha}(\omega) = \frac{1}{N^2}\sum_{qp} e^{iqn+ipl} L_{qp}^{\alpha}(\omega), \quad \alpha = \{\sigma, \bar{\sigma}, z\}. \quad (20)$$

The higher Green functions $L_{nl}^{\sigma\bar{\sigma}}(\omega)$ and $L_{ln}^{z}(\omega)$ are calculated using a similar procedure. We omit cumbersome intermediate calculations and only give the final result.

The equation for the $L_{qp}^{\sigma\bar{\sigma}}(\omega)$ Green function can be written in the form

$$\omega L_{q,p}^{\sigma\bar{\sigma}}(\omega) = 2(K_q - K_p) + G_{q+p}^{\sigma\bar{\sigma}}(\omega)\left\{\frac{2}{U}\frac{1}{N}\sum_k t_k\left[t_q\left(K_k + \frac{2}{3}C_{k+q}\right) - t_p\left(K_k + \frac{2}{3}C_{k+p}\right)\right] + \frac{2}{U}t_q t_p (K_q - K_p) + \frac{n}{2}(J_0 + J_{q+p})(K_q - K_p) + \left(n\frac{t_q + t_p}{U} - 2\right)(t_p K_p - t_q K_q)\right\}$$
$$+ M_{q,p}^{\sigma\bar{\sigma}}(\omega)\left\{\left(1 - \frac{n}{2}\right)\left(t_p - t_q + \frac{n}{2}\frac{t_q^2 - t_p^2}{U}\right) + \frac{1}{N}\sum_k \frac{t_k}{U}\left[t_q\left(K_k - \frac{2}{3}C_{k+q}\right) - t_p\left(K_k - \frac{2}{3}C_{k+p}\right)\right] + \frac{1}{2}\frac{1}{N}\sum_k J_k (K_{k+p} - K_{k+q})\right\}$$
$$+ \left[\frac{n}{2}\left(1 - \frac{n}{2}\right)(t_p - t_q) + t_q K_q - t_p K_p\right] \times \frac{1}{U}\frac{1}{N}\sum_k t_k M_{q+p-k,k}^{\sigma\bar{\sigma}}(\omega) + (K_q - K_p) \times \frac{1}{N}\sum_k M_{q+p-k,k}^{\sigma\bar{\sigma}}(\omega)\left[\frac{t_k t_{q+p-k}}{U} - \frac{1}{2}J_{p-k}\right]$$
$$+ \frac{1}{U}\frac{1}{N}\sum_k t_k M_{q+p-k,k}^{\sigma\bar{\sigma}}(\omega)\frac{1}{N} \times \sum_{k_1} t_{k_1}\left[\frac{2}{3}C_{q+k_1} - \frac{2}{3}C_{p+k_1} + K_{q+k_1-k} - K_{p+k_1-k}\right], \quad (21)$$

where $M_q^{\sigma\bar{\sigma}}(\omega)$ is the Fourier transform of the higher Green function symmetrical with respect to the n and l indices,

$$M_{nl}^{\sigma\bar{\sigma}}(\omega) = \langle\langle X_{j+n}^{\sigma 0} X_{j+l}^{0\bar{\sigma}} | X_j^{\bar{\sigma}\sigma}\rangle\rangle_\omega + [n \longleftrightarrow l].$$

This equation can be simplified using the approximation applied in [25] to study the t–J model. As the contributions of the $M_{n,l}^{\sigma\bar{\sigma}}(\omega)$ functions to (21) are negligibly small if $n \neq l$ and

$$M_{l,l}^{\sigma\bar{\sigma}}(\omega) = 2G_n^{\sigma\bar{\sigma}}(\omega)$$

at $n = l$, we obtain the approximate equation

$$M_{n,l}^{\sigma\bar{\sigma}}(\omega) = 2\delta_{nl} G_n^{\sigma\bar{\sigma}}(\omega).$$

In the quasi-momentum representation, this equation is written as

$$M_{qp}^{\sigma\bar{\sigma}}(\omega) = 2G_{q+p}^{\sigma\bar{\sigma}}(\omega). \quad (22)$$

Equation (21) then takes the form that explicitly relates the higher Green function under consideration to the lower functions,

$$\begin{aligned}\omega L_{k,q-k}^{\sigma\bar{\sigma}}(\omega) &= 2(K_k - K_{q-k}) \\ &+ G_q^{\sigma\bar{\sigma}}(\omega)\Bigg[(t_k - t_{q-k})\Bigg(n - 2 + \frac{4}{U}\frac{1}{N}\sum_{k_1} K_{k_1} t_{k_1}\Bigg) \\ &+ n\Bigg(1 - \frac{n}{2}\Bigg)\frac{t_k^2 - t_{q-k}^2}{U} \\ &+ (K_k - K_{q-k})\Bigg(\frac{2t_k t_{q-k}}{U} + n\frac{J_0 + J_q}{2} + J_q\Bigg) \\ &+ (t_{q-k}K_{q-k} - t_k K_k)\Bigg(-2 + n\frac{t_k + t_{q-k}}{U}\Bigg)\Bigg].\end{aligned} \quad (23)$$

Applying approximation (22) to the Fourier transform of the third Green function $L_{q,k}^z(\omega)$ defined by (20) yields

$$\begin{aligned}\omega L_{q-k,k}^z(\omega) &= \frac{4}{3}(C_{q-k} - C_k) + \frac{4}{3}G_q^{\sigma\bar{\sigma}}(\omega) \\ &\times \frac{1}{N}\sum_{k_1} J_{k_1}(\tilde{C}_{q-k+k_1} - \tilde{C}_{k-k_1}) \\ &+ \frac{4}{3}G_q^{\sigma\bar{\sigma}}(\omega)[J_q\tilde{C}_{q-k} - J_{q-k}\tilde{C}_{q-k} - J_q\tilde{C}_k + J_k\tilde{C}_k] \\ &+ G_q^{\sigma\bar{\sigma}}(\omega)\Bigg[\frac{n}{2}(J_{q-k} - J_k) + \frac{2}{U}\frac{1}{N} \\ &\times \sum_{k_1} t_{k_1}(t_{k-k_1}K_{k-k_1} - t_{q-k-k_1}K_{q-k-k_1})\Bigg],\end{aligned} \quad (24)$$

where $\tilde{C}_q = \alpha C_q + (1 - \alpha)3n/4$. As in [24–26], we here introduced the vertex correction α according to the equation

$$\tilde{C}_{nl} = (\alpha(1 - \delta_{nl}) + \delta_{nl})C_{nl}.$$

This procedure allows the errors of uncoupling to be corrected by imposing the requirement of the fulfillment of the corresponding sum rules. In our problem, the sum rule role is played by the condition $C_{ll} = 3n/4$. Note that all terms in the last two rows of (24) appear because of the interaction $H_{(3)}$.

The system of three equations, (19), (23), and (24), allows us to find the quasi-spin correlator. To simplify the final analytic equations, we use the small parameters of the system $\lambda = 2|t|/U \ll 1$ and $\delta = (1 - n) \ll 1$ (below, we only consider the region of low hole doping). It is easy to see that $K_p = (1 - n/2)f(\varepsilon_p) \longrightarrow 1/2$ as $n \longrightarrow 1$, and all terms related to the kinematics of the problem are therefore canceled in (23) and (24). The remaining three rows of (24) describe spin correlations in the Heisenberg limit [24, 26]. Clearly, all terms related to kinematics are proportional to either the concentration of holes $\delta = 1 - n$ or kinematic correlators K_f ($f \neq 0$), which are also proportional to $(1 - n)$. These two parameters have equal orders of smallness at the characteristic λ value $\lambda = 0.25$ and the concentration of electrons $n > 0.75$. Calculations show that, in the region of low doping, the kinematic correlators K_f ($f \neq 0$) are much smaller than λ and magnetic correlators C_f (at $f \neq 0$).

The above considerations allow us to simplify equations by retaining terms of order δ, $\lambda\delta$, and λ^2 only and ignoring all terms of order $\lambda^2\delta$. The Green function $G_q^{\sigma\bar{\sigma}}(\omega)$ found from (19), (23), and (24) then takes the eventual form

$$\begin{aligned}C_q^{\sigma\bar{\sigma}}(\omega) &= \frac{A_q(\omega)}{\omega^2 - \omega^2(q)}, \\ A_q(\omega) &= \frac{1}{N}\sum_k\Bigg[\Bigg(-2t_k + \frac{n}{U}t_k^2\Bigg)(K_k - K_{q-k}) \\ &+ \frac{8}{3}J_k(C_{q-k} - C_k)\Bigg],\end{aligned} \quad (25)$$

where the spectrum of magnetic excitations $\omega(q)$ is determined by the equation

$$\begin{aligned}\omega^2(q) &= \frac{1}{N}\sum_k\Bigg[t_k(t_k - t_{q-k})\Bigg((2 - n)\Bigg(1 - n\frac{t_k}{U}\Bigg) \\ &+ 2nK_k\frac{t_k + t_{q-k}}{U} - \frac{4}{U}\frac{1}{N}\sum_{k_1} t_{k_1}K_{k_1}\Bigg) \\ &+ 2t_k(t_{q-k}K_{q-k} - t_k K_k) - t_k(K_k - K_{q-k}) \\ &\times\Bigg(\frac{2t_k t_{q-k}}{U} + n\frac{J_0 + J_q}{2} + J_q\Bigg) \\ &+ \frac{8}{3}J_k\frac{1}{N}\sum_{k_1} J_{k_1}(\tilde{C}_{q-k+k_1} - \tilde{C}_{k-k_1}) \\ &+ \frac{8}{3}J_k(J_q\tilde{C}_{q-k} - J_{q-k}\tilde{C}_{q-k} - J_q\tilde{C}_k + J_k\tilde{C}_k)\Bigg].\end{aligned} \quad (26)$$

Only the last two rows of (26) remain in the limit $n \longrightarrow 1$, which corresponds to the spectrum of magnons in the Heisenberg model [24, 26].

Applying the spectral theorem in the usual way leads to self-consistent equations for magnetic correlators. By numerically solving these equations, we can calculate correlator values. After this, spectrum renormalizations caused by magnetic fluctuations and their influence on the superconducting transition temperature are determined.

5. THE SPECTRUM AND DENSITY OF STATES OF FERMIONIC EXCITATIONS IN THE t–J^* MODEL

It is well known [3–5] that high-temperature superconductivity conditions are to a substantial extent determined by the special features of the energy spectrum of Fermi quasi-particles. These features can be responsible for the singular behavior of the density of states of charge carriers. Bearing this in mind, let us first comparatively analyze the renormalizations of the energy spectrum caused by simultaneous magnetic fluctuation and three-center interaction effects.

Numerical calculations were performed in the nearest-neighbor approximation (the influence of long-range hoppings is in part discussed in the concluding section). Generally, a system of ten self-consistent transcendental equations was solved. This system determined ten values: three magnetic and five kinetic correlators, the chemical potential, and the vertex renormalization α. For convenience of comparing the results, we first give the fermionic energy spectrum of the t–J model without the inclusion of magnetic correlators,

$$\tilde{\varepsilon}_H(\mathbf{k}) = -4|t|\left[1 - \frac{n}{2} + \frac{\lambda K_1}{1-n/2}\right]\gamma_1(\mathbf{k}),$$
$$\gamma_1(\mathbf{k}) = \frac{1}{2}(\cos(k_x a) + \cos(k_y a)), \tag{27}$$

where a is the distance between square lattice sites.

If three-center interactions are included (the t–J^* model), the quasi-momentum dependence of the energy spectrum $\tilde{\varepsilon}_{tJ^*}(\mathbf{k})$ is described by a linear superposition of three invariants for a square lattice. If magnetic fluctuations are ignored, the fermionic spectrum is determined by the equation

$$\tilde{\varepsilon}'_{tJ^*}(\mathbf{k}) = -4|t|\left[1 - \frac{n}{2} + \frac{(4-3n/2)\lambda K_1}{1-n/2}\right]\gamma_1(\mathbf{k})$$
$$-|t|\lambda n\left(1-\frac{n}{2}\right)[2\gamma_2(\mathbf{k}) + \gamma_3(\mathbf{k})]. \tag{28}$$

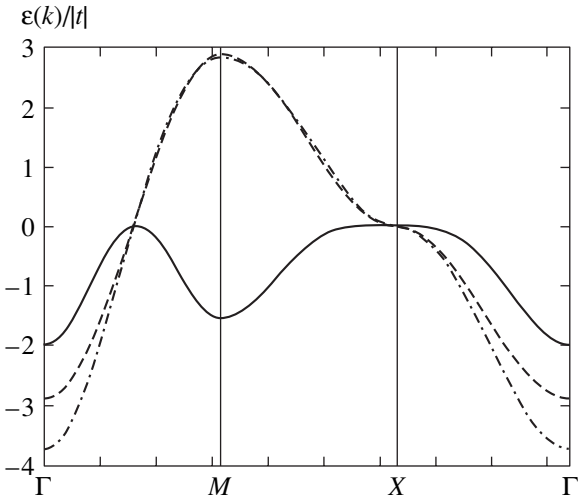

Fig. 1. Spectrum of fermionic excitations in the Hubbard model for various approximations under strong electron correlation conditions.

The appearance of two new invariants compared with (27),

$$\gamma_2(\mathbf{k}) = \cos(k_x a)\cos(k_y a),$$
$$\gamma_3(\mathbf{k}) = \frac{1}{2}(\cos(2k_x a) + \cos(2k_y a)), \tag{29}$$

formally corresponds to the presence of effective hoppings between sites of distant coordination spheres in the system. The physical origin of such hoppings is fairly simple to explain if the operator structure of three-center interactions is taken into account.

The self-consistent calculations of the kinetic correlator for spectrum (28) at $n = 0.92$ give $K_1 = 0.0628$. The corresponding calculated dispersion dependence of excitation energy is shown by the dot-and-dash line in Fig. 1. The dispersion dependences are given with the standard denotations of the distinguished Brillouin zone points, $\Gamma = (0, 0)$, $X = (\pi, 0)$, and $M = (\pi, \pi)$. A comparison of the dot-and-dash line with the dashed line corresponding to spectrum (27) shows that, quantitatively, spectrum renormalizations are insignificant under these conditions. This conclusion fully corresponds to the results reported in [10, 11].

A quite different result is obtained if not only three-center interactions but also magnetic correlations are included. The equation for the energy spectrum then takes the form

$$\tilde{\varepsilon}_{tJ^*}(\mathbf{k}) = -4|t|\left[1 - \frac{n}{2} + \frac{(4-3n/2)\lambda K_1 + C_1}{1-n/2}\right]\gamma_1(\mathbf{k})$$
$$-2|t|\lambda\left[n\left(1-\frac{n}{2}\right) - 4C_1 + \frac{nC_2}{1-n/2}\right]\gamma_2(\mathbf{k}) \tag{30}$$
$$-|t|\lambda\left[n\left(1-\frac{n}{2}\right) - 4C_1 + \frac{nC_3}{1-n/2}\right]\gamma_3(\mathbf{k}),$$

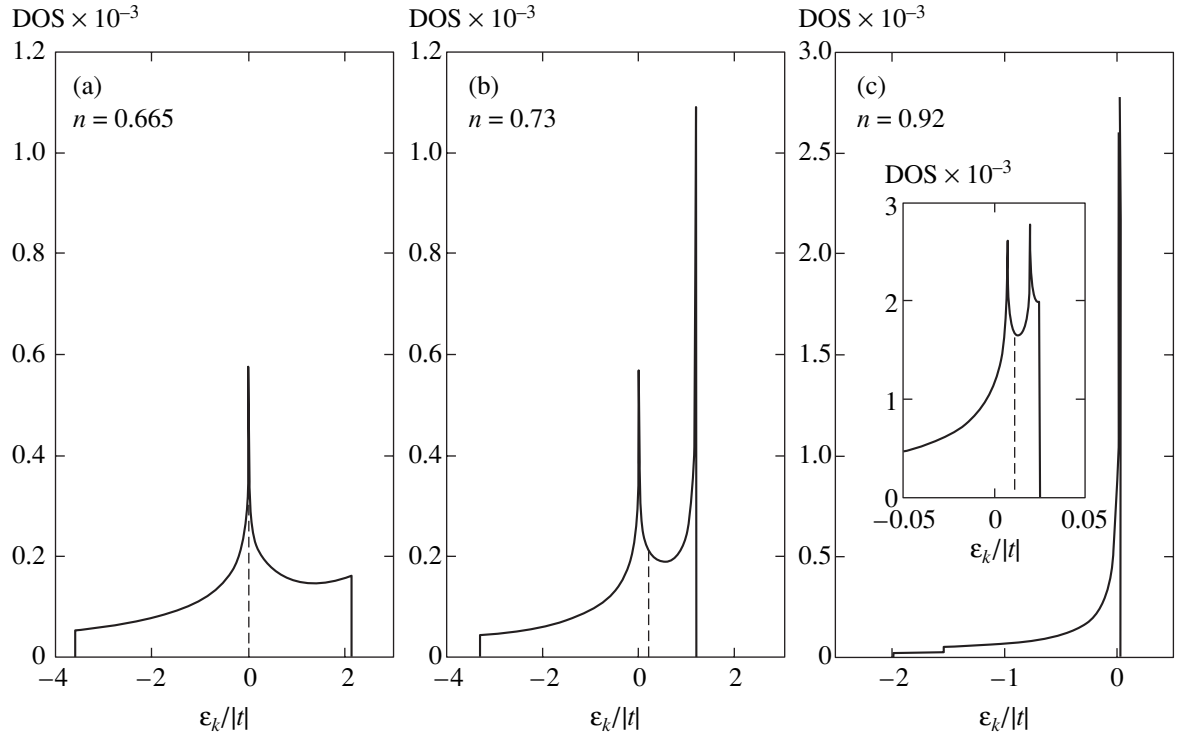

Fig. 2. Evolution of the density of fermionic states in the t–J* model depending on the concentration n. The structure of the density of states at $n = 0.92$ (two closely spaced peaks) is shown in the inset in Fig. 2c on a smaller scale. The vertical dashed line corresponds to the chemical potential position.

where C_2 and C_3 are the magnetic correlators for the second and third coordination sphere, respectively [26]. The self-consistent calculations performed with this energy spectrum for $n = 0.92$ give the correlator values $C_1 = -0.2639$, $C_2 = 0.1347$, $C_3 = 0.1115$, and $K_1 = 0.0037$. The corresponding quasi-momentum dependence of the energy spectrum is shown by the solid line in Fig. 1. The principal feature that qualitatively distinguishes this spectrum from the two previous ones is the presence of a minimum at the M Brillouin zone point. This minimum only appears at $n > n_1$. The n_1 value depends on the model parameters. According to the self-consistent calculations for spectrum (30) with $\lambda = 0.25$, $n_1 \approx 0.72$.

To describe the renormalizations of the spectrum mentioned above, let us consider the evolution of the density of states as n varies. The density of states calculated self-consistently for $n = 0.665$ is shown in Fig. 2a. As $n < n_1$, there is only one well-known Van Hove singularity corresponding to the Brillouin zone saddle points X. It is present in all the spectra considered above at all concentrations n. The states with the energies corresponding to this Van Hove singularity are populated at $n \approx 0.66$. This is why the theoretical concentration dependences $T_c(n)$ contain maxima at $n \approx 0.66$ (see Fig. 3). The structure of the density of states at $n > n_1$ is shown in Fig. 2b. The second Van Hove singularity then appears close to the top of the zone. It is formed when the spectral curve has a minimum at the M Brillouin zone points provided $n = n_1$ (at $n < n_1$, the kinetic and spin correlator values are such that dispersion dependence (30) has a maximum rather than a minimum at the M Brillouin zone points). Just the appearance of the local minimum at $n \approx n_1$ induces the new logarithmic singularity of the density of states. This Van Hove singularity is retained as the concentration increases up to $n = 1$. At $n \approx n_1$, the two singularities are spaced fairly far apart on the energy scale. The distance between them, however, decreases as n increases because the top of the zone lowers. The structure of the density of states shown in Fig. 2c corresponds to the concentration at which T_c is maximum (see solid curve in Fig. 3). The distance between the peaks of the density of states is then commensurate with the critical temperature T_c. In addition, the chemical potential and temperature then have values at which the contribution to thermodynamics is determined by all states in the vicinity of the top of the zone. To conclude this section, note again that the new peak of the density of states in the nearest-neighbor approximation only appears when both spin correlations and three-center interactions are included simultaneously.

6. THE CONCENTRATION DEPENDENCE OF THE SUPERCONDUCTING TRANSITION TEMPERATURE

Clearly, the special features of the energy spectrum mentioned above should manifest themselves in many

t–J* model characteristics. By way of example, let us consider the concentration dependence of the transition temperature to the superconducting state with the order parameter of $d_{x^2-y^2}$-type symmetry.

The superconducting state can be formed if integral equation (15) has a nontrivial solution $\Delta_{\mathbf{k}} \neq 0$. The kernel of this equation is the sum of four terms. The first term, which is proportional to $2t_{\mathbf{q}}$, corresponds to the kinematic mechanism of pairing [2]. The origin of the second kernel term is both exchange and three-center interactions. Exchange interaction gives the contribution proportional to $(J_{\mathbf{k}+\mathbf{q}} + J_{\mathbf{k}-\mathbf{q}})$, and three-center interaction introduces a correction proportional to $(-1 + n/2)(J_{\mathbf{k}+\mathbf{q}} + J_{\mathbf{k}-\mathbf{q}})$, which tends to suppress superconductivity. Because of the superposition of these contributions, the coefficient of the term proportional to $(J_{\mathbf{k}+\mathbf{q}} + J_{\mathbf{k}-\mathbf{q}})$ equals the renormalization factor $n/2$ [13] rather than one as in the t–J model. Precisely this renormalization of the coupling constant in the t–J* model substantially suppresses T_c [14].

As is well known, (15) reduces to a transcendental equation in the nearest-neighbor approximation. In what follows, we only consider superconducting phases with order parameter of $d_{x^2-y^2}$ symmetry. Its quasi-momentum dependence has the form $\Delta(\mathbf{k}) = \Delta_0(\cos k_x a - \cos k_y a)$. The equation determining the temperature of the superconducting transition then takes the form

$$1 = \frac{n}{2}\frac{J}{N}\sum_{\mathbf{q}} \frac{(\cos q_x a - \cos q_y a)^2}{\tilde{\varepsilon}_{\mathbf{q}} - \mu} \tanh\left(\frac{\tilde{\varepsilon}_{\mathbf{q}} - \mu}{2T_c}\right). \quad (31)$$

We see that, for $d_{x^2-y^2}$ symmetry, the first, third, and fourth terms of the kernel of (15) make no contribution. Three-center interactions therefore manifest themselves only by renormalizing the coupling constant and modifying the spectrum of fermionic excitations.

The solid curve shown in Fig. 3 was obtained by numerically solving (31). To better visualize the effects under consideration, we plotted the concentration dependence of the critical temperature T_c for spectrum (27) of the t–J model (dashed line) and the $T_c(n)$ dependence for the t–J* model without taking magnetic correlations into account [14] (dot-and-dash line) in the same figure. We see that simultaneously including three-center interactions and magnetic correlations changes the $T_c(n)$ dependence qualitatively, namely, a second maximum appears in the low doping region, which corresponds to a substantial increase in the number of electrons participating in Cooper pairing. Figure 2c shows that this increase is related to a considerable growth of the density of fermionic states in the vicinity of the Fermi level and to effective broadening of the energy region that makes the major contribution to the integration. The second $T_c(n)$ maximum at $n = 0.92$ is much higher than the first one ($n = 0.66$) because of a much larger number of electrons in the vicinity of the chemical potential and a large increase in the effective coupling constant value caused by its dependence on n.

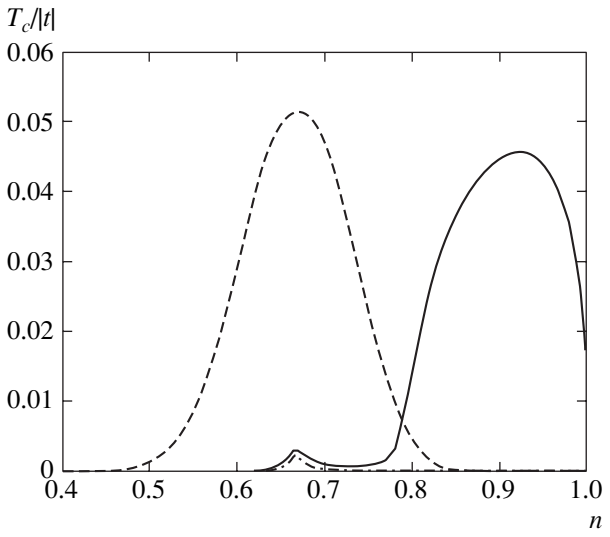

Fig. 3. Concentration dependences of the critical temperature for various approximations.

7. CONCLUSIONS

The results presented above show that joint three-center interaction and magnetic correlation effects play an important role in the t–J* model. A new Van Hove singularity is then induced in the density of states of Fermi quasi-particles at high energies. As this singularity appears at low hole concentrations, the suggestion can be made of the magnetopolaron nature of the induction of the new Van Hove singularity. This suggestion correlates with recent results obtained in studies of the electronic structure of a strongly correlated spin-fermion liquid with invoking an extended basis set for including magnetopolaron states [27]. The second argument in favor of this hypothesis is the structure of $H_{(3)}$, which relates electron hopping to the spin dynamics of neighboring sites. It is clear in view of these considerations why the inclusion of $H_{(3)}$ without taking magnetic correlations into account did not cause substantial changes in the electron energy spectrum and the density of states of fermionic excitations.

Changes in the energy spectrum were considered for the example of the concentration dependence of the superconducting transition temperature. Clearly, such qualitative changes in the density of fermionic states should also result in other noticeable changes in the thermodynamic properties of the system. We also think it important that the reproduction of energy spectrum parameters, for instance, from the angular resolution photoemission spectra (ARPES) requires taking into account the renormalizations described above. Note that, although our analysis was limited to the use of the nearest-neighbor approximation, long-range hoppings

were effectively included. This circumstance should substantially influence the values of the parameters being reproduced.

Including the kinetic energy of long-range hoppings (t' and t'') into the Hamiltonian can strengthen as well as suppress the new Van Hove singularity in the region of low hole concentrations. The corresponding calculations are a separate problem; they must include new magnetic correlators between sites from more distant coordination spheres (up to the ninth sphere inclusive). Solving the self-consistent problem for the values of these correlators is beyond the scope of the present work. As concerns the influence of long-range hoppings on the conditions of superconducting phase existence, changes in the kernel of the integral equation should primarily be taken into account [28]. The renormalizations of effective hoppings considered in this work can then also influence the concentration dependence $T_c(n)$. Note one more subtlety. Formally, the effect of three-center interactions related to inducing hoppings between sites from distant coordination spheres can be modeled by going beyond the scope of the nearest-neighbor approximation but without taking $H_{(3)}$ into account. However, we then lose the second effect of $H_{(3)}$, essential to superconductivity, which is related to the renormalization of the effective coupling constant [13] and responsible for the dependence of T_c on n just such as is shown in Fig. 3.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research jointly with KKFN "Eniseĭ" (project no. 02-02-97705), the Russian Foundation for Basic Research (project no. 03-02-16124), INTAS (grant no. 01-0654), and the Complex Program "Quantum Macrophysics" (Russian Academy of Sciences). One of us (D.M.D.) thanks the Charity Foundation for Promoting Science and Lavrent'ev competition of projects of young scientists of the Siberian Division of the Russian Academy of Sciences.

REFERENCES

1. P. W. Anderson, Science **235**, 1196 (1987).
2. R. O. Zaĭtsev and V. A. Ivanov, Pis'ma Zh. Éksp. Teor. Fiz. **46**, 140 (1987) [JETP Lett. **46**, S116 (1987)].
3. N. M. Plakida, *High-Temperature Superconductors* (Springer, Moscow, 1995; Mezhdun. Progr. Obraz., Moscow, 1996).
4. Yu. A. Izyumov, Usp. Fiz. Nauk **167**, 465 (1997) [Phys. Usp. **40**, 445 (1997)].
5. S. G. Ovchinnikov, Usp. Fiz. Nauk **167**, 1043 (1997) [Phys. Usp. **40**, 993 (1997)].
6. J. Hubbard, Proc. R. Soc. London, Ser. A **276**, 238 (1963).
7. M. A. Baranov, A. V. Chubukov, and M. Yu. Kagan, Int. J. Mod. Phys. B **6**, 2471 (1992).
8. L. P. Bulaevskiĭ, É. L. Nagaev, and D. L. Khomskiĭ, Zh. Éksp. Teor. Fiz. **54**, 1562 (1968) [Sov. Phys. JETP **27**, 836 (1968)].
9. K. A. Chao, J. Spalek, and A. M. Oles, J. Phys. C **10**, L271 (1977).
10. Q. L. Li, E. C. Koltenbah, and R. Joynt, Phys. Rev. B **48**, 437 (1993).
11. V. Yu. Yushankhai, V. S. Oudovenko, and R. Hayn, Phys. Rev. B **55**, 15562 (1997).
12. J. E. Hirsch, Phys. Lett. A **136**, 153 (1989).
13. V. Yu. Yushankhai, G. M. Vujicic, and R. B. Zakula, Phys. Lett. A **151**, 254 (1990).
14. V. V. Val'kov, T. A. Val'kova, D. M. Dzebisashvili, and S. G. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 450 (2002) [JETP Lett. **75**, 378 (2002)].
15. A. P. Kampf, Phys. Rep. **249**, 219 (1994).
16. Yu. A. Izyumov, Usp. Fiz. Nauk **169**, 225 (1999) [Phys. Usp. **42**, 215 (1999)].
17. N. M. Plakida, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 38 (2001) [JETP Lett. **74**, 36 (2001)].
18. R. O. Kuzian, R. Hayn, A. F. Barabanov, and L. A. Maksimov, Phys. Rev. B **58**, 6194 (1998).
19. M. M. Korshunov, S. G. Ovchinnikov, and A. V. Sherman, Pis'ma Zh. Éksp. Teor. Fiz. **80**, 45 (2004) [JETP Lett. **80**, 39 (2004)].
20. N. N. Bogolyubov, *Lectures on Quantum Statistics* (Naukova Dumka, Kiev, 1949; Gordon and Breach, London, 1967, 1970), Vols. 1, 2.
21. G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987).
22. N. M. Plakida, V. Yu. Yushankhay, and I. V. Stasyuk, Physica C (Amsterdam) **162–164**, 787 (1989).
23. V. Yu. Yushankhay, N. M. Plakida, and P. Kalinay, Physica C (Amsterdam) **174**, 401 (1991).
24. H. Shimahara and S. Takada, J. Phys. Soc. Jpn. **60**, 2394 (1991).
25. H. Shimahara and S. Takada, J. Phys. Soc. Jpn. **61**, 989 (1992).
26. A. F. Barabanov and V. M. Berezovskiĭ, Zh. Éksp. Teor. Fiz. **106**, 1156 (1994) [JETP **79**, 627 (1994)].
27. A. F. Barabanov, A. A. Kovalev, O. V. Urazaev, and A. M. Belemouk, Phys. Lett. A **265**, 221 (2000).
28. V. V. Val'kov and D. M. Dzebisashvili, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 450 (2003) [JETP Lett. **77**, 381 (2003)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 3, 2005, pp. 617–634.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 3, 2005, pp. 696–716.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Calculation of the Equation of State of a Dense Hydrogen Plasma by the Feynman Path Integral Method

S. V. Shevkunov

St. Petersburg State Polytechnic University, St. Petersburg, 195251 Russia
e-mail: root@shevk.hop.stu.neva.ru

Received June 20, 2004

Abstract—A method is developed for calculating the equation of state of a system of quantum particles at a finite temperature, based on the Feynman formulation of quantum statistics. A general analytical expression is found for the virial estimator for the kinetic energy of a system with rigid boundaries at a finite pressure. An effective method is developed for eliminating the unphysical singularity in the electrostatic potential between a discretized Feynman path of an electron and a proton. It is shown that the "refinement" of an expansion of a quantum-mechanical propagator by addition of high powers of time exacerbates, rather than eliminates, the divergence of a Feynman path integral. A brief summary of the current status of the problem is presented. The proposed new approaches are presented in relation to progress made in this field. Path integral Monte Carlo simulations are performed for nonideal hydrogen plasmas in which both indistinguishability and spin of electrons are taken into account under conditions preceding the formation of the electron shells of atoms. The electron permutation symmetry is represented in terms of Young operators. It is shown that, owing to the singularity of the Coulomb potential, quantum effects on the behavior of the electron component cannot be reduced to small corrections even if the system must be treated as a classical system according to the formal de Broglie criterion. Quantum-mechanical delocalization of electrons substantially weakens the repulsion between electrons as compared to protons. In relatively cold plasmas, many-body correlations lead to complex behavior of the potential of the average force between particles and give rise to repulsive forces acting between protons and electrons at distances of about 5 angstroms. Plasma pressure drops with decreasing plasma temperature as the electron shells of atoms begin to form, and the electron kinetic energy reaches a minimum at a temperature of about 31 000 K. The minimum point weakly depends on plasma density. Owing to quantum effects, the electron component is "heated" well before electrons are completely bound in the field of protons. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Thermodynamic properties of hydrogen plasmas are of special importance for controlled fusion research. At the same time, the calculation of equilibrium properties of a dense plasma is one of the most difficult fundamental problems in statistical mechanics.

Substantial progress in studies of the ion-plasma properties has recently been made by applying computer simulation methods [1, 2]. The ion–electron plasma has been much less thoroughly studied. The presence of the electron component precludes the use of classical statistical mechanics and substantially complicates modeling procedures. The formulation of a well-founded method for simulating quantum many-particle systems is hampered by fundamental difficulties, and its numerical implementation requires enormous computing resources. The principal difficulty lies in the quantum behavior of hydrogen plasmas, whose properties do not approach the classical limit with increasing temperature. This is clear even from the fact that a classical system of point charges is unstable at any temperature T and the corresponding partition function is divergent at the singularity points of the Coulomb potential:

$$\int_0^\infty \exp\left(\frac{e^2}{k_B T r}\right) 4\pi r^2 dr = \infty.$$

Convergence of the partition function for a hydrogen plasma is entirely due to the quantum behavior of its electron component. Accordingly, the quantum contributions cannot be treated as small corrections at any temperature. The energy of the system is mainly determined by short-range (unscreened) interaction between protons and electrons, which strongly depend on the quantum nature of electron motion.

The difficulties of rigorous quantum statistical treatment of many-particle systems stimulate the development of various approximate approaches. The most widely known example is the density functional theory (DFT) [3], where the energy of a system is written as a functional of particle density. The terms that represent exchange-correlation energy and quantum delocalization are expressed by phenomenological formulas known as local-density approximation [4] and gradient-

corrected local-density approximation [5]. In numerous versions of this approach [6–9] and related theories (so-called jellium models), the electron–electron interaction is replaced by the interaction of an electron with a continuum; i.e., a many-body problem is reduced to the simplified model of a single quantum particle moving in the field generated by the remaining particles. The one-electron approximation makes it possible to apply the occupation number formalism. However, rigorous treatment of exchange-correlation energy and spin states lies, a fortiori, outside the scope of this approach. Since the many-body wavefunction is factorized into the product of single-particle ones in the one-electron approximation, the Coulomb particle–particle correlations vanish. The correlations are usually taken into account by introducing corrections. As the electron-gas density decreases from values characteristic of electron shells to values characteristic of the space between ions, the electron kinetic energy decreases. This leads to a higher role played by electron–electron correlations in the plasma as compared to those in electron shells. Therefore, detailed description of electron–electron correlations is of particular importance for statistical thermodynamics of plasmas.

Currently, the only alternative to the one-electron approximation that can be used to obtain numerical results for quantum many-particle systems at finite temperatures and that explicitly takes into account both exchange and spin state is the path integral method. Path integrals are calculated numerically as high-dimensional integrals by Monte Carlo methods. Importance sampling is used to overcome the difficulties due to the high dimension of the domain of integration and obtain essentially exact results. In this respect, path integral computation can be classified as a reference method. However, path integral simulations are labor-intensive and may require hundreds of hours of CPU time. Wide application of the method is limited by certain unresolved issues. The most serious difficulties are discussed in the next section.

2. STATUS OF THE PROBLEM

The first attempts at computing thermodynamic properties by numerical path integration were made by Fosdick and Jordan [10, 11]. The first applications of path integral methods in analyses of equilibrium properties of hydrogen plasmas were reported in [12–19]. Slater determinants were used to allow for permutation symmetry. The effects of plasma nonideality on electron-gas degeneracy were analyzed in [16, 17]. The numerical results obtained in [18] were used to obtain phase diagrams for hydrogen plasmas. In [15], an extension of the path integral method to an open statistical system was proposed. Contributions of alternating sign to the partition function for a system of indistinguishable particles were discussed in [19].

In [20], a finite-dimensional approximation of Green's function was used to increase the time step without compromising statistical accuracy. In essence, the integral form proposed therein is a short-time approximation of a matrix element of the evolution operator, in which interaction is distributed over a path. Distributed interaction was originally used in computations of electron shells by a path integral method [21]. A similar problem was solved by a similar method in [22], where an alternative short-time approximation of a matrix element of the propagator was proposed.

The most difficult problem in the development of path integral methods is the indistinguishability of quantum particles. The problem has two facets: consistent representation of the partition function as a linear combination of connected-path diagrams and the so-called "problem of negative signs." A wavefunction symmetrized with respect to particle permutations is a linear combination of terms corresponding to respective connected-path diagrams in operator matrix elements. The coordinate part of a wavefunction that is antisymmetric under simultaneous permutation of spin variables and coordinates may be neither symmetric nor antisymmetric. Different spin states are associated with different permutation symmetries of the coordinate part. The partition function must be calculated by using a complete set of linearly independent basis functions. Equilibrium averages cannot be computed by using wavefunctions that are only symmetrized over permutations, because completeness and linear independence of basis functions are key requirements. Even though an orthonormal basis in the space of spin wavefunctions can in principle be constructed by using Young operators [23, 24], the complexity of a direct numerical implementation of the corresponding algorithm increases with the particle number N faster than $N!$. For this reason, the exact algorithm using Young operators to represent wavefunctions symmetrized under permutation is almost never employed in computations. When a pure quantum state is computed, the permutation symmetry of the coordinate part of the wavefunction is represented by an approximate expression. Electron-gas wavefunctions are most frequently expressed in terms of Slater determinants. A Slater determinant consists of two blocks of dimensions N_1 and N_2, which correspond to electrons in the spin-up and spin-down states, respectively. However, in general quantum theory [23], the permutation symmetries of the mutually complementary spin- and coordinate-dependent parts of a wavefunction are determined by an eigenvalue of a spin-squared operator rather than spin-projection operator. For a system of spin-1/2 fermions, there is one-to-one correspondence between each eigenfunction of the spin-squared operator and a Young tableau specifying a type of permutation symmetry. Since a Slater determinant is not equivalent to any Young operator, the corresponding wavefunction is not an eigenfunction of the spin-squared operator. Moreover, it may not even be a linear combination of such eigenfunctions, because a permutation of coordinates that belong to different blocks in a Slater determinant

cannot be reduced to the mere reversal of the sign of a total wavefunction, as dictated by the Pauli principle. Thus, the representation of wavefunctions in terms of Slater determinants is an uncontrollable approximation. Such functions do not make up a complete set of linearly independent basis functions, and the trace of the statistical operator calculated by using such functions is not a partition function in the strict sense of the term.

Until recently, construction of complete sets of basis wavefunctions for particles with definite spin in terms of path integrals presented a serious problem. Most results reported in the literature were obtained either for spin-zero bosons [25–33] or for hypothetical spinless fermions [26, 27, 34–37]. In both cases, there is no spin dependence, and the coordinate-dependent wavefunction is either completely symmetric or completely antisymmetric under permutation. However, spinless fermions do not exist in nature. The primary importance of the spin variable for stability of a system is clear from the simplest example of a pair of hydrogen atoms, which are either bound up into a molecule in the singlet state ($S = 0$) or mutually repelled in the unstable triplet state ($S = 1$).

An approximate representation in terms of Slater determinants was used in [38] to analyze ferromagnetic polarization in an electron gas at low temperatures. The Hamiltonian of the system is

$$\hat{H} = -\frac{1}{r_s^2}\sum_i \nabla_i^2 + \frac{2}{r_s}\sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \text{const},$$

with $r_s = a/a_0$ denoting a density parameter. Here, $4\pi\rho a^3/3 = 1$, ρ is the electron gas density, and a_0 is the Bohr radius. In the case of a small r_s (high density), the Hamiltonian is dominated by the kinetic energy term, and electrons behave as an ideal gas. In the opposite limit of a large r_s, electrons condense into a Wigner crystal [39]. The first-order phase transition between these states should occur at $r_s \approx 100$. As the system approaches the crystallization transition, the electron spins freeze into a ferromagnetic state, since the lowest energy may correspond to a nonzero total spin of the system. In [38], the ground state was sought by varying the spin-up and spin-down block sizes, N_1 and N_2, within the class of wavefunctions having the form of Slater determinants. It was found that the fully polarized state ($N_2 = 0$) is most stable at $r_s > 60$. The numerical results reported therein cannot be used to estimate the error due to approximate representation of permutation symmetry, which plays a key role in the problem.

An exact procedure for introducing spin into a statistical model of a system of spin-1/2 fermions in terms of path integrals was developed in [21, 40–47], where a complete set of symmetrized basis wavefunctions and the corresponding set of connected-path diagrams were obtained by using Young operators. In [45], the diagrams were classified and their combinatorial weights were calculated. Direct computations are feasible only for few-body systems ($N < 10$), because their computational complexity increases faster than $N!$. The limit in particle number was eliminated in [46], where the sampling technique was extended to the calculation of combinatorial weights by random walk over the set of diagrams. Numerical tests were performed, and the mechanism underlying the spin state of a disordered system at a finite temperature was analyzed. In [21], the method was tested by simulating thermally excited states of the electron shells of the simplest ions. The ground state of the electron shell of the hydrogen molecule was modeled in [41, 43]. In [46], correlation functions were computed for a dense hydrogen plasma.

The difficulties of representation of permutation symmetry stimulate the development of various approximate schemes. An approach combining DFT with a path integral method was proposed in [48]. In DFT calculations, a system of interacting particles is replaced by a noninteracting system where each particle moves in the average field generated by the remaining particles and the corresponding Hamiltonian is a functional of the total density matrix:

$$\mathcal{H} = T([\rho]) + V_{\text{ion}}([\rho]) + V_{\text{H}}([\rho]) + V_{\text{XC}}([\rho]).$$

Here, the kinetic energy $T([\rho])$ and exchange-correlation energy $V_{\text{XC}}([\rho])$ are represented by empirical formulas; $V_{\text{ion}}([\rho])$ and $V_{\text{H}}([\rho])$ are the potentials of the electron–ion and electron–electron Coulomb interactions, respectively. The matrix elements of the statistical operator for a system of noninteracting fermions are written by using the expression for occupation numbers in the coordinate representation:

$$\rho(x, x'; \beta) = \left\langle x\left|\frac{1}{\exp[\beta(\hat{H} - \mu) + 1]}\right|x'\right\rangle,$$

where β is the inverse temperature. The chemical potential μ is calculated iteratively by using this expression for the diagonal matrix element $\rho = \rho(x, x; \beta)$. A property of meromorphic functions is used to calculate the off-diagonal elements $\rho(x, x'; \beta)$. A path-integral representation is applied after changing to complex temperature. Thus, a self-consistent calculation of μ would allow one to avoid summation of the diagrams corresponding to connected paths, which is required in the exact theory. Despite the appealing simplicity of this approach, it cannot be considered as a final solution to the exchange problem. In the DFT approximation, since empirical formulas are used to calculate the exchange-correlation energy, the information about particle–particle correlations is lost. This invalidates subsequent cumbersome path-integral calculations, which are supposed to provide a standard of exactitude. The use of the DFT approximation makes it possible to apply the Gibbs distribution to single-particle states, but the assumption of weak exchange effects reduces

the scope of the method to that of conventional mean-field approximations.

3. NONINTEGRABLE SINGULARITY OF THE COULOMB POTENTIAL

A path integral is defined as a limit of the sequence of finite-dimensional integrals obtained by replacing the virtual path $\tilde{R}(t)$ that joins points R_a and R_b in the coordinate space $R = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ of an N-particle system with a polygonal path $R_a, R(1), \ldots, R(M-1), R_b$. In particular, a matrix element of the evolution operator is written in the coordinate representation as follows [49]:

$$\begin{aligned}\langle R_a|\exp\left(-\frac{i}{\hbar}\hat{H}t\right)|R_b\rangle &= \int_{R_a}^{R_b}\exp\left(\frac{i}{\hbar}S([\tilde{R}(t)])\right)D\tilde{R}(t)\\ &= \lim_{M\to\infty}\left(\frac{Mm_0}{2\pi i t\hbar}\right)^{3NM/2}\int dR(1)\ldots dR(M-1)\\ &\times\exp\left[\frac{i}{\hbar}S(R_a, R(1), \ldots, R(M-1), R_b)\right],\end{aligned} \quad (1)$$

where $S(R_a, R(1), \ldots, R(M-1), R_b)$ is the action functional over a polygonal path and m_0 is the particle mass. An analogous expression for the density matrix associated with the operator $\exp(-\beta\hat{H})$ is given by (1) with time replaced by the inverse temperature, $t \longrightarrow -i\hbar\beta$.

To date, no general proof has been found for the existence and uniqueness of this limit, i.e., of the path integral [50]. In the case of a smooth interaction potential, the existence of the limit does not seem to be questionable. The case of a singular (e.g., Coulomb) potential calls for a special analysis. The existence of a center of attraction in the Coulomb field generated by an atomic nucleus leads to an unphysical singularity in the discrete approximations of the path integrals for the electron gas (with $M < \infty$), and the resulting path integral is divergent. The singularity presents serious problems in numerical simulations, because representative electron paths are "engulfed" by the Coulomb field of the nucleus, collapsing to a point. The introduction of a rigid core eliminates the nonintegrable singularity, but simultaneously removes the contributions of the paths that intersect a neighborhood of the nucleus. This leads to a systematic error that increases with the core radius. Since the core radius must be greater than the segment length in the polygonal path, one has to increase the number of segments, which leads to a higher computational complexity and poorer statistics. Thus, a systematic error translates into a statistical one.

The singularity is obtained by dropping the contribution of the action functional that does not vanish even in the high-temperature limit in the discrete approximation. The unphysical singularity of the functional can be eliminated by using a more accurate discrete approximation that cannot be reduced to addition of high powers of t/M.

The asymptotic expression for short-time matrix elements of the evolution operator (with $\tau = t/M$, $M \longrightarrow \infty$) can be represented as a series in powers of τ [51]:

$$\langle\mathbf{r}|\exp\left(-\frac{i}{\hbar}\tau\hat{H}\right)|\mathbf{r}_0\rangle = \tau^{-3N/2}\exp\left[\frac{i}{\hbar\tau}\sum_{n=0}\tau^n W_n(\mathbf{r})\right]. \quad (2)$$

The first two terms in the series, W_0 and W_1, are the matrix elements of the free-particle evolution operator satisfying the equation

$$i\hbar\frac{\partial\rho_0(\mathbf{r}_0, \mathbf{r}, \tau)}{\partial\tau} = -\frac{\hbar^2}{2m_0}\nabla^2\rho_0(\mathbf{r}_0, \mathbf{r}, \tau), \quad (3)$$

whose solution is [52]

$$\rho_0(\mathbf{r}_0, \mathbf{r}, \tau) = \left(\frac{m_0}{2\pi i\hbar\tau}\right)^{3N/2}\exp\left(i\frac{m_0}{2\hbar\tau}(\mathbf{r}-\mathbf{r}_0)^2\right). \quad (4)$$

The remaining terms W_n are sought by substituting expression (2) into the operator Bloch equation

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + U(\mathbf{r}) - i\hbar\frac{\partial}{\partial\tau}\right] = 0,$$

where $U(\mathbf{r})$ is the particle–particle interaction operator, and equating the coefficients of like powers of τ. A closed equation is obtained for W_2, whereas the higher order terms W_n are determined by an infinite system of coupled equations:

$$(\mathbf{r}-\mathbf{r}_0)\nabla W_n + (n-1)W_n = \begin{cases}-U(\mathbf{r}), & n = 2,\\ \dfrac{i\hbar}{2m_0}\nabla^2 W_{n-1}, & n \geq 3,\end{cases} \quad (5)$$

which yields

$$W_2(\mathbf{r}) = -\int_0^1 d\xi U(\mathbf{r}_0 + (\mathbf{r}-\mathbf{r}_0)\xi), \quad (6)$$

and

$$W_n(\mathbf{r}) = \frac{i\hbar}{2m_0}\int_0^1 d\xi\,\xi^{n-2}\nabla^2 W_{n-1}(\mathbf{r}_0 + (\mathbf{r}-\mathbf{r}_0)\xi) \quad (7)$$

for larger n. By substituting $\varepsilon = i\tau/\hbar$, analogous asymptotic expressions are obtained for the matrix elements of the statistical operator $\exp(-\beta\hat{H})$ in the high-temperature limit (as $\varepsilon = \beta/M$, $M \longrightarrow \infty$). The diagonal matrix elements are obtained by successively solving Eqs. (6) and (7) for $W_n(\mathbf{r})$ in the limit of $\mathbf{r} \longrightarrow \mathbf{r}_0$. In the one-dimensional case,

$$\langle r|\exp(-\varepsilon\hat{H})|r\rangle = \left(\frac{m_0}{2\pi\hbar^2\varepsilon}\right)^{1/2}$$

$$\times \exp\left[-\varepsilon U(r) - \frac{\hbar^2\varepsilon^2}{12m_0}U''(r)\right. \tag{8}$$

$$\left. + \frac{\hbar^2\varepsilon^3}{24m_0}\left(U'^2(r) - \frac{\hbar}{10m_0}U'''(r)\right) + O(\varepsilon^4)\right].$$

When the matrix elements are replaced by the high-temperature asymptotic expressions, the path integrals in the matrix elements converge as $\varepsilon = \beta/M \longrightarrow 0$ if the remainder term is on the order of at least $O(\varepsilon^2)$. It is sufficient to retain only the linear term $\varepsilon U(r)$ in the exponent in (8), as commonly done in both diagonal and off-diagonal matrix elements [26–52]. However, Eqs. (14)–(19) written out below show that integral (6) should be used in off-diagonal elements instead of $\varepsilon U(r)$. The frequent hypothesis that accuracy is improved by retaining higher order terms of the expansion in (8) is true only for nonsingular potentials $U(r)$. Indeed, approximation (8) of an off-diagonal matrix element is supposed to rely on an expansion of $U(r)$ about a point lying on a path, but the corresponding series obtained for singular points are divergent. For example, even the first-order term in the expansion corresponding to the Coulomb potential $U(r) \propto -(r - r_0)^{-1}$ gives rise to an exponential singularity of the matrix element, $\exp(\kappa(r - r_0)^{-1})$, where κ is a dimensional factor, and the degree of singularity increases when the expansion includes higher order terms. Therefore, integral expressions (6) and (7) must be retained even in an asymptotic theory. Integral (6) corresponds to a uniform distribution of an electron over a segment in a polygonal path. Calculated analytically for the Coulomb potential over a line segment, integral (6) yields a logarithmic (i.e., weakly singular) term $W_2(r)$ and similar higher order terms $W_n(r)$. At the origin of the Coulomb potential, the resulting short-time (high-temperature) matrix element has an integrable singularity, $(r - r_0)^{-1}$, and the corresponding discrete approximation of the integral is convergent. Substituting (6) into the high-temperature matrix elements, one obtains a weakly singular functional $S(R_a, R(1), R(2), \ldots, R(M-1), R_b)$. It formally differs from the commonly used expression in that interaction with uniformly charged segments of a polygonal path is substituted instead of Coulomb interaction with point charges.

Even though the nonintegrable singularity is eliminated by representing the interaction term as (6), this representation does not improve numerical efficiency, because the arithmetic complexity of each step in a Monte Carlo procedure used to calculate the Coulomb interaction between a uniformly charged segment of a polygonal path and a nucleus is relatively high. The computation is substantially simplified if this interaction is replaced by the interaction with a segment of length $\delta = \hbar\sqrt{\beta/Mm_0}$ (thermal wavelength) oriented perpendicular to the line that joins a vertex in the polygonal path with the nucleus. Since the charge distributed over the segment reduces to a point with increasing number of segments, the potentials of both interactions follow a similar asymptotic behavior as $M \longrightarrow \infty$. In this limit, the potential of the interaction between an electron path segment and a nucleus having a charge e over a distance r is

$$u(r) = -\frac{e^2}{M\delta}\ln\left(\frac{\sqrt{\delta^2 + 4r^2} + \delta}{\sqrt{\delta^2 + 4r^2} - \delta}\right). \tag{9}$$

Numerical tests using functional (9) were performed in [21, 41–44] for atomic and molecular hydrogen. Excellent agreement with ground-state solutions was obtained even for low-temperature systems, which are most difficult to calculate. Representation (9) was also used in the calculations discussed below.

4. CALCULATION OF INTERNAL ENERGY AND EQUATIONS OF STATE

4.1. Theoretical Basis

The equilibrium average of a quantum-mechanical operator $\hat{A}$ for a system of N distinguishable particles characterized by an inverse temperature $\beta = 1/k_BT$ is

$$\bar{F} = \frac{\mathrm{Tr}[\hat{A}\exp(-\beta\hat{H})]}{\mathrm{Tr}[\exp(-\beta\hat{H})]}. \tag{10}$$

When matrix elements are represented as path integrals [49], the average has the form

$$\bar{A} = \frac{\int\tilde{A}([R_1(t; a, b)])\exp(\tilde{S}([R_2(t; b, a)]))DR_1(t)DR_2(t)}{\int\exp(\tilde{S}([R(t; a, a)]))DR(t)}, \tag{11}$$

where $\tilde{S}([R(t)])$ is the action functional over a path $R(t)$ evolving in imaginary time $t = -i\beta\hbar$ and $\tilde{A}([R(t)])$ is a functional over a path $R(t)$ in the coordinate space $R\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ of an N-particle system. The path $R(t; a, b)$ exits from point a and enters point b. The diagonal matrix elements correspond to closed paths $R(t; a, a)$ in (10). In the general case, the numerator in (11) contains paths $R(t; a, b)$ from a to b that are connected with reverse paths $R(t; b, a)$ joining the same points. In most cases, the functional $\tilde{A}([R(t)])$ can be expressed in such a form that the numerator in (11) contains only the nonzero contributions of closed paths (for which $a = b$),

$$R_1(t; a, b) = R_2(t; b, a) \equiv R(t; a, a).$$

In a discrete-time path integral, $R(t; a, a)$ is a closed polygon with M vertices, and the functional $\tilde{A}([R(t)])$ is a function $A(\{R_i\})$ of the coordinates of the same M vertices in the closed polygon $\{R_i\}$ as those used in the function $S(\{R_i\})$ corresponding to the functional $\tilde{S}([R(t)])$. The function $A(\{R_i\})$ is called the estimator for the observable associated with the operator $\hat{A}$. The same observable may admit the use of a variety of estimators given by different functions and characterized by different variances when calculated by stochastic methods, such as estimators for kinetic energy [53–57].

In computer simulations of macroscopic systems, the space is divided into identical cubic boxes, and periodic boundary conditions are imposed. The locations of particles in the basic box are periodically repeated in other boxes [19]. The current coordinates of the particles confined in the basic box are stored in computer memory. In the nearest-image convention, the interactions of each particle with the remaining particles in the basic box are taken into account. In addition to the nearest-image convention, the interactions with the particles located outside the basic box are modeled by the interactions with all of their images. In this model, all points in space are equivalent, and boundary effects are eliminated. Equivalence of different estimators in a model with periodic boundary conditions warrants a special analysis.

To calculate an equation of state, both pressure and kinetic-energy estimators are required. Consider a system consisting of N_p protons of mass m_p treated as a classical statistical ensemble and N electrons of mass m_e treated as an ensemble of indistinguishable particles. The pressure p for such a system confined in a volume V, with Hamiltonian $\hat{H}$ expressed in terms of the Helmholtz free energy F and partition function $Z = \mathrm{Tr}[\exp(-\beta\hat{H})]$ for a canonical ensemble as follows:

$$p = -\frac{\partial F}{\partial V} = k_B T \frac{\partial \ln Z}{\partial V}. \tag{12}$$

The partition function can be represented as the discretized path integral [45, 49]

$$Z = \lim_{M\to\infty} \frac{1}{N_p!}\left(\frac{h^2}{2\pi m_p k_B T}\right)^{-3N_p/2} \frac{1}{N!}\left(\frac{m_e M}{2\pi\beta\hbar^2}\right)^{3MN/2}$$

$$\times \sum_{\{\nu_i\}} \omega(\{\nu_i\}) \int_V \ldots \int_V \exp\left(-\sum_{k=1}^{N}\sum_{i=1}^{M} \frac{m_e M}{2\beta\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2\right.$$

$$-\frac{\beta}{M}\sum_{i=1}^{M}\left[-\sum_{k=1}^{N} e\varphi(\mathbf{r}_i^k)\right.$$

$$\left.\left. + \sum_{l<k=1}^{N} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|} - \sum_{k=1}^{N}\sum_{l=1}^{N_p} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right]\right. \tag{13}$$

$$\left. - \beta\left[\sum_{k=1}^{N_p} e\varphi(\mathbf{r}^k) + \sum_{l<k=1}^{N_p} \frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right]\right)$$

$$\times d\mathbf{r}^1 \ldots d\mathbf{r}^{N_p} d\mathbf{r}_1^1 \ldots d\mathbf{r}_1^N \ldots d\mathbf{r}_M^1 \ldots d\mathbf{r}_M^N$$

subject to the path connection condition $\mathbf{r}_0^k = \tilde{\mathbf{r}}_M^k$, with $\{\tilde{\mathbf{r}}_M^k\} = \hat{\Pi}_{\{\nu_i\}}\{\mathbf{r}_M^k\}$. The permutation operator $\hat{\Pi}_{\{\nu_i\}}$ is defined on a set of N ordered elements, $\{\mathbf{r}_M^k\} \equiv \mathbf{r}_M^1, \mathbf{r}_M^2, \ldots, \mathbf{r}_M^N$. The multiindex $\{\nu_i\} \equiv \nu_1, \nu_2, \ldots, \nu_N$ specifies a decomposition of N paths into cycles of connected paths, where ν_i denotes the number of cycles consisting of i connected paths; the weight

$$\omega(\{\nu_i\}) = \sum_S (2S+1)\omega_S(\{\nu_i\})$$

of a particular configuration of connected paths is obtained after performing the summation over spin states [45]; e is the elementary charge; $\varphi(\mathbf{r})$ is the potential of the external electric field; M is the number of segments in the polygonal path that represents an electron in the discretized path integral; $\mathbf{r}_i^k$ is the coordinate of the ith vertex in the kth path; and $\mathbf{r}^k$ is the coordinate of the kth proton. The sum in (13) is taken over all sets $\{\nu_i\}$ such that $\sum_{i=1}^{N} i\nu_i = N$, i.e., over all configurations of connected paths. Expression (13) does not involve integration over the proton momentum subspace, because it is independent of the system's volume and, therefore, does not contribute to pressure.

To avoid differentiation with respect to integration limits in expression (13) substituted into (12), the following change of variables is performed:

$$\mathbf{r}_i^k \longrightarrow \gamma \mathbf{r}_i^k, \quad \mathbf{r}^k \longrightarrow \gamma \mathbf{r}^k \quad (0 < \gamma \leq 1).$$

Then,

$$V \longrightarrow \gamma^3 V, \quad \frac{\partial}{\partial V} = \frac{\gamma}{3V}\frac{\partial}{\partial \gamma}.$$

According to (12), the pressure is

$$p = \frac{1}{\beta Z}\frac{\partial Z}{\partial V} = \frac{1}{3V\beta Z}\frac{\partial Z}{\partial \gamma}\bigg|_{\gamma=1}$$

$$= \frac{N_p k_B T}{V} + \frac{MNk_B T}{V} + \frac{1}{3VZ}$$

$$\times \left\{ \lim_{M\to\infty} \frac{1}{N_p!}\left(\frac{h^2}{2\pi m_p k_B T}\right)^{-3N_p/2} \frac{1}{N!}\left(\frac{m_e M}{2\pi\beta\hbar^2}\right)^{3MN/2} \right.$$

$$\times \sum_{\{\nu_i\}} \omega(\{\nu_i\}) \int_V \ldots \int_V \left[-\sum_{k=1}^{N}\sum_{i=1}^{M} \frac{m_e M}{\beta^2\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2 \right.$$

$$-\frac{1}{M}\sum_{i=1}^{M}\left(-\sum_{k=1}^{N} e\mathbf{r}_i^k \nabla\varphi(\mathbf{r}_i^k) \right.$$

$$\left. -\sum_{l<k=1}^{N} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|} + \sum_{k=1}^{N}\sum_{l=1}^{N_p} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right)$$

$$\left. -\sum_{k=1}^{N_p} e\mathbf{r}^k \nabla\varphi(\mathbf{r}^k) \right.$$

$$\left. + \sum_{l<k=1}^{N_p} \frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right] \exp\left[-\sum_{k=1}^{N}\sum_{i=1}^{M} \frac{m_e M}{2\beta\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2 \right.$$

$$-\frac{\beta}{M}\sum_{i=1}^{M}\left(-\sum_{k=1}^{N} e\varphi(\mathbf{r}_i^k) + \sum_{l<k=1}^{N} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|}\right. \qquad (14)$$

$$\left. -\sum_{k=1}^{N}\sum_{l=1}^{N_p} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right)$$

$$\left. -\beta\left(\sum_{k=1}^{N_p} e\varphi(\mathbf{r}^k) + \sum_{l<k=1}^{N_p} \frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right)\right]$$

$$\left. \times d\mathbf{r}^1 \ldots d\mathbf{r}^{N_p} d\mathbf{r}_1^1 \ldots d\mathbf{r}_M^N \right\}$$

$$= \lim_{M\to\infty}\left\{ \frac{N_p k_B T}{V} + \frac{MNk_B T}{V} + \frac{1}{3V} \right.$$

$$\times \left\langle -\sum_{k=1}^{N}\sum_{i=1}^{M} \frac{m_e M}{\beta^2\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2 \right.$$

$$-\frac{1}{M}\sum_{i=1}^{M}\left[-\sum_{k=1}^{N} e\mathbf{r}_i^k \nabla\varphi(\mathbf{r}_i^k) - \sum_{l<k=1}^{N} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|} \right.$$

$$\left. + \sum_{k=1}^{N}\sum_{l=1}^{N_p} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right]$$

$$\left. \left. - \sum_{k=1}^{N_p} e\mathbf{r}^k \nabla\varphi(\mathbf{r}^k) + \sum_{l<k=1}^{N_p} \frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right\rangle \right\},$$

where $\langle \ldots \rangle$ denotes averaging in the space of proton coordinates $\{\mathbf{r}^l\}$, electron paths $\{\mathbf{r}_i^k\}$, and their different connections $\{\nu_i\}$ over the distribution function

$$\rho(\{r_i^k\}, \{r^l\}, \{\nu_i\}; \beta)$$

$$\propto \exp\left[-\sum_{k=1}^{N}\sum_{i=1}^{M} \frac{m_e M}{2\beta\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2 \right.$$

$$-\frac{\beta}{M}\sum_{i=1}^{M}\left(\sum_{k=1}^{N} e\varphi(\mathbf{r}_i^k) + \sum_{l<k=1}^{N} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|}\right. \qquad (15)$$

$$\left. -\sum_{k=1}^{N}\sum_{l=1}^{N_p} \frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right)$$

$$\left. -\beta\left(\sum_{k=1}^{N_p} e\varphi(\mathbf{r}^k) + \sum_{l<k=1}^{N_p} \frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right)\right]$$

under the path connection condition used in (13).

In the absence of external fields, the pressure in a Coulombic quantum system is directly expressed in terms of the mean kinetic and potential energies, $\langle K \rangle$

and $\langle U\rangle$. Indeed, the general estimator for the kinetic energy of a quantum system has the form [53–56]

$$\langle K\rangle = \frac{3}{2}MNk_{\mathrm{B}}T - \left\langle \sum_{k=1}^{N}\sum_{i=1}^{M}\frac{m_e M}{2\beta^2\hbar^2}(\mathbf{r}_i^k - \mathbf{r}_{i-1}^k)^2 \right\rangle, \quad (16)$$

and the estimator for the total potential energy in the case of $\varphi(\mathbf{r}) = 0$ is

$$\langle U\rangle = \left\langle \frac{1}{M}\sum_{i=1}^{M}\left(\sum_{l<k=1}^{N}\frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}_i^l|} - \sum_{k=1}^{N}\sum_{l=1}^{N_p}\frac{e^2}{|\mathbf{r}_i^k - \mathbf{r}^l|}\right) + \sum_{l<k=1}^{N_p}\frac{e^2}{|\mathbf{r}^k - \mathbf{r}^l|}\right\rangle. \quad (17)$$

A comparison of (16) and (17) with (14) for $\varphi(\mathbf{r}) = 0$ leads to the following result:

$$p = \frac{N_p k_{\mathrm{B}}T}{V} + \frac{2}{3V}\left(\langle K\rangle + \frac{1}{2}\langle U\rangle\right). \quad (18)$$

Expression (18) is valid for a Coulombic system of quantum and classical particles of volume V in the absence of external fields. If the system contains a single immobile classical particle, then the first term in (18) vanishes. When an unbounded system is in a quantum state whose stability is ensured only by interactions between particles, variation of volume in (14) does not change the total kinetic energy and $p = 0$. In such a state,

$$\langle K\rangle = -\frac{1}{2}\langle U\rangle. \quad (19)$$

In particular, (19) holds for discrete energy states of the electron shells of atoms in free space. Note that this equality is not valid when the system contains more than one immobile charged particle, in which case the derivative with respect to γ in (14) is ill-defined, and expression (14) for pressure cannot be rewritten as (18). In the general case of arbitrary potentials $\varepsilon_{lk}(r)$, $E_{lk}^{\mathrm{cl}}(r)$, $\xi_{lk}(r)$, $\Phi_k(\mathbf{r})$, and $\Phi_n^{\mathrm{cl}}(\mathbf{r})$ of interaction between the lth and kth quantum particles, the lth and kth classical particles, the lth classical and kth quantum particles, external field and the kth quantum particle, and external field and the nth classical particle, respectively, a similar analysis leads to a more general analog of expression (18):

$$p = \frac{N_p k_{\mathrm{B}}T}{V} + \frac{2}{3V}(\langle K\rangle - \langle K_{\mathrm{vir}}\rangle), \quad (20)$$

with

$$K_{\mathrm{vir}} = \frac{1}{2M}\sum_{i=1}^{M}\left[\sum_{k=1}^{N}\mathbf{r}_i^k\nabla\Phi_k(\mathbf{r}_i^k) + \sum_{l<k}^{N}|\mathbf{r}_i^l - \mathbf{r}_i^k|\varepsilon'_{lk}(|\mathbf{r}_i^l - \mathbf{r}_i^k|) + \sum_{k=1}^{N}\sum_{l=1}^{N_p}|\mathbf{r}_i^k - \mathbf{r}^l|\xi'_{lk}(|\mathbf{r}_i^k - \mathbf{r}^l|)\right] + \frac{1}{2}\left[\sum_{k=1}^{N_p}\mathbf{r}^k\nabla\Phi_k^{\mathrm{cl}}(\mathbf{r}^k) + \sum_{l<k}^{N_p}|\mathbf{r}^l - \mathbf{r}^k|E_{lk}^{\mathrm{cl}'}(|\mathbf{r}^l - \mathbf{r}^k|)\right], \quad (21)$$

where primed quantities denote derivatives with respect to coordinates. For an unbounded system that does not contain classical particles, the following analog of (19) is valid:

$$\langle K\rangle = \langle\tilde{K}_{\mathrm{vir}}\rangle,$$

where

$$\tilde{K}_{\mathrm{vir}} = \frac{1}{2M}\sum_{i=1}^{M}\left[\sum_{k=1}^{N}\mathbf{r}_i^k\nabla\Phi_k(\mathbf{r}_i^k) + \sum_{l<k}^{N}|\mathbf{r}_i^i - \mathbf{r}_i^k|\varepsilon'_{lk}(|\mathbf{r}_i^l - \mathbf{r}_i^k|)\right] \quad (22)$$

is a well-known particular case of the virial estimator for the kinetic energy of a quantum system [53–56], which is commonly used in numerical analyses. Thus, calculations of kinetic energy based on (22) are valid only in the absence of boundaries. In the general case, including systems with periodic boundary conditions, expressions (20) and (21) should be employed.

In stochastic simulation methods, virial estimator (21) should be used instead of general estimator (16), because the amplitude of fluctuations of expression (21) decreases roughly as $1/\sqrt{M}$, whereas the amplitude of fluctuations of (16) scales with $\sqrt{M}$. When M is sufficiently large, the statistical error in the average value of the estimator (16) calculated by using finite samples may be on the order of the average value itself; i.e., the general estimator K is useless. Similar problems concerning the application of the general estimator arise when pressure is calculated by using expression (20), because this leads to realistic results only in the least interesting case of extremely high temperature. However, it follows from (20) that the virial estimator can be employed instead of the general one, as is frequently

done in numerical analyses based on path integral methods [53–56], only for zero-pressure systems. The condition of zero pressure can be applied to a finite-temperature system only when its stability in the absence of boundaries is guaranteed by virtue of indefinite increase in potential energy with interparticle distance. In such systems, continuum states do not exist. Conversely, all realistic particle–particle interaction potentials tend to vanish at large distances together with their derivatives. This explains why such systems are stable at finite temperatures only when they are bounded by walls. For such systems, pressure is always finite and straightforward substitution of the virial estimator for the general one is incorrect. This is true for systems with periodic boundary conditions. These problems do not arise in analyses of systems described by classical statistical mechanics, because the classical counterpart of (20) contains the constant $(3/2)Nk_BT$ instead of $\langle K\rangle$ and the calculation of the pressure variance does not yield a divergent result.

The divergence of the general estimator for kinetic energy presents serious difficulties in applications of path integral methods. The energy could be calculated if one found an independent method for calculating the pressure and substitute the result into (20) to obtain an equation for $\langle K\rangle$. An independent pressure calculation can be performed by making use of interactions between particles and walls. Suppose that these interactions can be described in terms of functions $\Phi_k(\mathbf{r})$ and $\Phi_k^{\mathrm{cl}}(\mathbf{r})$ representing the potential barrier that holds the system in a volume V. Then, any rigid walls set outside this volume do not affect the state of the system, which entails zero pressure on the rigid wall on the left hand side in (14). Therefore, expression (21) can be used to rewrite (20) as

$$\frac{1}{3V}\left(\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{N}\mathbf{r}_i^k\nabla\Phi_k(\mathbf{r}_i^k)+\sum_{k=1}^{N_p}\mathbf{r}^k\nabla\Phi_k^{\mathrm{cl}}(\mathbf{r}^k)\right) = \frac{N_pk_BT}{V}+\frac{2}{3V}(\langle K\rangle-\langle K_{\mathrm{vir}}^0\rangle), \tag{23}$$

where

$$K_{\mathrm{vir}}^0 \equiv \frac{1}{2M}\sum_{i=1}^{M}\left[\sum_{l<k}^{N}\left|\mathbf{r}_i^l-\mathbf{r}_i^k\right|\varepsilon'_{lk}\left(\left|\mathbf{r}_i^l-\mathbf{r}_i^k\right|\right) + \sum_{k=1}^{N}\sum_{l=1}^{N_p}\left|\mathbf{r}_i^k-\mathbf{r}^l\right|\xi'_{lk}\left(\left|\mathbf{r}_i^k-\mathbf{r}^l\right|\right)\right] + \frac{1}{2}\sum_{l<k}^{N_p}\left|\mathbf{r}^l-\mathbf{r}^k\right|E_{lk}^{\mathrm{cl}'}\left(\left|\mathbf{r}^l-\mathbf{r}^k\right|\right).$$

According to (20), in a similar system, but without potential barriers $\Phi_k(\mathbf{r})$ and $\Phi_k^{\mathrm{cl}}(\mathbf{r})$, the pressure on rigid walls is

$$p = \frac{N_pk_BT}{V}+\frac{2}{3V}(\langle K\rangle-\langle K_{\mathrm{vir}}^0\rangle). \tag{24}$$

Comparing (23) to (24), one obtains

$$p = \frac{1}{3V}\left(\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{N}\mathbf{r}_i^k\nabla\Phi_k(\mathbf{r}_i^k)+\sum_{k=1}^{N_p}\mathbf{r}^k\nabla\Phi_k^{\mathrm{cl}}(\mathbf{r}^k)\right). \tag{25}$$

Using estimators for the spatial distributions of quantum and classical particles [46],

$$\rho(\mathbf{r}) = \left\langle\frac{1}{M}\sum_{k=1}^{N}\sum_{i=1}^{M}\delta(\mathbf{r}_i^k-\mathbf{r})\right\rangle$$

and

$$\rho_p(\mathbf{r}) = \left\langle\sum_{k=1}^{N_p}\delta(\mathbf{r}^k-\mathbf{r})\right\rangle,$$

one can rewrite expression (25) for a hydrogen plasma as

$$p = \frac{1}{3V}\int_V\mathbf{r}\cdot(\nabla\Phi(\mathbf{r})\rho(\mathbf{r})+\nabla\Phi_p(\mathbf{r})\rho_p(\mathbf{r}))dV, \tag{26}$$

where $\nabla\Phi(\mathbf{r})$ and $\nabla\Phi_p(\mathbf{r})$ are the forces exerted by a potential barrier on electrons and protons, respectively. Even though the dominant contribution to integral (26) is due to a narrow near-wall layer, transformation of volume integral (26) into a surface one is reasonable only for a classical system, because it would involve the introduction of a function $\Phi(\mathbf{r})$ that is discontinuous across the boundary. For a quantum system, the discontinuous $\Phi(\mathbf{r})$ must be introduced after taking the limit $M\longrightarrow\infty$, which makes it impossible to calculate the pressure numerically by this method. Note also that the density $\rho(\mathbf{r})$ of a quantum system on the wall vanishes as $M\longrightarrow\infty$ by virtue of the Schrödinger equation.

Consider a macroscopic system confined in a $d\times d\times d$ cubic vessel, with a face area $S=d^2$. A potential barrier set at a distance b from the walls prevents contact of particles with the walls. Since there exists an intermediate layer of thickness σ, the system is bounded by a soft barrier rather than a rigid wall; i.e., an undetermined δV is added to the volume $V=S(d-b)=Sa$. Pressure can be calculated by using (26) in the limit of $\delta V/V\longrightarrow 0$, i.e., $\sigma/a\longrightarrow 0$. A system is macroscopic if its dimensions are much greater than the correlation radius of fluctuations. Expression (26) means that the interactions between particles and walls are completely

described by $\Phi(\mathbf{r})$ and $\Phi_p(\mathbf{r})$. However, no limitation is imposed on the specific form of these functions. The most efficient numerical algorithm corresponds to $\Phi(\mathbf{r})$such that $\mathbf{r}\cdot\nabla\Phi(\mathbf{r})$ = const in the barrier region:

$$\Phi(\mathbf{r}) = \Phi(x) + \Phi(y) + \Phi(z),$$

where

$$\Phi(x) = \begin{cases} \dfrac{k_B T a}{\sigma}\ln\left(-\dfrac{x}{a}\right), & x < -a, \\ 0, & -a \le x \le a, \\ \dfrac{k_B T a}{\sigma}\ln\left(\dfrac{x}{a}\right), & x > a. \end{cases} \quad (27)$$

For a classical system, $\Phi_p(\mathbf{r})$ has a similar form. The ratio $k_B T/\sigma$ in (27) is interpreted as the force exerted on a particle by a barrier. In the first approximation, the probability density of particle penetration to a depth $x - a$ into the layer can be estimated as

$$\exp\left(-\frac{\Phi}{k_B T}\right) = \left(\frac{a}{x}\right)^{a/\sigma}.$$

According to this expression, the particle density in the layer is a decreasing function of the penetration depth that can be approximated by a power law. The characteristic distance at which the density is lower by a factor of e is

$$x - a = a(e^{\sigma/a} - 1) \approx \sigma.$$

The value of b is assumed to be sufficiently large to ensure that characteristic radius of interaction with the barrier is smaller than the distance to the wall, $\sigma \ll b$, which guarantees negligible particle density at a rigid wall and applicability of (25) and (26).

When a potential barrier is described by (27), the integrand in (26) does not vanish only in a near-wall layer of thickness on the order of σ. Since the system is macroscopic, the particle density in the near-wall layer depends only on the distance to the wall, the variables in the integrand separate, and triple integral (26) can be represented as the sum of similar one-dimensional integrals of the form

$$I_x = I_y = I_z$$

$$= \frac{S}{3V}\lim_{\sigma/a\to 0}\int_{-(a+b)}^{\alpha+b} x\left[\frac{\partial\Phi}{\partial x}\rho(x) + \frac{\partial\Phi_p}{\partial x}\rho_p(x)\right]dx \quad (28)$$

$$= \frac{1}{3a}\lim_{\sigma/a\to 0}\int_{a}^{\alpha+b} x\left[\frac{\partial\Phi}{\partial x}\rho(x) + \frac{\partial\Phi_p}{\partial x}\rho_p(x)\right]dx.$$

Since the second-right equality in (28) is valid for even functions $\Phi(x)$, the product $x\partial\Phi/\partial x$ is an even function. Substituting (27) into (28) yields

$$p = 3I_x = k_B T\int_a^{\alpha+b}\left(\frac{\rho(x)}{\sigma} + \frac{\rho_p(x)}{\sigma_p}\right)dx = \frac{k_B T}{S}\left(\frac{\langle\Delta N\rangle}{\sigma} + \frac{\langle\Delta N_p\rangle}{\sigma_p}\right), \quad (29)$$

where $\langle\Delta N\rangle$ and $\langle\Delta N_p\rangle$ are the mean numbers of quantum particles within $[a, a+b]$ (under the barrier). As $\sigma/a \longrightarrow 0$, only the contribution of leading order in $x'/a = (x-a)/a$ to the integrand in (28) does not vanish, and $\Phi(x)$ can be replaced by its asymptotic form in the limit of $x'/a \longrightarrow 0$:

$$\tilde{\Phi}(x') = \begin{cases} -\dfrac{k_B T x'}{\sigma}, & x < -a, \\ 0, & -a \le x \le a, \\ \dfrac{k_B T x'}{\sigma}, & x > a. \end{cases} \quad (30)$$

Thus, the calculation of pressure reduces to the straightforward averaging over distribution function (15) of particles under the barrier defined by (30). This can be done numerically by a Monte Carlo method.

Simulations of macroscopic systems make use of periodic boundary conditions, which correspond to the absence of boundaries. On the other hand, the calculation of pressure as given by (26) requires the presence of a surface. The problem is solved by imposing boundary conditions only in the yz plane in the absence of boundary along the x axis. The pressure given by (29) is calculated by a Monte Carlo method within a periodic box. Since only the particles contained in the barrier region contribute to (29), the calculation may be performed by simulating only the surface layer. The numerical results presented here were obtained by simulating the surface-layer fragment spanning the interval of $x \in [c, a+b]$, while periodic boundary conditions were used in the yz plane. The rigid plane wall at $x = c$ guarantees the required particle density within the interval $[c, a]$. By assuming that this layer is much greater than the correlation radius of fluctuations, it is ensured that the rigid wall does not affect the statistical characteristics of particles in the barrier region.

The layer $[c, a]$, where $\nabla\Phi(\mathbf{r}) = 0$, serves to eliminate the effect of the plane boundary set at $x = c$ on the barrier region $[a, a+b]$. The decay of this effect in the layer is controlled by calculating the appropriate correlation functions. In the present computations, $b = a - c$. The characteristic radius $\sigma = b/15$ guarantees that particle density at the rigid wall under the barrier is lower by a factor of $e^{15} \approx 3\times 10^6$, while the barrier region contains 7% of the total number of particles.

Expression (29) is only outwardly similar to the expression for partial pressure in a perfect gas, because the quantities $\langle \Delta N \rangle$ and $\langle \Delta N_p \rangle$ are not mean particle numbers and are not mutually independent.

Combining (20) with (26), one obtains an expression for the kinetic energy of the electron subsystem at a finite pressure:

$$\begin{aligned}\langle K \rangle \equiv \langle K^p_{\rm vir} \rangle &= \frac{3}{2}(pV - N_p k_{\rm B} T) + \langle K_{\rm vir} \rangle \\ &= \frac{1}{2}\int_V \mathbf{r} \cdot (\nabla \Phi(\mathbf{r})\rho(\mathbf{r}) + \nabla \Phi_p(\mathbf{r})\rho_p(\mathbf{r}))dV \qquad (31) \\ &\quad - \frac{3}{2}N_p k_{\rm B} T + \langle K_{\rm vir} \rangle.\end{aligned}$$

It follows from (31) that the virial estimator for kinetic energy at a finite pressure in a system of N quantum particles and N_p classical particles is

$$\begin{aligned}K^p_{\rm vir} &= \frac{1}{2M}\sum_{i=1}^{M}\Bigg(\sum_{l<k}\left|\mathbf{r}_i^l - \mathbf{r}_i^k\right|\varepsilon'_{lk}\left(\left|\mathbf{r}_i^l - \mathbf{r}_i^k\right|\right) \\ &\quad + \sum_{k=1}^{N}\mathbf{r}_i^k \cdot \nabla \Phi_k(\mathbf{r}_i^k) \\ &\quad + \sum_{l=1}^{N}\sum_{k=1}^{N_p}\left|\mathbf{r}_i^l - \mathbf{r}^k\right|\xi'_{lk}\left(\left|\mathbf{r}_i^l - \mathbf{r}^k\right|\right) + \sum_{k=1}^{N}\mathbf{r}_i^k \cdot \nabla \Phi_k(\mathbf{r}_i^k)\Bigg) \qquad (32) \\ &\quad + \frac{1}{2}\Bigg(\sum_{l<k}\left|\mathbf{r}^l - \mathbf{r}^k\right|E'_{lk}\left(\left|\mathbf{r}^l - \mathbf{r}^k\right|\right) + \sum_{k=1}^{N_p}\mathbf{r}^k \cdot \nabla \Phi_k^{\rm cl}(\mathbf{r}^k) \\ &\quad + \sum_{k=1}^{N_p}\mathbf{r}^k \cdot \nabla \Phi_p(\mathbf{r}^k)\Bigg) - \frac{3}{2}N_p k_{\rm B} T,\end{aligned}$$

where $\Phi(\mathbf{r})$ and $\Phi_p(\mathbf{r})$ describe the interactions between quantum and classical particles with a wall, respectively. For a Coulombic system of quantum and classical particles in the absence of external fields other than the near-wall barrier, expression (31) reduces to

$$\begin{aligned}K^p_{\rm vir} &= \frac{1}{2M}\sum_{i=1}^{M}\sum_{k=1}^{N}\mathbf{r}_i^k \cdot \nabla \Phi(\mathbf{r}_i^k) \\ &\quad + \frac{1}{2}\sum_{k=1}^{N_p}\mathbf{r}^k \cdot \nabla \Phi_p(\mathbf{r}^k) - \frac{1}{2}U - \frac{3}{2}N_p k_{\rm B} T,\end{aligned} \qquad (33)$$

where U is the estimator for potential energy defined by (17). In the case of a system confined between two planes separated by the distance $2b$ with barrier function (27), one obtains

$$\begin{aligned}K^p_{\rm vir} &= \frac{3}{2}k_{\rm B} T b\left(\frac{\Delta N}{\sigma} + \frac{\Delta N_p}{\sigma_p}\right) \\ &\quad + \frac{1}{2M}\sum_{i=1}^{M}\Bigg(\sum_{l<k}\left|\mathbf{r}_i^l - \mathbf{r}_i^k\right|\varepsilon'_{lk}\left(\left|\mathbf{r}_i^l - \mathbf{r}_i^k\right|\right) \\ &\quad + \sum_{l=1}^{N}\sum_{k=1}^{N_p}\left|\mathbf{r}_i^l - \mathbf{r}^k\right|\xi'_{lk}\left(\left|\mathbf{r}_i^l - \mathbf{r}^k\right|\right)\Bigg) \\ &\quad + \frac{1}{2}\left(\sum_{l<k}\left|\mathbf{r}^l - \mathbf{r}^k\right|E'_{lk}\left(\left|\mathbf{r}^l - \mathbf{r}^k\right|\right)\right) - \frac{3}{2}N_p k_{\rm B} T.\end{aligned} \qquad (34)$$

For a Coulombic system, expression (34) has the form

$$\begin{aligned}\langle K^{\rm vir}_p \rangle &= \frac{3}{2}k_{\rm B} T b\left(\frac{\langle \Delta N \rangle}{\sigma} + \frac{\langle \Delta N_p \rangle}{\sigma_p}\right) \\ &\quad - \frac{1}{2}\langle U \rangle - \frac{3}{2}N_p k_{\rm B} T.\end{aligned} \qquad (35)$$

Each of the expressions (31)–(35) consists of a surface term and a volume one, the former containing either the functions $\Phi(\mathbf{r})$ and $\Phi_p(\mathbf{r})$ or the quantities ΔN and ΔN_p. The presence of a surface does not modify the volume terms only in the macroscopic limit. For a system with periodic boundary conditions, the surface and volume terms cannot be calculated simultaneously. For this reason, they are calculated separately: a system with periodic boundary conditions is simulated, the quantity $\langle U \rangle/2$ is evaluated, then a system with periodic boundary conditions is simulated only in the yz plane, and the surface contributions per proton are evaluated for a macroscopic system characterized by the same overall density ρ:

$$\begin{aligned}\frac{\langle K^p_{\rm vir} \rangle - \langle K_{\rm vir} \rangle}{N_p} &= \frac{3}{2}\left(\frac{p}{\rho} - k_{\rm B} T\right) \\ &= \frac{3}{2}k_{\rm B} T\left[\frac{1}{S\rho}\left(\frac{\langle \Delta N \rangle}{\sigma} + \frac{\langle \Delta N_p \rangle}{\sigma_p}\right) - 1\right].\end{aligned} \qquad (36)$$

To calculate the kinetic energy of a quantum system per particle, the right-hand side of (36) is added to the ratio $\langle K_{\rm vir} \rangle/N_p$ calculated for the system in the volume. When the surface contributions are evaluated, the volume density of particles is not known a priori and is calculated by averaging over the interval $[c + \delta, c + a - \delta]$. Variation of δ is performed to ensure the absence of density perturbations due to the proximity of a boundary. To obtain the required density and ensure neutrality

in the near-barrier region, both dimensions of the periodic box and parameters σ and σ_p are varied within an initial stage of a Markov process. This stage is excluded from the averaging procedure. A number of special feedback loops are introduced into the computer algorithm to ensure automatic calculation of the prescribed density. Pressure is calculated by simulating only a fragment of the surface layer. To calculate energy, the system in the volume must also be simulated.

4.2. Numerical Results

Monte Carlo simulations of dense hydrogen plasmas were performed on a Pentium IV 3.2 GHz PC. A CPU time of about 20 hours was required to compute a thermodynamic state. Periodic boundary conditions were employed. A periodic box contained 1000 protons and 1000 electrons. The electrostatic interactions between particles were computed under the nearest-image convention [19]. Protons and electrons were treated as classical and quantum particles, respectively. The path of each electron was represented by a closed polygonal one consisting of $M = 20$ segments. In some computations performed for low-temperature systems containing one or two electrons, the number of vertices was $M = 320$. The electrostatic interactions between electrons and protons were calculated by applying the method of distributed quantum-particle charge, which makes it possible to reduce the number of path segments by about an order of magnitude. The effects due to the indistinguishability of electrons (exchange) were modeled explicitly by using a spin variable. The combinatorial weight factors of diagrams representing connected paths were computed by random sampling performed simultaneously with Markov steps. Quantum-mechanical observables were calculated by averaging the corresponding estimators over a Gibbs distribution of particle configurations and paths. Configuration sequences were generated numerically by executing Markov processes. Each macroscopic step in the process included an attempt to translate and rotate a randomly picked path of an electron, and M attempts to displace a vertex in the path and to translate a randomly picked proton in space. Such macroscopic steps were executed alternately with attempts to connect and disconnect Feynman paths in accordance with the combinatorial weights of the emerging diagrams and Gibbs weights of microscopic states. The temperature of the system was gradually decreased by using a procedure in which the initial configuration for computing a particular thermodynamic state was the configuration obtained by computing the preceding state. The initial stage of a Markov process (thermalization) was excluded from the calculation of equilibrium averages. When a series of thermodynamic states was computed, the computation of the thermalization stage for the first state included a vertex multiplication procedure, which increased the rate of relaxation to thermodynamic equilibrium, particularly in computations involving a large number of vertices. In the simulations of subsequent thermodynamic states, the length of the thermalization stage was reduced by a factor of 5. The length of the Markov process executed to compute a thermodynamic state was 10^6 macroscopic steps, which corresponds to a random sample consisting of 1.2×10^7 configurations. The first 200000 steps were excluded from the averaging procedure. In simulations performed for systems with one and two electrons, 2×10^7 macroscopic steps involving about 2×10^9 configurations were executed. Statistical errors were estimated by analyzing fluctuations of partial averages. This was done by dividing the Markov process into ten equal segments and calculating the corresponding mean values of estimators.

The magnitudes of the largest translations and rotations in space executed at the thermalization stage were automatically corrected by the computer program, depending on particular simulation parameters. For example, the magnitude of the largest translation of a path in space at $T = 18\,200$ K and $\rho = 0.370 \times 10^{20}$ cm^{-3} was $\Delta r^e_{\max} = 0.1$Å for an electron and $\Delta r^p_{\max} = 0.08$Å for a proton, the largest angle of path rotation was $\Delta\varphi_{\max} = 62°$, and the largest translation of a vertex in a path was $\Delta r^v_{\max} = 0.33$Å. The corresponding probabilities of adopting a new configuration were 60.3, 63.9, 61.9, and 59.4%, respectively. In the computation of this thermodynamic state, 4.71×10^5 attempts were made to connect and disconnect Feynman paths, and new configurations were adopted in some 270 instances. At each macroscopic step, proximate pairs of vertices were sought among 20 randomly picked candidates. A connection or disconnection was attempted if the distance between two vertices was smaller than 3.8 Å. The procedure of uniform path extension described in [46] was applied simultaneously with a connection or disconnection. The statistical weight of a configuration, including the contributions of diagrams with negative combinatorial weights, was 0.94 times the weight evaluated by neglecting exchange. The parameters specified above strongly depend on simulation conditions. For example, $\Delta r^e_{\max} = 75$ Å, $\Delta r^p_{\max} = 150$ Å, and $\Delta r^v_{\max} = 0.30$ Å were found to be optimal at a temperature of 30700 K and the same pressure. The weights of unbound electronic states rapidly increase with temperature, and so do the probabilities of larger translations of particles in space.

Figure 1a shows the results obtained by computing the states of hydrogen plasmas on the isochore corresponding to $\rho = 0.156 \times 10^{20}$ cm^{-3}. With increasing temperature, the compressibility factor approaches unity, which corresponds to a perfect gas. At temperatures below 20000 K, the curve abruptly turns downwards. This is explained by an increase in the statistical weights of bounded electronic states. In the low-temperature limit, electron–proton pairs combine into neu-

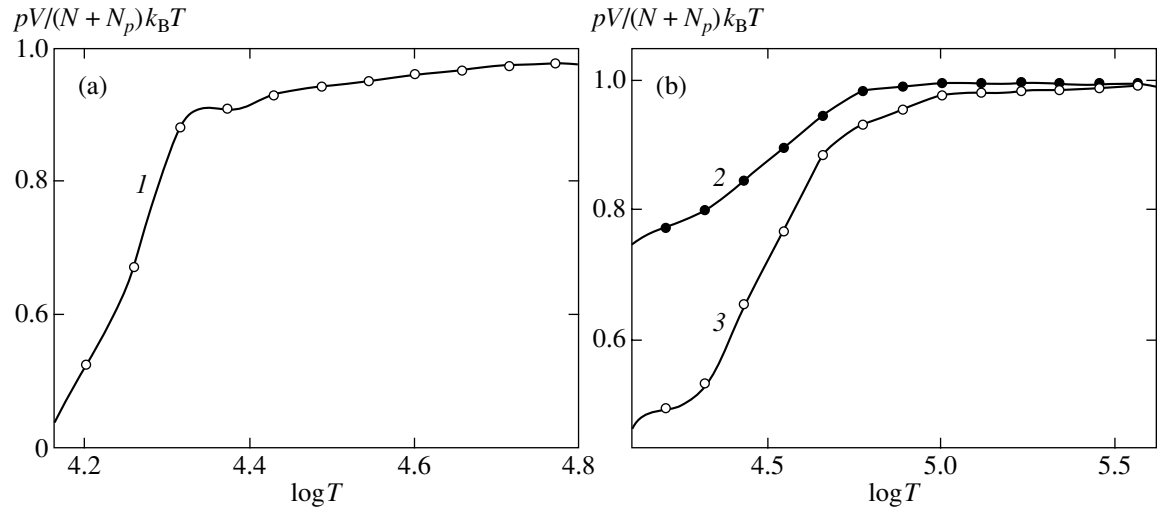

Fig. 1. Compressibility factor vs. temperature for hydrogen plasmas on the isochores corresponding to overall densities of (*1*) 0.156×10^{20}, (*2*) 0.370×10^{20}, and (*3*) 1.25×10^{20} cm^{-3}.

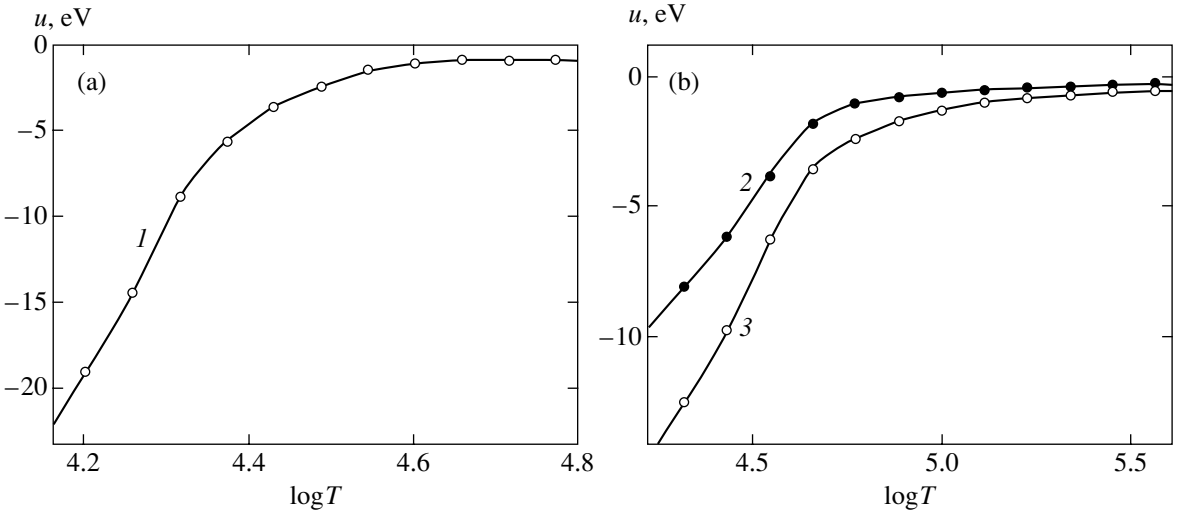

Fig. 2. Mean electron kinetic energy vs. temperature for hydrogen plasmas on the isochores corresponding to an overall density of (*1*) 0.156×10^{20}, (*2*) 0.370×10^{20}, and (*3*) 1.25×10^{20} cm^{-3}.

tral hydrogen atoms, and the number of free particles drops. A simultaneous increase in proton–electron correlations shifts the equilibrium toward states dominated by attraction, additionally reducing the pressure. Analogous trends are predicted for denser plasmas with $\rho = 0.370 \times 10^{20}$ cm^{-3} and 1.25×10^{20} cm^{-3} (see Fig. 1b). Abrupt deviations from the high-temperature behavior are observed at temperatures of 45000 and 90000 K, respectively. The key role played by the formation of bound proton–electron states in the pressure drop at these temperatures is confirmed by an analysis of the behavior of the mean electron potential energy (see Fig. 2), even though the corresponding kink (temperature boundary) in the curve is less pronounced. The sharp decrease in potential energy is due to stronger interactions between protons and electrons, which reflect an increase in correlations between them. The qualitative change in the behavior of the system observed at these points should be attributed to the formation of the electron shells of atoms, and the states located far from these points should be interpreted as totally ionized. This conclusion is confirmed by an analysis of proton–electron correlation functions.

Figure 3 shows the probability density function of location of an electron at a radial distance R from a proton,

$$D(R) = 4\pi R^2 \rho(R),$$

where $\rho(R)$ is the local electron density. Curve *1* corresponds to a temperature above the kink (total ionization); curve *2*, to the temperature boundary; curve *3*, to

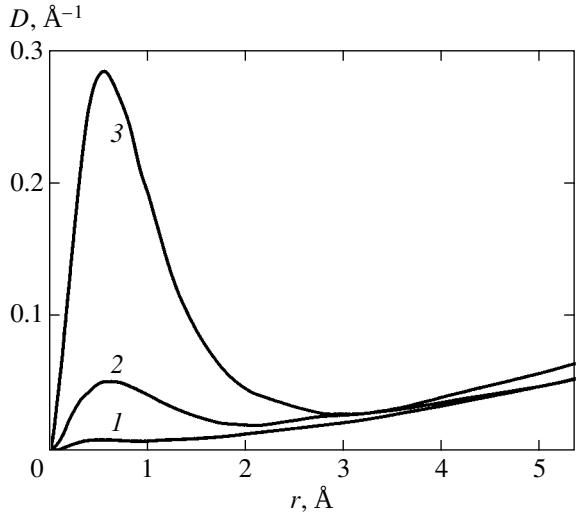

Fig. 3. Radial probability density function for an electron near a proton in a hydrogen plasma of density 0.156×10^{20} cm^{-3} at a temperature of (*1*) 100000, (*2*) 45500, and (*3*) 26900 K.

a temperature below the temperature boundary. The maximum in $D(R)$ located in the vicinity of the Bohr radius ($a_B \approx 0.5$ Å) reflects the formation of the electron shells of atoms. The maximum point corresponds to the atomic ground state. The height of the maximum of curve *3* in Fig. 3 is almost five times smaller than the probability corresponding to the ground state of the hydrogen atom. The contributions of the nearest excited states have almost no effect on the location of the maximum. This is an expected result, because the energy gap between the ground and lowest excited quantum states of the hydrogen atom is about 10 eV, whereas the temperature for which curve *3* in Fig. 3 was calculated is only slightly higher than 2 eV. Under these conditions, the statistical weight of the lowest excited state is two orders of magnitude lower than that of the ground state, and the higher excited states transform into collective states of an electron gas because of interactions with adjacent particles. At temperatures of 50000 to 80000 K, the contributions of the lowest excited states are comparable to that of the ground state, but the statistical weights corresponding to individual electron shells forming around protons are negligible as compared to those of electron-gas states.

The formation of electron shells is the key effect responsible for the drop in pressure observed with decreasing plasma temperature. Figure 4 shows the average number of electrons inside a sphere of radius R centered at a proton. At a temperature of 27000 K, about 20% of electrons are bound up in the electron shells around protons; i.e., the total number of free particles decreases by 10%. A similar decrease is observed in the compressibility factor as compared to that at the temperature boundary of 45000 K (see Fig. 1b).

Figure 5 shows the radius $R_e(T)$ of the sphere centered at a proton that contains one electron on average. Curve *1* was obtained for the mean overall density corresponding to a volume of 64000 Å per electron. The radius of the sphere having this volume is $R_e^\infty = 24.8$ Å. For curve *2*, the corresponding value is $R_e^\infty = 18.6$ Å. In the absence of spatial correlations, the sphere of radius R_e^∞ centered at a proton contains one electron on average. The deviation of $R_e(T)$ from R_e^∞ illustrates the degree of spatial correlation between protons and fluctuations of electron density in a plasma. Figure 5 dem-

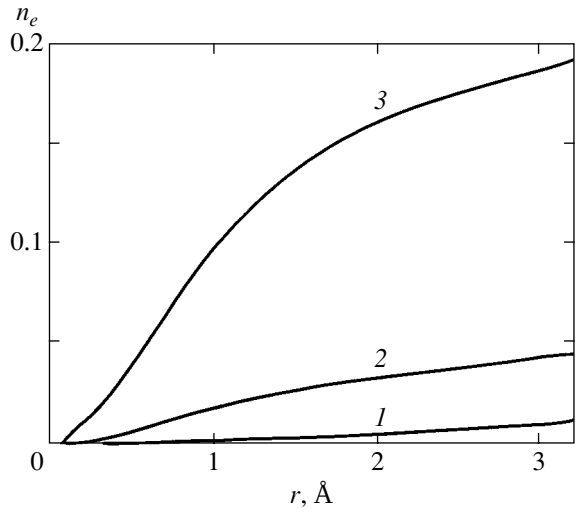

Fig. 4. Average number of electrons within a distance R from a proton for a hydrogen plasma of density 0.156×10^{20} cm^{-3} at a temperature of (*1*) 100000, (*2*) 45500, and (*3*) 26900 K.

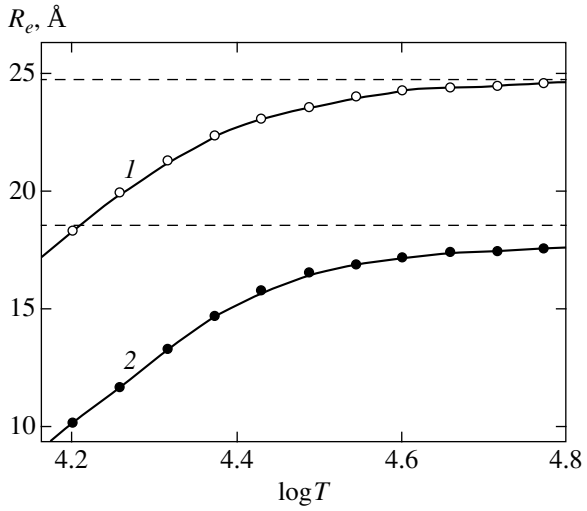

Fig. 5. Average radius of the sphere centered at a proton that contains one electron vs. temperature for plasmas of density (*1*) 0.156×10^{20} and (*2*) 0.370×10^{20} cm^{-3}.

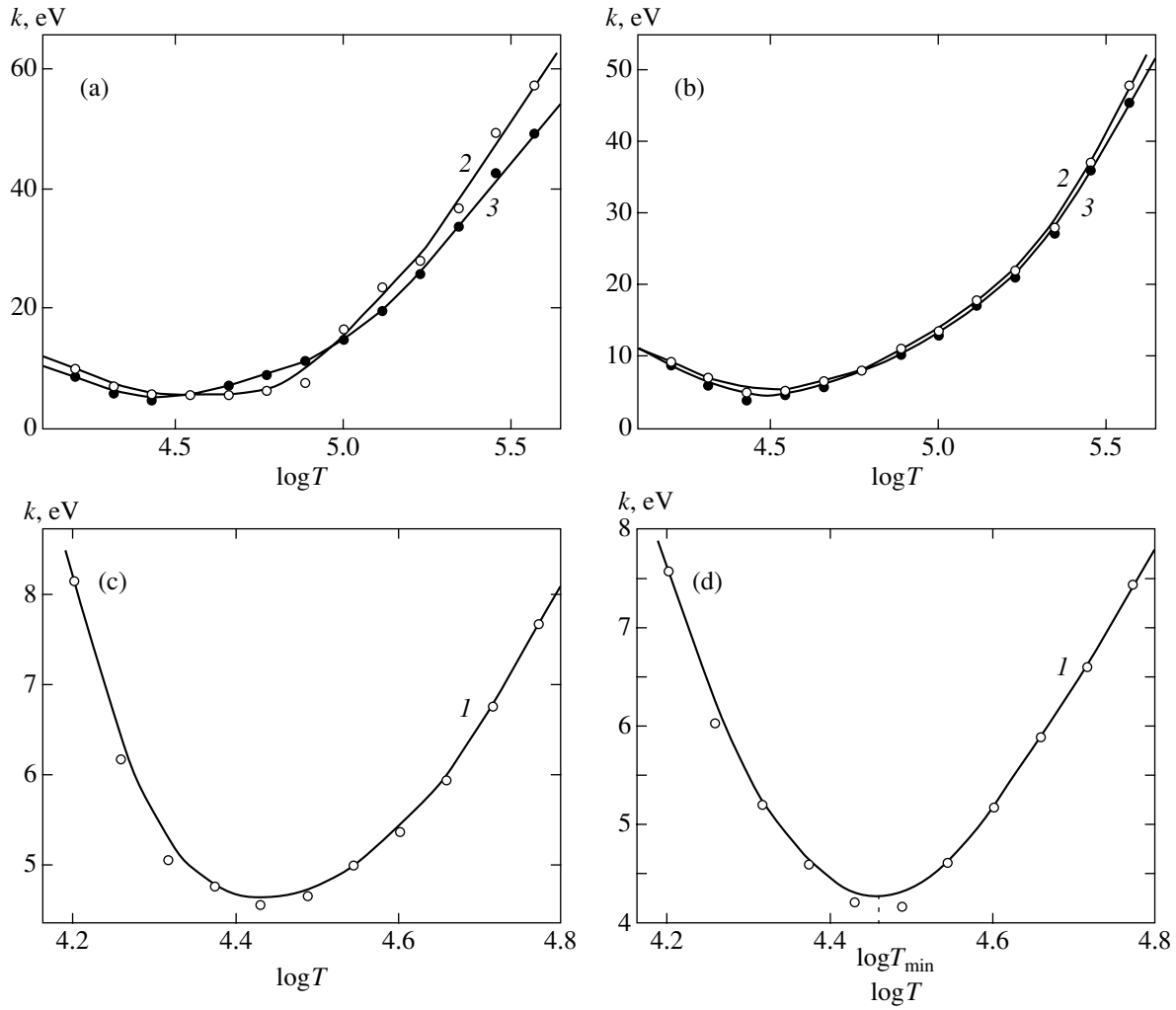

Fig. 6. Mean electron kinetic energy in a hydrogen plasma of density (*1*) 0.156×10^{20}, (*2*) 1.25×10^{20}, and (*3*) 0.370×10^{20} cm^{-3}: (a, c) general estimator (16); (b, d) virial estimator (35).

onstrates that the correlation increases with the overall density.

Figures 6a and 6c show the electron kinetic energy calculated as a function of temperature by averaging the general estimator given by (16); Figs. 6b and 6d, the results based on the virial estimator given by (35). The consistency of the results obtained by both methods demonstrates the accuracy of calculations of not only kinetic energy, but also the equation of state, because the calculation of kinetic energy based on the virial estimator includes the calculation of pressure. The calculations based on the virial estimator involve lower statistical errors, which is clear from the narrower scatter of data points about the approximating curve. As the plasma temperature decreases, a decrease in the electron kinetic energy is followed by an increase. The location of the energy minimum, $T_{\min} \approx 31000$ K, weakly depends on plasma density, shifting by about 5% toward lower temperatures as the density decreases by an order of magnitude. The increase in electron kinetic energy with decreasing temperature is a purely quantum effect explained by an increase in the statistical weights of electron states localized in the vicinity of protons. The dominant contribution is due to the ground state of the hydrogen atom, with a kinetic energy of 13.6 eV. The mean kinetic energy approaches this value well before the formation of electron shells is complete, because thermal motion in delocalized quantum states contributes 3 to 4 eV to the total kinetic energy per electron. The electron kinetic energy substantially deviates from the classical limit value $3k_BT/2$ at temperatures below 40000 K (see Fig. 7). The electron kinetic energy at a plasma temperature of 16000 K corresponds to the kinetic energy of classical particles at 100000 K. Owing to quantum effects, the electron subsystem is "heated": the mean electron energy is higher than the mean ion energy, and the difference amounts to several electronvolts with decreasing temperature (see Fig. 8). The particle energies plotted in Fig. 8 change sign from positive to negative (the plasma becomes strongly nonideal) at a temperature of about 30000 K, i.e., almost

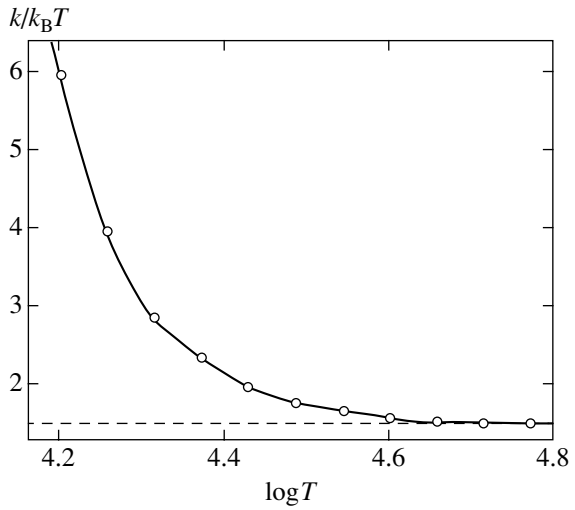

Fig. 7. Normalized mean electron kinetic energy vs. temperature calculated for a hydrogen plasma of density 0.156×10^{20} cm^{-3} by using general estimator (16).

simultaneously with the transition to quantum behavior of the electron component. A correct description of a hydrogen plasma must rely on a quantum statistical treatment of the electron component.

The results obtained in this study show that the temperature boundary separating the domains of quantum and classical behavior of electrons is located approximately at 40000 K for plasmas with $\rho \approx 10^{19}$–10^{20} cm^{-3}. Using the conventional criterion, based on comparison of the thermal de Broglie wavelength $\lambda = h/\sqrt{2\pi m_e k_B T}$ with the mean interparticle distance $\rho^{-1/3}$, one finds that the temperature boundary of the domain of quantum behavior of electrons is 12000 K. The curve shown in Fig. 7 demonstrates that the electron kinetic energy at this temperature differs from the classical limit value by a factor of several tens. The de Broglie criterion is inapplicable under these conditions, because it ignores strong Coulomb correlations between particles. On the other hand, a system of classical point charges is inherently unstable, because the corresponding partition function diverges at the singular points of the Coulomb potential. The convergence of the partition function and equilibrium averages for a hydrogen plasma is entirely due to the quantum-mechanical delocalization of its electron component. Irrespective of any criterion, "classical behavior" of the electron subsystem can be discussed only tentatively, because the very existence of the plasma at any temperature is essentially due to quantum behavior of electrons.

Electrostatic screening and quantum-mechanical delocalization of electrons determine particle–particle interactions and pressure in hydrogen plasmas. Figure 9 compares typical potentials of the mean force,

$$W(R) = -k_B T \ln[g(R)] + \text{const},$$

where $g(R)$ is the corresponding correlation function, calculated for electron–electron and proton–proton interactions with the Coulomb interaction potential for two elementary point charges in free space. The Coulomb interactions over distances larger than 18 Å are almost completely screened. Quantum delocalization manifests itself at distances smaller than 3 Å as a weaker electron–electron interaction as compared to proton–proton interaction. The potential W of the mean force of electron–proton interaction has a distinct maximum at 3 to 5 Å (see Fig. 10). To elucidate the nature of this maximum, a detailed study of correlations was

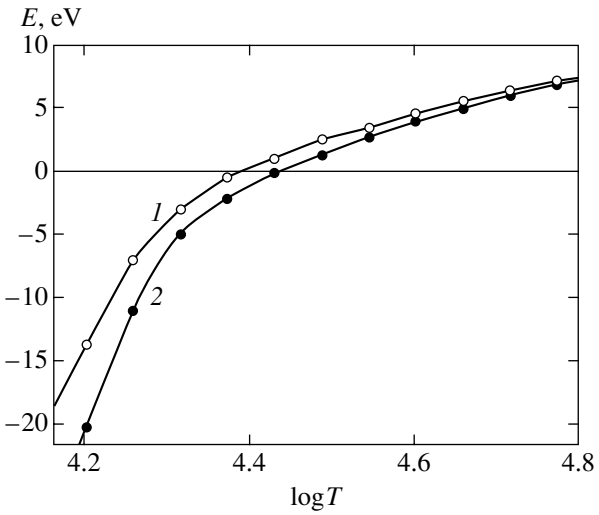

Fig. 8. Mean electron (*1*) and proton (*2*) total energy vs. temperature for a hydrogen plasma of density 0.156×10^{20} cm^{-3}.

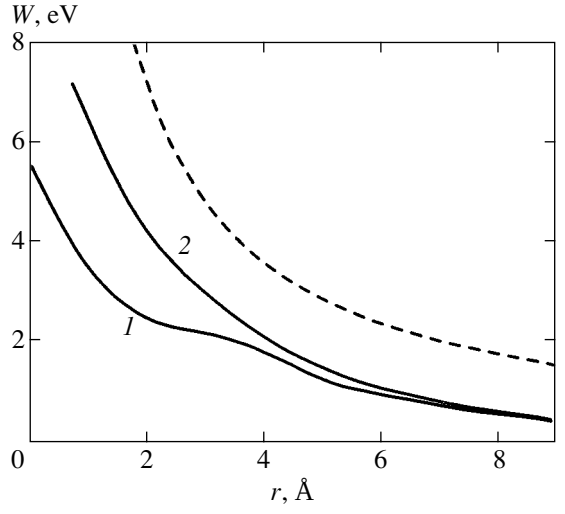

Fig. 9. Potential of the mean force of electron–electron (*1*) and proton–electron (*2*) interaction for a hydrogen plasma of density 0.370×10^{20} cm^{-3} at a temperature of 30700 K. Dashed curve: Coulomb interaction potential for two elementary point charges in free space.

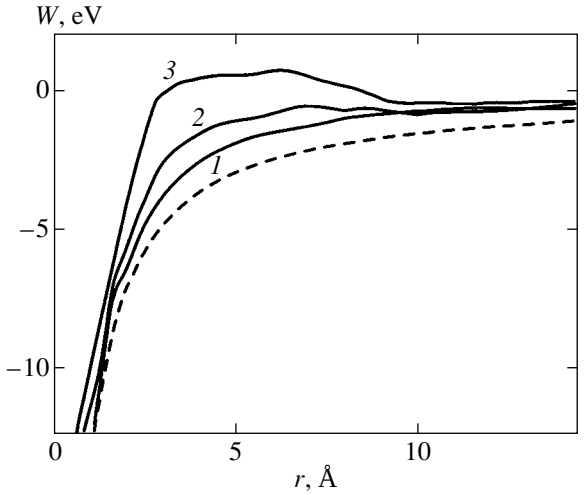

Fig. 10. Potential of the mean force of electron–proton interaction for a hydrogen plasma of density 0.370×10^{20} cm^{-3} at a temperature of (*1*) 39900, (*2*) 23600, and (*3*) 18200 K. Dashed curve: Coulomb interaction potential for two opposite point charges in free space.

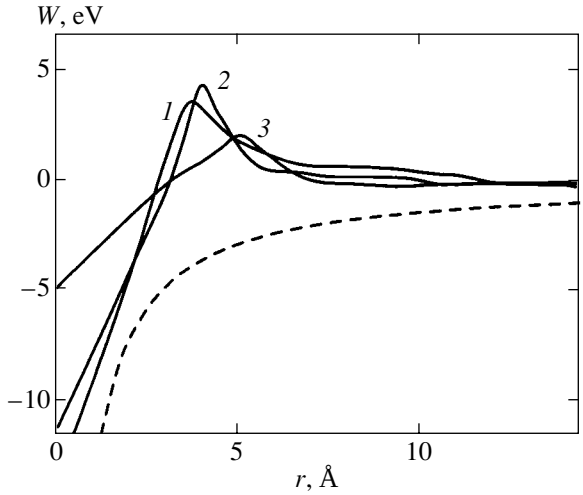

Fig. 11. Potential of the mean force of electron–proton interaction for a system of two protons and two electrons inside a sphere of radius 23.45 Å at a temperature of (*1*) 14000, (*2*) 11000, and (*3*) 4900 K.

performed for an isolated proton–electron pair and for a similar pair in the presence of another electron and in the presence of another electron and another proton. In the first case, the only proton was at rest at the center of a spherical volume of radius 18.6 Å. The corresponding density of a macroscopic system is 0.370×10^{20} cm^{-3}. The electron was freely moving in the electrostatic field generated by the proton. Long Markov chains were generated, and the Feynman path of the electron was discretized into 40 to 320 segments. None of the simulations revealed any maximum in the potential of the mean force analogous to that observed in Fig. 10 for a many-body system. A monotonic curve was obtained, albeit substantially different from that corresponding to the Coulomb interaction between point charges. The maxima exhibited by the curves plotted in Fig. 10 are obviously due to many-body correlations. To determine the minimal number of particles required to obtain these maxima, a system of two protons and two electrons was simulated. A proton was at rest at the center of a sphere of radius 23.45 Å, which corresponds to the same density as in the first case. Another proton was located at the periphery of the sphere. The potential of the interaction between the latter proton and two electrons was calculated. The results shown in Fig. 11 demonstrate (slightly narrower) maxima of approximately the same height, as in a macroscopic system. To ascertain the role played by the second proton in the formation of the maximum, additional computations were performed for a system consisting of a proton and two electrons in a sphere having the same volume. In this case, no maximum of the potential of the mean force was observed. Therefore, the second proton is required for the potential $W(r)$ to have a maximum. An analysis of the spatial configurations of proton and electron paths calculated in the course of simulations suggest the following mechanism responsible for the shape of the $W(r)$ curve. At lower temperatures, each electron tends to be increasingly bound to one of the protons. As the correlations grow stronger, the electron cloud tends to follow the motion of one of the protons, i.e., proton–electron correlations are modified by the effect of proton–proton correlations. At distances larger than about 5 Å, the proton–proton interaction is strongly screened and the corresponding energy is lower than $k_B T$ (see Fig. 9). Conversely, the protons are mutually repelled at distances smaller than 5 Å. Thus, when both electrons are localized near respective protons, an electron localized within a distance of about 5 Å from one proton cannot be attracted by the other proton; i.e., many-body correlations lead to repulsion between a proton and an electron. This effect is weaker at shorter distances, where the background electron density is higher.

REFERENCES

1. J. M. Caillol, J. Chem. Phys. **111**, 6528 (1999); **111**, 6538 (1999).
2. R. Rousseau and D. Marx, J. Chem. Phys. **111**, 5091 (1999).
3. W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
4. S. Fahy, X. W. Wang, and S. G. Louie, Phys. Rev. Lett. **61**, 1631 (1988).
5. D. C. Langreth and M. J. Mehl, Phys. Rev. B **28**, 1809 (1983).
6. X.-G. Zhang and D. M. C. Nicholson, Phys. Rev. B **60**, 4551 (1999).
7. Sh. Liu, P. W. Ayers, and R. G. Parr, J. Chem. Phys. **111**, 6197 (1999).
8. R. D. Muino and A. Salin, Phys. Rev. B **60**, 2074 (1999).

9. M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **111**, 911 (1999).
10. L. D. Fosdick and H. F. Jordan, Phys. Rev. **143**, 58 (1966).
11. H. F. Jordan and L. D. Fosdick, Phys. Rev. **171**, 128 (1968).
12. V. M. Zamalin and G. É. Norman, Zh. Vychisl. Mat. Mat. Fiz. **13**, 408 (1973).
13. V. S. Filinov, Teplofiz. Vys. Temp. **11**, 871 (1973).
14. V. S. Filinov, Teplofiz. Vys. Temp. **13**, 251 (1975).
15. V. S. Filinov, Teplofiz. Vys. Temp. **13**, 1146 (1975).
16. V. S. Filinov, Teplofiz. Vys. Temp. **14**, 245 (1976).
17. V. S. Filinov, Phys. Lett. A **54A**, 259 (1975).
18. V. S. Filinov, Phys. Lett. A **55A**, 219 (1975).
19. V. M. Zamalin, G. É. Norman, and V. S. Filinov, *Monte Carlo Method in Statistical Thermodynamics* (Nauka, Moscow, 1977; Mir, Moscow, 1977).
20. M. Mella, G. Morosi, and D. Bressanini, Phys. Rev. E **61**, 2050 (2000).
21. S. V. Shevkunov, Teplofiz. Vys. Temp. **29**, 45 (1991).
22. J. S. Kole and H. De Raedt, Phys. Rev. E **64**, 16704 (2001).
23. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, New York, 1977).
24. H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950; Nauka, Moscow, 1985).
25. M. C. Gordillo, Phys. Rev. B **60**, 6790 (1999).
26. P.-N. Roy and G. A. Voth, J. Chem. Phys. **110**, 3647 (1999).
27. N. B. Blinov and P.-N. Roy, J. Chem. Phys. **116**, 4808 (2002).
28. E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 2555 (1984).
29. D. M. Ceperley and E. L. Pollock, Phys. Rev. Lett. **56**, 351 (1986).
30. E. L. Pollock and D. M. Ceperley, Phys. Rev. B **36**, 8343 (1987).
31. D. M. Ceperley and E. L. Pollock, Phys. Rev. B **39**, 2084 (1989).
32. N. V. Blinov, P.-N. Roy, and G. A. Voth, J. Chem. Phys. **115**, 4484 (2001).
33. D. M. Ceperley and E. Manousakis, J. Chem. Phys. **115**, 10111 (2001).
34. P.-N. Roy, S. Jang, and G. A. Voth, J. Chem. Phys. **111**, 5303 (1999).
35. Sh. Miuza and S. Okazaki, J. Chem. Phys. **115**, 5353 (2001).
36. Ch. Chakravarty, J. Chem. Phys. **99**, 8038 (1993).
37. P. Beran, Phys. Rev. B **59**, 9725 (1999).
38. F. H. Zong, C. Lin, and D. M. Ceperley, Phys. Rev. E **66**, 36703 (2002).
39. E. Wigner, Phys. Rev. **46**, 1002 (1934).
40. S. V. Shevkunov, Available from VINITI, No. 6370-V87 (1987).
41. S. V. Shevkunov and P. N. Vorontsov-Velyaminov, Mol. Simul. **7**, 249 (1991).
42. S. V. Shevkunov, in *Modern Problems in Statistical Physics*, Ed. by I. R. Yukhnovskiĭ (Naukova Dumka, Kiev, 1989), Vol. 1, p. 379 [in Russian].
43. S. V. Shevkunov, Mat. Model. **2**, 3 (1990).
44. S. V. Shevkunov, Dokl. Akad. Nauk **369**, 43 (1999) [Dokl. Phys. **44**, 730 (1999)].
45. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **118**, 36 (2000) [JETP **91**, 31 (2000)].
46. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **121**, 1098 (2002) [JETP **94**, 943 (2002)].
47. S. V. Shevkunov, Dokl. Akad. Nauk **382**, 615 (2002) [Dokl. Phys. **47**, 109 (2002)].
48. M. Skorobogatiy and J. D. Joannopoulos, Phys. Rev. B **60**, 1433 (1999).
49. R. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965; Mir, Moscow, 1968).
50. F. A. Berezin, *Method of Second Quantization* (Nauka, Moscow, 1986; Academic, New York, 1966).
51. N. Marki and W. H. Miller, J. Chem. Phys. **90**, 904 (1989).
52. R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1978).
53. M. F. Herman, E. J. Bruskin, and B. J. Berne, J. Chem. Phys. **76**, 5150 (1982).
54. M. Takahashi and M. Imada, J. Phys. Soc. Jpn. **53**, 963 (1983).
55. M. Parinello and A. Rahman, J. Chem. Phys. **80**, 860 (1984).
56. J. D. Doll and D. L. Freeman, J. Chem. Phys. **83**, 768 (1985).
57. S. Jang and G. A. Voth, J. Chem. Phys. **115**, 7832 (2001).

Translated by A. Betev