Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 1–4.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 5–9.
Original Russian Text Copyright © 2002 by Korkin, Khriplovich.

GRAVITATION, ASTROPHYSICS

The Relation between the Maximum Entropy of a Quantized Surface and Its Area

R. V. Korkin[a, b, *] **and I. B. Khriplovich**[a, b, **]

[a]*Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,*
pr. akademika Lavrent’eva 11, Novosibirsk, 630090 Russia
[b]*Novosibirsk State University, ul. Pirogova 2, Novosibirsk, 630090 Russia*
**e-mail: rvkorkin@mail.ru*
***e-mail: khriplovich@inp.nsk.su*

Received December 27, 2001

Abstract—It is shown that the maximum entropy of a quantized surface in the classical limit is proportional to its area. The result is valid for the loop quantum gravitation as well as for a more general class of approaches to surface quantization. For some special cases, the maximum entropy is calculated in explicit form. © *2002 MAIK “Nauka/Interperiodica”.*

Let us consider the relation between the maximum entropy of a quantized surface and its area. For the sake of definiteness, we start our analysis with the approach to surface quantization based on loop quantum gravitation [1–5]. In this case, the surface geometry is determined by a set of ν punctures on this surface. In the general case, each puncture is supplied by two integral or half-integral “angular momenta” j^u and j^d:

$$j^u, j^d = 0, 1/2, 1, 3/2, \ldots . \qquad (1)$$

j^u and j^d are associated with the edges directed up and down relative to the normal to the surface, respectively, and add up to form the angular momentum j^{ud}:

$$\mathbf{j}^{ud} = \mathbf{j}^u + \mathbf{j}^d, \quad |j^u - j^d| \le j^{ud} \le j^u + j^d. \qquad (2)$$

The area of the surface is given by

$$A = \alpha l_p^2 \times \sum_{i=1}^{\nu} \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}. \qquad (3)$$

It is appropriate to make the following remark concerning the last formula. It is quite natural that the unit of area is the square of the Planck length:

$$l_p^2 = \frac{G\hbar}{c^3}. \qquad (4)$$

Each radical

$$\sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}$$

has the same order of magnitude as j_i^u and j_i^d. In this case, however, the area A will be finite in the classical limit of a large sum of quantum numbers if the power of

$$N = \sum_{i=1}^{\nu} \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)} \qquad (5)$$

is equal to the power of $\hbar$ in l_p^2. This argument, which was formulated in [6], is also quite natural. It can easily be verified by considering in quantum mechanics any expectation value which does not vanish in the classical limit. On the other hand, since large values of j are generally admissible, the expression for the surface area must contain the sum of square roots and not the sum of $j(j+1)$ or $\sqrt[4]{j(j+1)}$.

As regards the common numerical factor α in formula (3), it cannot be determined without using some additional physical considerations. This ambiguity appears due to the presence of a free parameter (so-called Immirzi parameter) [7, 8] corresponding to a family of nonequivalent quantum theories each of which can be used without resorting to such consideration. We believe that the value of this factor in formula (3) will be determined from an analysis of the entropy of a black hole. This idea (which was mentioned earlier in [9]) was investigated by one of the authors [10] for somewhat simplified models under the assumption that the entropy of an eternal black hole has the maximum value in equilibrium. This assumption stems from the work by Vaz and Witten [11], who used it in the model of a quantum black hole emerging during dust collapse. In this communication, we will confine our analysis to the calculation of the maximum entropy of the surface whose area is defined by formula (3) (or its generalization).

The entropy S of a surface is defined as the logarithm of the number of states of this surface with a fixed area A, i.e., with the fixed sum (5). Let ν_i be the number of punctures with a given set of angular momenta j_i^u, j_i^d, and j_i^{ud}. The total number of punctures is

$$\nu = \sum_i \nu_i.$$

We ascribe the statistical weight g_i to each puncture *i*. Since

$$\mathbf{j}_i^{ud} = \mathbf{j}_i^u + \mathbf{j}_i^d,$$

this statistical weight is equal (in the absence of additional constraints) to the number of possible projections $\mathbf{j}_i^{ud}$; i.e.,

$$g_i = 2j_i^{ud} + 1.$$

Then, the entropy is given by

$$\begin{aligned} S &= \ln\left[\prod_i \frac{(g_i)^{\nu_i}}{\nu_i!}\nu!\right] \\ &= \sum_i \nu_i \ln g_i - \sum_i \ln \nu_i! + \ln \nu!. \end{aligned} \tag{6}$$

The structure of expressions (3) and (6) is so different that, in the general case, the entropy obviously cannot be proportional to the surface area (see [10] for a more detailed analysis). It will be shown below, however, that this is the case for the maximum entropy in the classical limit.

It is natural to expect from combinatorial considerations that the absolute maximum of entropy is attained when all the values of quantum numbers j_i^u and j^{ud} are present. This guess is also confirmed by calculations for specific model cases [see [10]). We also assume that the characteristic numbers of punctures ν_i in the classical limit are large. Using Stirling's formula for factorials, we can transform expression (6) to

$$\begin{aligned} S &= \sum_i \left[\nu_i \ln(2j+1) - \left(\nu_i + \frac{1}{2}\right)\ln \nu_i\right] \\ &\quad + \left(\sum_i \nu_i + \frac{1}{2}\right)\ln\left(\sum_{i'} \nu_{i'}\right). \end{aligned} \tag{7}$$

We have omitted here the terms containing $\ln\sqrt{2\pi}$, which are close to unity in order of magnitude. The applicability of this approximation as well as Stirling's formula in the given problem will be discussed below.

We seek the extremum of expression (7) under the condition

$$\begin{aligned} N &= \sum_i \nu_i r_i = \text{const}, \\ r_i &= \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}. \end{aligned} \tag{8}$$

The problem is reduced to the solution of the system of equations

$$\ln g_i - \ln \nu_i + \ln\left(\sum_{i'} \nu_{i'}\right) = \mu r_i, \tag{9}$$

or

$$\nu_i = g_i \exp(-\mu r_i) \sum_{i'} \nu_{i'}. \tag{10}$$

Here, μ is a Lagrangian multiplier for the additional condition (8). Summing expressions (10) over *i*, we arrive at the following equation for μ:

$$\sum_i g_i \exp(-\mu r_i) = 1. \tag{11}$$

On the other hand, multiplying Eq. (9) by ν_i, summing over *i*, and taking into account condition (8), we obtain the following result for the maximum entropy for a given value of *N*:

$$S_{\max} = \mu N. \tag{12}$$

In formula (7), we have omitted the terms

$$-\frac{1}{2}\sum_i \ln \nu_i, \quad \frac{1}{2}\ln \nu.$$

The accuracy of this approximation will be considered again below.

Thus, it is the maximum entropy of the surface that is proportional to its area in the classical limit. This proportionality obviously takes place for a classical black hole. This is a strong argument in favor of the assumption that the entropy of a black hole has the maximum value.

It should be emphasized that relation (12) is valid not only for the loop quantum gravitation, but also for a wide class of approaches to surface quantization. In actual practice, the following requirements should be met. The surface must contain regions of different types so that there exist ν_i regions of each type *i* with the generalized effective quantum number r_i and the degeneracy g_i. Then, in the classical limit, the maximum entropy of the surface is proportional to its area.

From our point of view, the explicit form of the radicals in formula (3) (i.e., the specific choice of j_i^u, j_i^d, and j_i^{ud}) for black holes remains unclear. For this reason, we will confine our analysis to a few examples of a reasonable specific choice of radicals in the general quantization rule (3) for the surface area.

We begin with the case when formula (3) can be reduced to

$$A = \alpha l_p^2 \sum_{i=1}^{\nu} \sqrt{j_i(j_i+1)} = \alpha l_p^2 \sum_{j=1/2}^{\infty} \sqrt{j(j+1)}\nu_j. \quad (13)$$

The entropy for this case was determined by us earlier [10] under the assumption that the sum of quantum numbers

$$n = \sum_{j=1/2}^{\infty} j\nu_j$$

is specified. We will now solve this problem for a given

$$N = \sum_{j=1/2}^{\infty} \sqrt{j(j+1)}\nu_j. \quad (14)$$

Here, the degeneracy of a puncture with the quantum number j is $g_j = 2j + 1$, and Eq. (11) can be written in the form

$$\sum_{p=1}^{\infty} (p+1) z^{\sqrt{p(p+1)}} = 1, \quad p = 2j, \quad z = e^{-\mu/2}. \quad (15)$$

The solution to this equation is

$$z = 0.423 \quad \text{or} \quad \mu = -2\ln z = 1.722. \quad (16)$$

Thus, the maximum entropy in this case is given by

$$S_{\max} = 1.722N = 2.515\nu. \quad (17)$$

The mean value of angular momentum is

$$\langle j \rangle = \frac{1}{\nu} \sum_{j=1/2}^{\infty} j\nu_j = 1.059. \quad (18)$$

In a certain sense, the simplest way to choose the quantum numbers j_i in this model is to set all of them equal to 1/2. Then, $\nu_j = \nu\delta_{j,\,1/2}$ and

$$S = \ln 2\nu \quad \text{or} \quad S = \frac{2\ln 2}{\sqrt{3}}N. \quad (19)$$

In fact, this value of the entropy for a black hole was obtained earlier in [12] in the framework of the Chern–Simons field theory. The characteristic value of j_i determined in [12] is also equal to 1/2.

Let us consider another example. For each puncture, we obtain

$$j_i^u = j_i^d = j_i, \quad j_i^{ud} = 0.$$

In this case,

$$N = 2\sum_{j=1/2}^{\infty} \sqrt{j(j+1)}\nu_j, \quad (20)$$

and $g_j = 1$. Then, the equation for $z = \exp(-\mu/2)$ has the form

$$\sum_{p=1}^{\infty} z^{\sqrt{p(p+1)}} = 1, \quad (21)$$

and its solution is given by

$$z = 0.602 \quad \text{or} \quad \mu = -2\ln z = 0.508. \quad (22)$$

The maximum entropy and the average angular momentum obtained in this way are

$$S_{\max} = 0.508N = 1.655\nu, \quad (23)$$

$$\langle j \rangle = 1.224. \quad (24)$$

Finally, let us consider the general case when N is defined by formula (5), $g_i = 2j_i^{ud} + 1$, and all values of j_i^u, j_i^d, and j_i^{ud} are allowed. The solution to Eq. (11) in this case is

$$z = 0.202, \quad (25)$$

and the maximum entropy amounts to

$$S_{\max} = 3.120N = 4.836\nu. \quad (26)$$

The mean values of quantum numbers in this case are given by

$$\langle j^u \rangle = \langle j^d \rangle = 1.072, \quad \langle j^{ud} \rangle = 2.129. \quad (27)$$

It should be emphasized that we always arrive in this way at the effective condition of quantization with integral quantum numbers ν for the entropy (and area) of a black hole, which was proposed in the pioneering work [13] (see also [14, 15]).

Let us now analyze the accuracy of our result for the maximum entropy. In order to lend more weight and substance to our arguments, let us consider the second of the above models with quantity N defined by formula (21) and with the results described by relations (22)–(24). It can be easily verified, however, that the above estimates are qualitatively correct both for the first model described by formulas (14)–(18) and for the third, most general, case. For $\langle j \rangle \sim 1$, the number of punctures ν is of the same order of magnitude as N. Thus, $\nu \sim N \gg 1$ in the classical limit. In accordance with formula (10), numbers ν_j for $z \lesssim 1$ satisfy the condition $\nu_j > 1$ as long as the quantum numbers j are limited by the condition

$$j \lesssim J \sim \ln N. \quad (28)$$

Obviously, the typical values of ν_j which make significant contributions to N are large, and Stirling's approximation for S is quite justified.

On the other hand, the number of terms in the sum over j in expressions of the type (7) is effectively limited by inequality (28). Thus, the contribution from the terms with $\ln\sqrt{2\pi}$, which have been omitted in formula (7), as well as from the term $(1/2)\ln\nu$ retained in expression (7) but rejected in the final expression (12), amounts to only $\ln N$ in order of magnitude. The leading correction to our result (12) is due to the term

$$-\frac{1}{2}\sum_i \ln\nu_i$$

and amount to a value of the order of $\ln^2 N$. By the way, it differs from the leading correction in the model considered in [16, 17], where it is of the order of $\ln N$.

It should be noted in conclusion that attempts to calculate the entropy of the surface in loop quantum gravitation were also made in [18, 19]. In these publications, the distribution of angular momenta j over punctures is disregarded altogether. It is hard to imagine how the entropy of the surface can be determined without this information.

ACKNOWLEDGMENTS

We are obliged to G.G. Kirilin for his interest in this research and fruitful discussions.

This work was partly financed by the Russian Foundation for Basic Research (project no. 01-02-16898), the Leading Scientific School (grant no. 40-15-96811), the Ministry of Education of the Russian Federation (grant no. E00-3.3-148), and the Federal Program “Integration-2001.”

REFERENCES

1. C. Rovelli and L. Smolin, Nucl. Phys. B **442**, 593 (1995); Erratum: **456**, 753 (1995); gr-qc/9411005.
2. A. Ashtekar and J. Lewandowski, Class. Quantum Grav. **14**, 55 (1997); gr-qc/9602046.
3. R. Loll, Phys. Rev. Lett. **75**, 3048 (1995); gr-qc/9506014; R. Loll, Nucl. Phys. B **460**, 143 (1996); gr-qc/9511030.
4. R. De Pietri and C. Rovelli, Phys. Rev. D **54**, 2664 (1996); gr-qc/9602023.
5. S. Frittelli, L. Lehner, and C. Rovelli, Class. Quantum Grav. **13**, 2921 (1996); gr-qc/9608043.
6. I. B. Khriplovich, Phys. Lett. B **431**, 19 (1998); gr-qc/9804004.
7. G. Immirzi, Class. Quantum Grav. **14**, L177 (1997); gr-qc/9701052.
8. C. Rovelli and T. Thiemann, Phys. Rev. D **57**, 1009 (1998); gr-qc/9705059.
9. M. Bojowald and H. A. Kastrup, Class. Quantum Grav. **17**, 3009 (2000); hep-th/9907042; M. Bojowald and H. A. Kastrup, hep-th/9907043.
10. I. B. Khriplovich, gr-qc/0109092.
11. C. Vaz and L. Witten, Phys. Rev. D **64**, 084005 (2001); gr-qc/0104017.
12. A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov, Phys. Rev. Lett. **80**, 904 (1998); gr-qc/9710007; A. Ashtekar, C. Beetle, and S. Fairhurst, Class. Quantum Grav. **16**, L1 (1999); gr-qc/9812065; A. Ashtekar, A. Corichi, and K. Krasnov, Adv. Theor. Math. Phys. **3**, 419 (2000); gr-qc/9905089; A. Ashtekar, J. Baez, and K. Krasnov, Adv. Theor. Math. Phys. **4**, 1 (2001); gr-qc/0005126.
13. J. D. Bekenstein, Lett. Nuovo Cimento **11**, 467 (1974).
14. V. F. Mukhanov, Pis’ma Zh. Éksp. Teor. Fiz. **44**, 50 (1986) [JETP Lett. **44**, 63 (1986)].
15. Ya. I. Kogan, Pis’ma Zh. Éksp. Teor. Fiz. **44**, 209 (1986) [JETP Lett. **44**, 267 (1986)]; hep-th/9412232.
16. R. Kaul and P. Majumdar, Phys. Rev. Lett. **84**, 5255 (2000); gr-qc/0002040.
17. S. Carlip, Class. Quantum Grav. **17**, 4175 (2000); gr-qc/0005017.
18. C. Rovelli, Phys. Rev. Lett. **77**, 3288 (1996); gr-qc/9603063.
19. K. V. Krasnov, Phys. Rev. D **55**, 3505 (1997); gr-qc/9603025; K. V. Krasnov, Gen. Relativ. Gravit. **30**, 53 (1998); gr-qc/9605047; K. V. Krasnov, Class. Quantum Grav. **16**, 563 (1999); gr-qc/9710006.

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 101–105.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 116–121.
Original Russian Text Copyright © 2002 by Zalesskiĭ, Frolov, Zvezdin, Gippius, Morozova, Khozeev, Bush, Pokatilov.

SOLIDS
Structure

Effect of Spatial Spin Modulation on the Relaxation and NMR Frequencies of ^{57}Fe Nuclei in a Ferroelectric Antiferromagnet $BiFeO_3$

A. V. Zalesskiĭ[a], *, A. A. Frolov[a], A. K. Zvezdin[b], A. A. Gippius[c], **, E. N. Morozova[c], D. F. Khozeev[c], A. S. Bush[d], and V. S. Pokatilov[d]

[a] *Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninskiĭ pr. 59, Moscow, 117333 Russia*
** e-mail: nmr@ns.crys.ras.ru*
[b] *Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 117942 Russia*
[c] *Department of Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
*** e-mail: gippius@mail.ru*
[d] *Moscow State Institute of Radio Engineering, Electronics, and Automation (Technical University), pr. Vernadskogo 78, Moscow, 117454 Russia*

Received February 6, 2002

Abstract—The ^{57}Fe NMR spectra of the ^{57}Fe-enriched (95.43%) ceramics of a ferroelectric antiferromagnet $BiFeO_3$ with a spatially modulated magnetic structure have been studied. It is established that a cycloidal spin modulation in $BiFeO_3$ causes a spatial modulation of the spin–spin relaxation rate along the magnetic cycloid period and results in inhomogeneous broadening of the local NMR lineshape. It is also found that the local magnetic moments of Fe ions in various parts of the cycloid depend differently on the temperature, which is indicative of a difference in the spin wave excitation. The observed phenomena can be explained in terms of the Shul–Nakamura indirect nuclear interaction which becomes effective at high concentrations of the magnetoactive nuclei and low temperatures. Similarity of the obtained experimental results to the regularities of NMR in the Bloch walls is discussed. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The ferroelectric antiferromagnet $BiFeO_3$ is a perovskite-like compound which exhibits both magnetic and electric long-range ordering. The compound possesses a rhombohedrally distorted bimolecular perovskite unit cell (space group $R3c$) with the parameters (in hexagonal system) $a_{\text{hex}} = 5.57874$ Å and $c_{\text{hex}} = 13.8688$ Å [1]. Below the Néel temperature ($T_N \approx 670$ K), $BiFeO_3$ becomes an antiferromagnet with a cycloidal spatially spin-modulated structure (SSMS) established by neutron diffraction [2]. The magnetic structure of a separate unit cell corresponds to the G-type aniferromagnetic ordering (each magnetic Fe^{3+} ion is surrounded by six Fe^{3+} ions with spins directed opposite to that of the central ion) and can be described in terms of a two-sublattice model by introducing the antiferromagnetic vector $\mathbf{L} = \mathbf{M}_1 - \mathbf{M}_2$. However, the magnetic moments $\mathbf{M}_1$, $\mathbf{M}_2$, and the vector $\mathbf{L}$ are rotated in the plane of propagation of a modulated wave normal to the basal plane. The wave vector of this cycloid is perpendicular to the threefold axis (c axis) and lies in the plane of rotation of $\mathbf{L}$. According to the data obtained in [2], the cycloid period is sufficiently large, $\lambda = 620 \pm 20$ Å, and is incommensurate with the lattice parameter.

Dependence of the angle θ between $\mathbf{L}$ and the c axis on the coordinate x in the cycloid propagation direction is given by the relation [3, 4]

$$\cos\theta(x) = \operatorname{sn}\left(\pm\frac{4K(m)}{\lambda}x, m\right), \qquad (1)$$

where $\operatorname{sn}(x, m)$ is the elliptic Jacobi function, m is the parameter, and $K(m)$ is the complete elliptic integral of the first order.

Indirect evidence of the existence of SSMS in $BiFeO_3$ was obtained in [5–7] from measurements of the magnetoelectric effect. Among other results, the occurrence of a field-induced phase transition from SSMS to uniform antiferromagnetic state was observed. The existence of SSMS was recently confirmed by the ^{57}Fe NMR data [8, 9] providing some new information about the specific features of SSMS.

The rotation of magnetic moments ($\mathbf{L}$ vector) in SSMS modifies the NMR lineshape so that the spectrum becomes extended over a certain frequency range $\delta\nu$ with two edge peaks of different height and a characteristic dip in between. The frequency range arises due to an anisotropic contribution to the local magnetic field H_n on the nuclei which for $BiFeO_3$ amounts to about 1% of the total magnetic field on the Fe nucleus.

1063-7761/02/9501-0101$22.00 © 2002 MAIK "Nauka/Interperiodica"

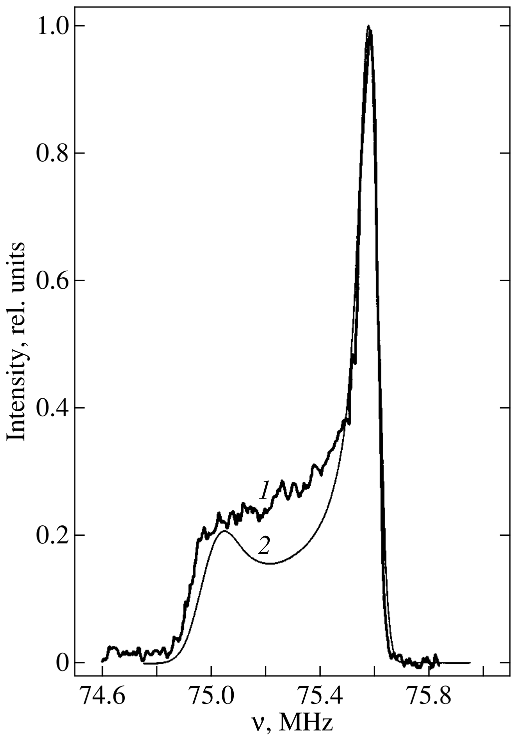


Fig. 1. (*1*) Experimental ^{57}Fe NMR spectrum for $BiFeO_3$ at 4.2 K and (*2*) its theoretical imitation by integral (11) with $m = 0.95$, $\alpha = 10$, and $\beta = 1$.

As shown in [8], anisotropy of the NMR frequencies in $BiFeO_3$ is described by the expression

$$\nu \approx \nu_{\parallel} - (\nu_{\parallel} - \nu_{\perp})\sin^2\theta = \nu_{\parallel} - \delta\nu\sin^2\theta, \qquad (2)$$

where $\nu_{\parallel}$ and $\nu_{\perp}$ are the frequencies corresponding to $\mathbf{L} \parallel c$ ($\theta = 0$) and $\mathbf{L} \perp c$ ($\theta = \pi/2$), respectively.

The profile of the NMR absorption spectrum $P(\nu)$ for SSMS in the δ-function approximation was calculated in [8]:

$$P(\nu) \sim \left\{ (\nu_{\parallel} - \nu_{\perp})^{1/2} (\nu - \nu_{\perp})^{1/2} \left[\frac{1}{m} - \frac{\nu - \nu_{\perp}}{\delta\nu} \right]^{1/2} \right\}^{-1}, \qquad (3)$$

where ν is the running frequency and m is the parameter of the elliptic Jacobi function in Eq. (1). The experiment shows that for $BiFeO_3$ $\nu_{\parallel} > \nu_{\perp}$.

It follows from Eq. (3) that for $m \longrightarrow 0$ the spectrum becomes symmetric with two edge peaks of equal intensity and a minimum between (a harmonic cycloid). If $m \longrightarrow 1$, the lineshape profile becomes asymmetric: the intensity of the high-frequency peak at $\nu_{\parallel}$ increases, the low-frequency peak at $\nu_{\perp}$ decreases, and the minimum of $P(\nu)$ between peaks shifts toward the low-frequency peak (the cycloid becomes more and more anharmonic). As follows from the experimental spectra measured in the temperature range from 77 to 300 K, SSMS is characterized by essentially anharmonic cycloid [8, 9].

In our previous studies [8, 9], the possible effect of nuclear relaxation on the NMR features in $BiFeO_3$ has been disregarded. Strictly speaking, Eq. (3) correctly describes the lineshape only if the relaxation processes are neglected.

We report the results of new NMR measurements in $BiFeO_3$ in the temperature range from 4.2 K to 77 K, which revealed new features of the relaxation of ^{57}Fe nuclei and spin wave excitations in SSMS.

2. SAMPLE PREPARATION AND EXPERIMENTAL PROCEDURE

The experiments were made on the same $BiFeO_3$ sample that was used in our previous studies [8, 9]. It was prepared by the standard ceramic technology using Fe_2O_3 oxide enriched in the ^{57}Fe isotope up to 95.43%. The parameters $a_{hex} = 5.589$ Å and $c_{hex} = 13.77(1)$ Å of a bimolecular rhombohedral unit cell coincide with the data for $BiFeO_3$ [1].

According to the experimental requirements, three types of pulsed NMR spectrometers were employed. One of these was a conventional superheterodyne spectrometer of the ISS-1-13 type with continuous recording of the spin echo amplitude as a function of the frequency, manufactured at the Institute of Radioengineering and Electronics of the Russian Academy of Sciences; the second was a pulsed coherent spectrometer with a point-by-point registration of the echo signal [10]; and the third spectrometer, also with a point-by-point registration, was a standard Bruker SXP-4-100 instrument. The measurements in the temperature range from 8 to 300 K were made with the use of a cryogenic closed-cycle helium refrigerator manufactured by CTI Cryogenics.

The spin–spin relaxation time T_2 was determined by measuring the decay of the echo amplitude A with increasing delay time τ_{12} between two exciting pulses according to the expression

$$A = A_0 \exp(-2\tau_{12}/T_2). \qquad (4)$$

The spin–lattice relaxation time T_1 was measured by the saturating pulse technique.

3. NUCLEAR RELAXATION IN A SPIN-MODULATED STRUCTURE

Figure 1 (curve *1*) shows the ^{57}Fe spin-echo spectrum of $BiFeO_3$ obtained at 4.2 K by continuously recording the echo amplitude versus frequency. The echo signal was excited by pulses of equal duration $\tau_1 = \tau_2 = 10$ μs. This mode allows one to “cut out” a frequency band of about 0.1 MHz in spectrum. The delay time between pulses was $\tau_{12} = 80$ μs. The spectrum retains the shape inherent in the spectra measured at higher temperatures [9]; however, it becomes more asymmetric with strong smearing of a low-frequency peak.

Detailed measurements of the spin–spin relaxation time T_2 by the above echo excitation technique revealed a strong frequency dependence of T_2. The exponential decay curves of the echo amplitude, measured at the frequencies $\nu_\parallel$ and $\nu_\perp$ and approximated by Eq. (4), are shown in Fig. 2. For the convenience of comparison, the initial amplitudes A_0 are taken equal to unity. The value of T_2 increases from $T_{2\perp} \approx 400$ μs at the frequency $\nu_\perp = 75.0$ MHz to $T_{2\parallel} \approx 1080$ μs at $\nu_\parallel = 75.58$ MHz.

The spectral variation of T_2 is retained at higher temperatures. The measurements made on different spectrometers give the following values of $T_{2\perp}$ and $T_{2\parallel}$: 250–260 μs and 500–520 μs (at a temperature of 30 K); 150–200 μs and 400–500 μs (77 K); 60–80 μs and 120–140 μs (300 K). With an increase in the temperature, the spin–spin relaxation time, as expected, becomes shorter, while the $T_{2\parallel}/T_{2\perp}$ ratio tends to decrease.

The measurements of the spin–lattice relaxation time T_1 revealed its constancy across the spectrum (3–4 ms at 4.2 K and 1.4–1.5 ms at 77 K).

For a theoretical analysis of the experimental NMR lineshape, one should pass from a simple approximation using the δ function as a local (intrinsic) lineshape [8, 9] to a model which takes into account the finite local linewidth and its variation with frequency along the cycloid period. These calculations can be made based on the ideas that were used for description of the NMR in the Bloch walls [11] possessing a similar 180° periodicity in the magnetic structure.

In contrast to the Bloch walls, there is no need in considering the enhancement factor in our case since, as was established in [8], the enhancement mechanism in $BiFeO_3$ is ineffective. The general expression for the NMR absorption lineshape for a periodic structure is given by

$$P(\nu) = \int_0^\pi I(\theta) f[\nu - \nu(\theta)] d\theta, \tag{5}$$

where $I(\theta)$ is the signal intensity in the part of the cycloid where **L** makes an angle θ with the c axis, and $\nu(\theta)$ is the anisotropic contribution to NMR frequency given by Eq. (2). The function $I(\theta)$ for $BiFeO_3$ was calculated in [8] and has the form

$$I(\theta) \propto (m^{-1} - 1 + \sin^2\theta)^{-1/2}. \tag{6}$$

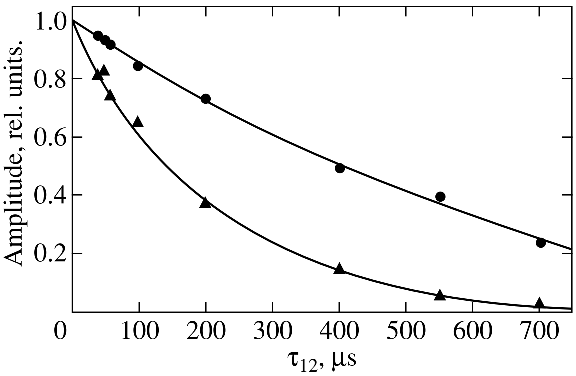


Fig. 2. Decay of the echo amplitude with an increase in the delay time τ_{12} between pulses measured at the frequencies (●) $\nu_\parallel$ and (▲) $\nu_\perp$ and approximated by Eq. (4). The initial amplitudes A_0 are taken equal to unity.

We take the Gaussian profile as a local lineshape function

$$F(\nu - \nu(\theta)) = \frac{2}{\sqrt{2\pi}\Delta(\theta)} \times \exp\left\{-\left(\frac{2}{\Delta(\theta)}\right)^2 \frac{(\nu - \nu(\theta))^2}{2}\right\}, \tag{7}$$

which assumes that the local linewidth (more precisely, a halfwidth) Δ is a function of the angle θ. The function $\Delta(\theta)$ is taken in the form

$$\Delta(\theta) = \Delta_\parallel + \delta_1 \sin^2\theta, \tag{8}$$

where $\Delta_\parallel$ is the minimum linewidth at the frequency $\nu_\parallel$ and δ_1 is the additional broadening at the frequency $\nu_\perp$. This choice of the function (8) will be justified below. In view of Eqs. (2), (6), and (8), we obtain the following expression for the integral lineshape:

$$P(\nu) \propto \int_0^\pi (m^{-1} - 1 + \sin^2\theta)^{-1/2} (\Delta_\parallel + \delta_1 \sin^2\theta)^{-1} \times \exp\left\{-\frac{2[\nu - (\nu_\parallel - \delta\nu \sin^2\theta)]^2}{(\Delta_\parallel + \delta_1 \sin^2\theta)^2}\right\} d\theta. \tag{9}$$

The integration is more conveniently performed with the dimensionless parameters

$$\xi = \frac{\nu - \nu_\parallel}{\delta\nu}, \quad \alpha = \frac{\delta\nu}{\Delta_\parallel}, \quad \beta = \frac{\delta_1}{\Delta_\parallel}. \tag{10}$$

From expressions (10), it follows that the value $\xi = -1$ corresponds to the frequency $\nu_\perp (\theta = \pi/2)$, and $\xi = 0$, to

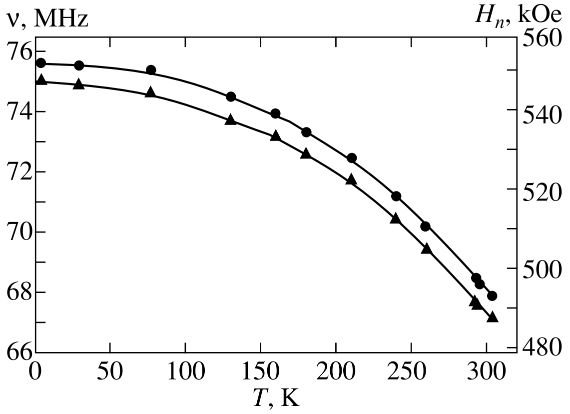


Fig. 3. Temperature dependence of the frequencies (●) $\nu_\parallel$ and (▲) $\nu_\perp$ and the corresponding local magnetic fields H_n on the ^{57}Fe nuclei.

the frequency $\nu_\parallel$ ($\theta = 0$), after which integral (9) reduces to

$$P(\nu) \propto \int_0^\pi (m^{-1} - 1 + \sin^2\theta)^{-1/2}(1 + \beta\sin^2\theta)^{-1} \times \exp\left\{-\alpha^2\frac{2(\zeta + \sin^2\theta)^2}{(1 + \beta\sin^2\theta)^2}\right\}d\theta. \tag{11}$$

Curve *2* in Fig. 1 shows the results of calculation of the integral in (11) at $m = 0.95$, $\alpha = 10$, and $\beta = 1$. In frequency units, the above parameters correspond to the local linewidths $\Delta_\parallel = 58$ kHz and $\Delta_\perp = \Delta_\parallel + \delta_1 = 116$ kHz. Taking into consideration that the line profile exhibits inevitable distortion due to a finite delay time between pulses during the experiment, i.e., that the spectral distribution of T_2 affects the echo amplitude (see Fig. 1), one may conclude that the theoretical lineshape quite satisfactorily describes specific features of the experimental NMR spectrum.

4. TEMPERATURE DEPENDENCE OF NMR FREQUENCIES

The temperature dependence of the frequencies $\nu_\parallel$, $\nu_\perp$, and the corresponding local magnetic fields H_n on the ^{57}Fe nuclei in BiFeO$_3$ is shown in Fig. 3. The numerical values of the frequencies $\nu_\parallel$ and $\nu_\perp$ and their difference $\delta\nu$ for the low-temperature range from 4.2 to 77 K and at 304 K are given in the table. Data for the range 77 K < *T* < 304 K are published in [9]. To within the experimental error of $\nu_\parallel$ and $\nu_\perp$, the frequency difference $\delta\nu$ in this temperature range can considered as constant. However, as can be seen from Fig. 3 and the table, the magnitude of $\delta\nu$ appreciably increases with the temperature due to a steeper drop in the $\nu_\perp$ frequency (corresponding to a low-frequency edge of the $\delta\nu$ band in the part of a cycloid where $\mathbf{L} \perp c$) in comparison with the frequency $\nu_\parallel$ (corresponding to $\mathbf{L} \parallel c$). Thus, on heating the sample from 4.2 to 77 K, the relative drop in $\nu_\perp$ (from 75.00 to 74.60 MHz) amounts to 0.53%, while the relative drop in $\nu_\parallel$ (from 75.58 to 75.35 MHz) is 0.30%. The temperature-induced variations of $\delta\nu$ in this temperature interval are significantly greater than the error of frequency measurements.

Table

T, K	$\nu_\perp$, MHz	$\nu_\parallel$, MHz	$\delta\nu$, MHz
4.2	75.00 ± 0.01	75.58 ± 0.01	0.58 ± 0.02
30	74.86 ± 0.01	74.54 ± 0.01	0.68 ± 0.02
77	74.60 ± 0.01	75.35 ± 0.01	0.75 ± 0.02
304	67.08 ± 0.01	67.88 ± 0.01	0.80 ± 0.02

In principle, the anisotropic contribution of $\delta\nu$ to NMR frequency can vary due to changes in the environment of Fe ions (the anisotropic contributions to NMR frequencies were analyzed for orthoferrites [12], for which the magnitude of $\delta\nu \approx 0.5$ MHz observed at the reorientation of spins is close to that given in the table for BiFeO$_3$). However, according to the structural data [13, 14], the numerical values of lattice parameters and interatomic distances in BiFeO$_3$ change in this temperature range only in fourth digit; thus, the structural factors cannot explain the temperature changes in $\delta\nu$.

Since a change in the ^{57}Fe NMR frequency is proportional to the change in the local magnetic moment of the Fe ion, the observed difference in the temperature dependence of the frequencies $\nu_\parallel$ and $\nu_\perp$ points to the different temperature dependence of magnetic moments in the parts of a cycloid where $\mathbf{L} \perp c$ and $\mathbf{L} \parallel c$.

5. DISCUSSION AND CONCLUSION

The experimental data on the NMR lineshape and the temperature dependence of NMR frequencies in BiFeO$_3$ strikingly resemble the NMR features observed in the Bloch walls [15]. These features are as follows: (1) broadening of the local NMR lineshape and the acceleration of nuclear relaxation processes toward the center of a wall in comparison with the nuclei in domains; (2) decrease in the local magnetization (the so-called "spin reduction" [15]) in a domain wall due to the excitation of "intrawall" spin waves (Winter magnons) which is manifested by an increase in the difference between the NMR frequencies in the domains and the wall center with increasing temperature.

The features of NMR in BiFeO$_3$ can be treated in a similar way, assuming the parts of a cycloid with the angles near $\theta = \pi/2$ as transition regions (the analog of walls) between more extended regions with the angles closer to $\theta = 0, \pi, \ldots$ (the analog of domains). Nonuniform rotation of the antiferromagnetic vector $\mathbf{L}$ is caused by anharmonicity of the cycloid: the rotation of **L** towards the basal plane near $\theta = \pi/2$ is steeper than in the regions where vector **L** is oriented close to the *c* axis (see Fig. 2a in [9]).

The analogy with domains and domain walls becomes more and more justified with decreasing temperature because of an increase in anharmonicity of the cycloid. This is evidenced by the tendency to an increase in the parameter m: $m \approx 0.48$ at 304 K [9], $m \approx 0.8$–0.9 at 77 K [8, 9], $m \approx 0.95$ at 4.2 K (curve *2* in Fig. 1).

At a high concentration of the ^{57}Fe isotope (95.43%) and a low temperature, one can expect the contribution δ_1 to the angular dependence of the NMR linewidth, resulting from the indirect nuclear spin–spin interaction (the Shul–Nakamura interaction) due to the excitation of spin waves similar to the Winter magnons in domain walls. The aforementioned transition regions near $\theta = \pi/2$ can play the role of domain walls. The amplitude of these excitations is proportional to $\sin^2\theta$ [11], which justifies use of the function $\Delta(\theta)$ in the form of Eq. (8). As to the absence of a noticeable spectral variation of $T_1(\theta)$, one would expect the effect of spin–lattice interaction on the lineshape to become important only for temperatures approaching T_N.

In addition to the broadening δ_1 of the local line at the frequency $\nu_\perp$, a significant difference is observed between the temperature dependences of $\nu_\parallel$ and $\nu_\perp$, which is caused by a decrease in the atomic magnetic moments of Fe at the center of a transition region due to the thermal intrawall-type excitations [15].

Thus, one more experimental result points to a deep analogy between processes observed in the ordinary domain walls and in SSMS.

The question concerning the nature of the intrawall excitations in SSMS remains open. In strongly anisotropic uniaxial and spatially modulated antiferromagnets like $BiFeO_3$, the mobile walls characterized by high values of the RF susceptibility and responsible for the enhancement of NMR absorption are absent. Such low-energy translation-like wall oscillations determine the thermodynamics of spins in the walls. However, the experimental and theoretical estimations of the enhancement factor show [8] that the enhancement mechanism in $BiFeO_3$ is ineffective. The spin waves in SSMS are very specific by themselves (see [16]): their properties and the properties of antiferromagnetic domains and walls in a complicated magnetic structure such as that of $BiFeO_3$ require a special theoretical consideration.

The main result of our study consists in the following. The spatial spin-modulated cycloidal structure in $BiFeO_3$ results not only in periodicity of the local fields on ^{57}Fe responsible for a specific NMR lineshape, but also in a spatial periodicity in the spin–spin relaxation rate, local linewidth, and local magnetic moment. The observed dynamic properties of a modulated structure are the consequence of the Shul–Nakamura interaction which becomes effective at low temperatures and high concentrations of the ^{57}Fe isotope.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 02-02-16369 and 99-02-17167.

REFERENCES

1. F. Kubel and H. Schmid, Acta Crystallogr. B **46**, 698 (1990).
2. I. Sosnowska, T. Peterlin-Neumaier, and E. Steichele, J. Phys. C: Sol. St. Phys. **15**, 4835 (1982).
3. I. Sosnowska and A. K. Zvezdin, J. Magn. Magn. Mat. **140–144**, 167 (1995).
4. M.-M. Tehranchi, N. F. Kubrakov, and A. K. Zvezdin, Ferroelectrics **204**, 181 (1997).
5. Yu. F. Popov, A. K. Zvezdin, G. P. Vorob’ev, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **57**, 65 (1993) [JETP Lett. **57**, 69 (1993)].
6. Yu. F. Popov, A. M. Kadomtseva, G. P. Vorob’ev, and A. K. Zvezdin, Ferroelectrics **162**, 135 (1994).
7. A. M. Kadomtseva, Yu. F. Popov, G. P. Vorob’ev, and A. K. Zvezdin, Physica B (Amsterdam) **211**, 327 (1995).
8. A. V. Zalessky, A. A. Frolov, T. A. Khimich, *et al.*, Europhys. Lett. **50**, 547 (2000).
9. A. V. Zalesskiĭ, A. K. Zvezdin, A. A. Frolov, and A. A. Bush, Pis’ma Zh. Éksp. Teor. Fiz. **71**, 682 (2000) [JETP Lett. **71**, 465 (2000)].
10. A. A. Gippius, Doctoral Dissertation in Physics and Mathematics (Mosk. Gos. Univ., Moscow, 2001).
11. M. A. Butler, Int. J. Magn. **4**, 131 (1973).
12. A. S. Karnachev, Yu. I. Klechin, N. M. Kovtun, *et al.*, Zh. Éksp. Teor. Fiz. **78**, 1176 (1980) [Sov. Phys. JETP **51**, 592 (1980)].
13. P. Fisher, M. Polomska, I. Sosnowska, and M. Szymanski, J. Phys. C: Sol. St. Phys. **13**, 1931 (1980).
14. A. J. Jacobson and B. E. F. Fender, J. Phys. C: Sol. St. Phys. **8**, 844 (1975).
15. E. A. Turov and M. P. Petrov, *Nuclear Magnetic Resonance in Ferro- and Antiferromagnets* (Nauka, Moscow, 1969; Wiley, New York, 1972).
16. A. I. Popov, G. A. Esina, and A. K. Zvezdin, Fiz. Tverd. Tela (St. Petersburg) **38**, 3091 (1996) [Phys. Solid State **38**, 1691 (1996)].

Translated by A. Zalesskiĭ

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 106–113.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 122–130.
Original Russian Text Copyright © 2002 by Buchel'nikov, Dan'shin, Linnik, Tsymbal, Shavrov.

SOLIDS
Electronic Properties

Static and Dynamic Properties of a Ferrite-Garnet Film in the Neighborhood of Orientational Phase Transitions

V. D. Buchel'nikov[a, *], N. K. Dan'shin[b, **], A. I. Linnik[b],
L. T. Tsymbal[b], and V. G. Shavrov[c, *]**

[a]*Chelyabinsk State University, Chelyabinsk, 454021 Russia*

**e-mail: buch@csu.ru*

[b]*Donetsk Institute of Physics and Technology, National Academy of Sciences of Ukraine,*
Donetsk, 340114 Ukraine

***e-mail: danshin@host.dipt.donetsk.ua*

[c]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences,*
Mokhovaya ul. 11, Moscow, 101999 Russia

****e-mail: shavrov@mail.cplire.ru*

Received December 24, 2001

Abstract—The static susceptibility and the spectrum of a magnetic-resonance mode of a single-crystal garnet-ferrite film with a domain structure are experimentally investigated. It is found that, in a magnetic field perpendicular to the film plane, these characteristics have features associated with the reorientation of the domain structure. The spin-wave spectrum of the film is calculated on the basis of the domain structure model proposed. It is shown that the experimentally observed features of the spectrum can be accounted for by the reorientation of magnetizations in the domains. A good agreement is obtained between experimental and theoretical values of the gap in the spin-wave spectrum at the "starting and end points of reorientation." © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that the activation in the spectra of low-frequency branches of spin-wave oscillations in ordered single- and multisublattice magnetic materials is mainly determined by the crystallographic magnetic anisotropy. At the points of orientational phase transitions of the second kind (OPT-2), the magnetic anisotropy vanishes. In this case, the activation in the spectra of low-frequency spin oscillations is attributed to other (as a rule, weaker) interactions. In a two-sublattice model of an antiferromagnetic material, an orientational transition may consist in a smooth turn through 90° of the spin subsystem as a whole. It is this kind of reorientation that is realized in the subsystem of iron spins in weak ferromagnetic materials—rare earth orthoferrites (REOFs). To date, the nature and the mechanisms of activation in the spectrum of low-frequency branches of quasi-spin waves in these compounds have been sufficiently well studied both near spontaneous [1] and field-induced [2] orientational transitions. It turns out that a gap in the spectrum of a soft quasi-ferromagnetic mode in REOFs represents a certain measure of dynamic interaction between various oscillatory subsystems of the magnetic substance.

The same role that the crystallographic magnetic anisotropy plays in "thick" three-dimensional samples can be played, both in statics and dynamics, by the shape anisotropy in thin magnetic films. This fact follows from extensive analogies that manifest themselves when juxtaposing static and dynamic experiments in three-dimensional samples, on the one hand, and in virtually two-dimensional samples, on the other; the latter samples are represented by films whose linear dimensions L in the plane are much greater than their thickness d ($L \gg d$). This fact can be verified by comparing the structure of the orientational transition induced by an external field $\mathbf{H}$ (this transition is called the "end point of spin reorientation") in a three-dimensional sample of a REOF and in a thin film. As is well known, the orientational transition consists in the following: under the external field directed strictly along one of the crystal axes (either $\mathbf{a}(x)$ or $\mathbf{c}(z)$), the ferromagnetic $\mathbf{M}$ and antiferromagnetic $\mathbf{L}$ vectors rotate until, at a certain $H = H_{\text{tr}}$, the vector $\mathbf{M}$ becomes parallel to the field, whereas $\mathbf{L} \perp \mathbf{H}$. This is the end point of reorientation. For example, when $\mathbf{H} \parallel \mathbf{a}$, this point corresponds to the transition

$$\Gamma_{24}(M_{xz}, L_{xz}) \longleftrightarrow \Gamma_2(M_x, L_z).$$

One usually observes a jump in the magnetic susceptibility at the transition point, while a soft (quasi-ferromagnetic) resonance mode has the minimal frequency, which corresponds to the energy gap in the spectrum of low-frequency spin waves. We should stress that even a small deviation of the field from the appropriate crystal axis transfers an REOF sample into an angular phase, and the above transition disappears.

Now, let us consider, for example, a magnetically ordered film of a garnet-ferrite-type cubic ferromagnet with a negative first constant of cubic anisotropy: $K < 0$. This ferromagnet has eight easy axes of type {111}. Let a normal **n** to the film surface coincide with one of these axes. Then, the remaining six axes make acute angles with the film plane. It is known experimentally that the saturation field in such films does not exceed 2000 Oe, which is less by two orders of magnitude than the first exchange field H_{E1} in garnet ferrites. Therefore, one can abstract away from the multisublattice structure of a garnet ferrite and consider it as a ferromagnet with magnetization **M** for magnetic fields of up to 2000 Oe. It is known [3] that, depending on the relation between the squared modulus of the vector of spontaneous magnetization **M** and the constant of induced uniaxial anisotropy K_1 (directed along the normal to the film plane for the easy axis), different configurations of the domain structure may be formed in such films. For example, when $K_1/2\pi M^2 > 1$, the vector of spontaneous magnetization in individual domains is perpendicular to the film surface and is directed along easy axis. In this case, the domain structure itself in a zero magnetic field represents an array of strongly bent narrow stripes with width of the order of the film thickness. When $K_1/2\pi M^2 < 1$ and the film thickness is less than a certain critical value, $d < d_{\text{cr}}$, the spontaneous magnetic moment in similar stripe domains lies strictly in the plane of the film [3]. It is quite obvious that switching on a field $\mathbf{H} \parallel [111]$ in the second case pulls up the magnetizations **M** in the domains from the plane (111) and turns them toward axis [111] until they become parallel to **H**. By analogy with REOFs, this is also the end point of reorientation. Since this is a full-fledged orientational transition, a question arises: To what extent can the theory and the results of [1, 2] be applied to describing the dynamics of this transition? Do such films exhibit the static and dynamic phenomena inherent in OPT-2 in three-dimensional magnetic materials? To answer these questions, we carried out static and resonance experiments on a doped yttrium iron garnet (YIG) film and performed appropriate theoretical computations.

2. EXPERIMENT

2.1. Sample, Ground State, and Reorientation Structure

We used a single-crystal $(\mathrm{YLaGd})_3(\mathrm{FeGa})_5\mathrm{O}_{12}$ garnet-ferrite film grown on a gallium gadolinium garnet substrate by the liquid-phase epitaxy technique. The sample was made in the form of a disk 5 mm in diameter. As we have already mentioned, in a zero magnetic field, such a film consists of stripe domains, with width of the order of the film thickness, each of which is magnetized up to the saturation, while the film as a whole is unmagnetized [3]. It should be noted that the aforementioned composition of the film was chosen to minimize the growth-induced uniaxial anisotropy in the direction of the normal **n** to the film surface [4]. The film thickness was $d \approx 0.1$ µm, and the axis [111] was nearly parallel to the normal **n**. This thickness is an order of magnitude less than d_{cr} [3]; hence, we can rightfully assume that the magnetization vector in this film lies strictly parallel to the film plane. In this case, we experimentally found that the saturation field in the film plane does not exceed 10–15 Oe. At the same time, the saturation field along axis [111] is 1300 Oe. Thus, we can assume that, in a zero external magnetic field, an easy-plane-type anisotropy is formed in this film.

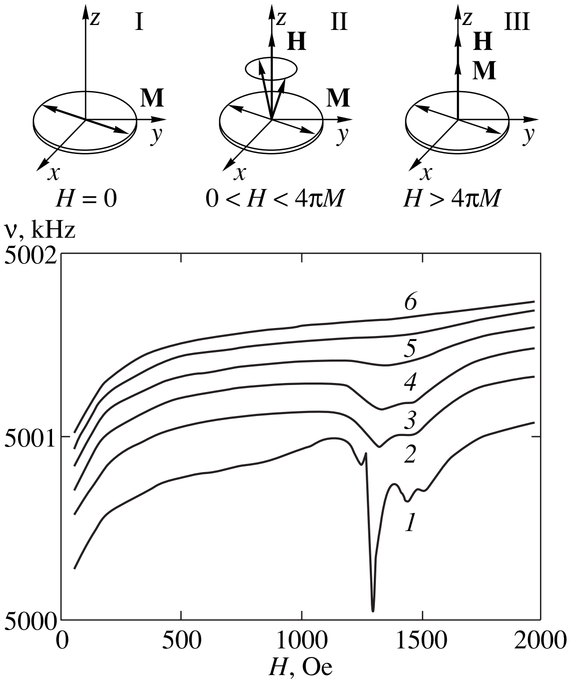


Fig. 1. Frequency of the resonance circuit of the autodyne oscillator with a garnet-ferrite film versus magnetic field H for various angles θ between the direction of H and the [111] axis: (*1*) 0°, (*2*) 0.13°, (*3*) 0.2°, (*4*) 0.8°, (*5*) 1.8°, and (*6*) 3.3°. The inset shows equilibrium magnetic states of the film.

The inset in Fig. 1 shows three basic states that are realized in a sample as the field $\mathbf{H} \parallel [111] \parallel z$ increases from 0 to $H \geq 4\pi M$ $(H \ll H_{E1})$. Transitions between these states can be represented as follows. When $H = 0$, a stripe domain structure of composite configuration is formed in the film with the width of domains of order d, while the spontaneous magnetic moment **M** in each domain lies in the film plane. In Fig. 1, this state is represented by an arbitrary pair of vectors of spontaneous magnetization **M** in adjacent domains. When a filed $\mathbf{H} \parallel$ [111] is switched on, the total magnetization in this direction grows predominantly due to the rotation of magnetization vectors in the domains. This rotation of magnetization vectors in the domains can be considered as the reorientation of magnetizations along the direction of the field. The reorientation starts at zero field and ends at a field of $H = H_{tr} \approx 4\pi M$. In our case, $H_{tr} \approx$ 1300 Oe. The field $H = 0$ is the "starting point of reorientation," while the field H_{tr} is the "end point of reorientation." The field H_{tr} makes the sense of a field induced by the shape anisotropy in such a film (the demagnetizing field). The presence of two points, the starting point and the end point of reorientation, makes the analogy between reorientations in an REOF and in a garnet-ferrite film even deeper. In a garnet-ferrite film, these points correspond to the points of spontaneous second-order orientational transitions in REOFs [1, 2]. Therefore, just as in REOFs, one should expect the softening of the low-frequency branch of quasi-spin waves in garnet-ferrite films at both above-mentioned points.

Note that, apparently, such a reorientation by no means requires that the film should be single-crystal. It is sufficient that it has a texture with the same distinguished axis [111] $\parallel$ **n**. Then, the film is virtually isotropic in the plane (111) in the sense of the distribution of the vectors **M** in state I. In state II, these vectors form a cone, a certain half-open umbrella that is closed for $H = H_{tr}$. Finally, to realize the above reorientation, the only important condition is that there exist three phases: a multidomain collinear, a multidomain angular, and a single-domain collinear phase. Even the number of sublattices in a specific magnetic material is unimportant, because, in statics, everything is reduced to the field dependence of the total magnetization, whereas, in dynamics, a ferromagnetic (acoustic) mode in weak fields results from the precession of the entire bundle of magnetizations as a single whole [5].

2.2. Static Properties (Magnetic Susceptibility)

It is known that the static susceptibility experiences a jump at the point of the magnetic OPT-2. We used an autodyne oscillator, which is usually applied in nuclear magnetic resonance experiments, to measure the static susceptibility of the film. The film was placed flat on a planar coil of the autodyne oscillator operating at a frequency of $\nu \sim 5$ MHz. In the experiment, we measured a dc voltage proportional to the difference between the frequencies of the autodyne oscillator and a reference oscillator, which was tuned to a frequency of 5 MHz. This technique allowed us to fix frequency variations of several hertz [6]. Here, we could set an external magnetic field near the orientation **H** $\parallel$ [111] accurate to within ±3 angular minutes.

Figure 1 represents a family of functions $\nu(H)$ for various angles θ of deviation of **H** from the axis [111]. The direction of **H** corresponding to the maximal variation in the resonance frequency ν of the autodyne circuit was determined from the symmetry of the graphs $\nu(H)$ for different values of θ. The frequency scale in Fig. 1 corresponds to the lowest graph. For the other graphs, one should move the frequency scale upward so that their initial regions coincide.

One can see that the family of graphs for different θ includes one that has the deepest minimum. This graph has the following characteristic features.

1. The frequency minimum is observed at the value of the external field equal to the saturation field of the film (about $4\pi M$) in the direction [111].

2. The depth of the minimum strongly depends on the orientation of the external magnetic field. For example, even a 0.1° deviation of **H** from the optimal direction corresponding to the deepest minimum results in a decrease in the depth of the minimum approximately by an order of magnitude. If the field deviates from the optimal direction by several degrees, then this minimum vanishes completely.

3. No hysteresis is observed in the position of the frequency minimum on the axis of fields.

With what can one associate the frequency minimum observed? Since the resonance frequency ν of the oscillator circuit coil with a sample is proportional to $\mu^{-1/2}$ (μ is the permeability of the film), the graphs in Fig. 1 represent the field dependence of the square root of the inverse permeability $(1 + 4\pi\chi)^{-1/2}$. Then, the minimum on the frequency–field diagrams can be considered as a natural jump in the susceptibility χ at the point of the magnetic OPT-2. The above-listed features of the behavior of this minimum show that here we deal with the "order–order" phase transition of the second kind. From this viewpoint, the lowest graph in Fig. 1 corresponds to $\theta = 0$, the most precise orientation of the constant magnetic field with respect to the axis [111]. The aforementioned OPT-2 takes place only in this case. When $\theta \neq 0$, the transition disappears, which is manifested in a sharp decrease in the jump of susceptibility (or the depth of the minimum in Fig. 1). This provides an obvious analogy with the properties of orientational transitions in REOFs [1, 2].

Notice that all graphs in Fig. 1 also exhibit a frequency minimum at very small fields. Apparently, this is associated with the fact that, as $H \longrightarrow 0$, the susceptibility attains its maximum associated with the second OPT-2 at the starting point of reorientation. One can see that the depth of the minimum in this case is virtually independent of the orientation of the external magnetic field. Such a behavior of the resonance frequency can be accounted for by the fact that any orientation of the external magnetic field (except for $\theta = \pi/2$) will obviously pull up the magnetization vectors of the domains from the plane of the film, i.e., will give rise to the OPT-2 at the starting point of reorientation.

Everywhere above, we pointed out that the exact orientation of the field is the direction **H** $\parallel$ [111] rather than **H** $\parallel$ **n**. This is not accidental. Experiments show that, when growing epitaxial garnet-ferrite films, the axis [111] does not always exactly coincide with the normal **n** to the film. On the other hand, it is known that it is the two-dimensionality of the film ($L \gg d$) that is responsible for the origin of the stripe structure; hence, the description of this structure always involves a geometric factor, the normal **n** to the surface. At the same time, an important moment for the realization of the orientational phase transition at the end point of reorientation is the orientation of the field with respect to a physical factor, a crystallographic axis (in our case, the axis [111]). Therefore, it is quite possible that, having found a sufficiently accurate orientation **H || [111]** from the symmetry of the graphs in Fig. 1, we certainly do not satisfy the condition **H || n**. Growth defects and the magnetic inhomogeneity of the film should also contribute to this fact; these factors can be manifested in the nonmonotonicity of the graphs on both sides of the frequency minimum, which is most clearly exhibited in the graph for $\theta = 0$. However, it is quite possible that these inhomogeneities are of physical, rather than technological, origin. A detailed investigation of these inhomogeneities is beyond the scope of this paper. It should only be noted that a visual observation of the domain structure in various epitaxial garnet-ferrite films shows that, as a rule, these films are not perfect. The most perfect structures (a stripe structure, an array of cylindrical magnetic domains, etc.) are observed when there is a strong growth-induced anisotropy in the direction of **n**, which, however, is minimal in the film under investigation.

Thus, the static properties of a garnet-ferrite film whose magnetic structure is mainly determined by the shape anisotropy show that two magnetic orientational phase transitions of the second kind are realized in this film. The properties of these transitions, such as a jump in the susceptibility, a strong field dependence in the case of the transition at the end point of reorientation, the absence of hysteresis, and the value of the transition field, are the features that are characteristic of ordinary OPT-2 in three-dimensional magnetic materials whose magnetic structure is determined by the crystallographic anisotropy.

2.3. Dynamic Properties (Quasi-Ferromagnetic Resonance)

It is known that the lowest frequency branch of the oscillation mode in the spectrum of an ordered magnetic material softens at the points of OPT-2. The dynamics in the neighborhood of various orientational transitions was analyzed in detail (theoretically and experimentally) in [1, 2] by an example of REOFs. It turns out that eight sublattices of REOFs can be reduced in dynamics to four sublattices, or, if one restricts the analysis to acoustic modes, even to two: ferromagnetic and antiferromagnetic ones, which correspond to the oscillations of the vectors **M** and **L**, respectively. A similar situation is possible when studying a multisublattice ferrimagnetic material, a garnet ferrite. Then, the dynamics in the isotropic case can also be described by the above two vectors. Just as in the case of REOFs, we are only interested in the quasi-ferromagnetic mode, associated with the precession of the vector **M** in the magnetic field, that should soften at the transition point (the frequency of the antiferromagnetic mode of a garnet ferrite lies in the infrared range). However, as we have already mentioned above, the cubic anisotropy of a garnet ferrite initially leads to a situation where several vectors **M** precess in the field **H** when the latter is directed along the easy axis [111] of the crystal. In this case, the film dynamics near the point of orientational transition at the starting and end points of reorientation is associated with the precession of the total magnetization along the axis [111] at an appropriate consolidated frequency.

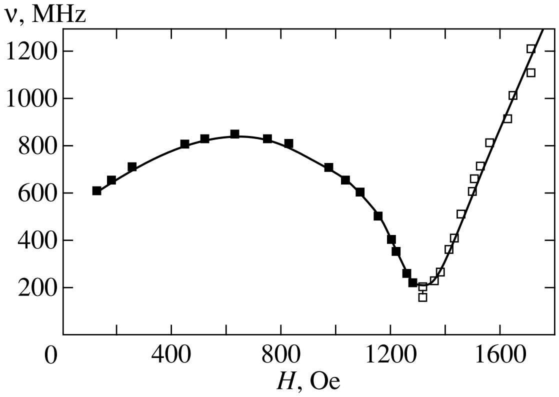


Fig. 2. Frequency–field dependence of the ferromagnetic resonance in a garnet-ferrite film in the field **H** || [111], **H** ⊥ **h**.

Resonance measurements were carried out on a reflecting direct-amplification microwave spectrometer with the range of operating frequencies from 10 to 8000 MHz under the modulation of the magnetic field at room temperature (295 K). In the experiments, we recorded the derivative of the absorbed signal with respect to the magnetic field. A microstrip line to which a sample in the form of a disk 3 mm in diameter was pasted served as a measurement cell. The external magnetic field was directed along the axis [111] by the method conventional in the investigation of soft magnetic-resonance modes in a magnetic field (see, for example, [7]) and normally to the magnetic component of the microwave field **h**. The experimental frequency–field dependence of the ferromagnetic resonance shown in Fig. 2 is obtained from a family of records made for a large number of fixed frequency points.

The magnetic-field dependence of the resonance frequencies represented in this figure has pronounced minima at low fields and at a field of about 1300 Oe, i.e., exactly at the fields ($H = 0$ and $H = H_{tr}$) where the susceptibility experiences a jump. This fact suggests that the spectrum observed corresponds to soft quasi-ferromagnetic modes because they always have the minimum frequency at the points of OPT-2. Just as in the REOF experiments, the prefix "quasi" implies that the magnetic-resonance spectrum near the transition may be formed not only due to the precession of magnetization. As is shown in [1, 2], actually, this spectrum is always associated with the dynamic interaction between various oscillatory subsystems of the magnetic material. This interaction in an REOF results in an energy gap at the transition point, which represents an additive sum of contributions of different collective oscillations. Figure 2 shows that a gap exists in our case as well, in particular, at the end point of reorientation $\nu_0 = 200 \pm 10$ MHz. This absolute value of gap is two orders of magnitude less than that usually observed in REOFs. Such a difference can be attributed to the fact that the spontaneous ferromagnetic moment in REOFs is so small that the primary role in dynamics is played by the antiferromagnetic structure of iron spins. In antiferromagnetic materials, the appearance of a gap at the point of OPT-2 represents an exchange-amplified effect [8]. In ferrimagnetic materials, the magnitude of M is of the order of the magnetizations M_1 and M_2 of the sublattices. Therefore, we actually deal with a ferromagnetic material where the aforementioned exchange amplification is absent. Nevertheless, the gap observed in this case can also be attributed to the dynamic interaction between various oscillatory subsystems, which are associated not only with the precession but also with the longitudinal oscillations of magnetization. Indeed, the ratio of the temperature of the induced OPT-2 realized here to the ferrimagnetic-ordering temperature $T_N = 560$ K of a garnet ferrite is rather large; it is approximately equal to 0.7. Under these conditions, the ratio of the longitudinal susceptibility to the transversal one ($\chi_\parallel/\chi_\perp$) can reach a value of 0.6 [9]. For example, in REOFs, the contribution of longitudinal oscillations of magnetization to the value of the gap for such value of $\chi_\parallel/\chi_\perp$ is quite significant [2, 9].

Notice that the experimental frequency–field dependence shown in Fig. 2 qualitatively agrees with that calculated in [10] for a spherical sample of a cubic ferromagnetic material with $K > 0$ for $\mathbf{H} \perp \mathbf{h}$ and $\mathbf{H} \parallel [110]$ without taking into account the displacement of domain walls. In these calculations, the gap in the spectrum is associated solely with the domain structure, which is preserved up to the saturation field of the sample. Despite obvious differences in the conditions and the geometry of our experiment from the initial conditions of calculation in [10], we can also speak of a soft mode in our case at the point of the orientational order–order transition (the reorientation occurs inside the domains). However, in contrast to usual transitions, for example, OPT-2 on REOFs, here the transitions of the type "start and end of reorientation" occur between inhomogeneous ordered (I and II) and between inhomogeneous (II) and homogeneous (III) ordered phases (Fig. 1), rather than between homogeneous phases. If we assume that a similar situation can be realized in our case (i.e., a domain structure can be preserved up to $H = H_{tr}$), then, in addition to the contributions of the known mechanisms, one should take into account the role of domains in the formation of the gap [1, 2, 9].

3. THEORY

It follows from the experiment that a composite easy-plane-type stripe domain structure is formed in the garnet-ferrite film in the absence of a constant external magnetic field. This structure may have 180°-, 120°-, and 60°-domain walls, which is associated with the effect of the projections (lying in the film plane) of six easy axes of the initial crystal of cubic anisotropy on the formation of the domain structure. It is virtually impossible to describe such a structure analytically; therefore, we simplify the problem of determining the spectrum of spin oscillations of the film. Suppose that, when $H = 0$, only 180°-domain walls exist in the film that separate alternating stripe domains of equal size and opposite magnetization. Such a situation may occur in a film that has a pronounced texture along one of six easy axes lying in the film plane. The analysis of the oscillation spectrum of such a domain structure shows that it consists of two branches; low- and high-frequency ones [10, 11]. If we consider a real domain structure of a film with regard to all types of domain walls, we can see that it consists of a large number of branches; this number is determined by the amount of different types of domains. Among these branches, only one branch will have the lowest frequency. It is this branch that should soften at the starting and end points of spin reorientation. It is obvious that the difference between the activations of low-frequency branches in the simplified 180°-domain structure and in a real domain structure is expressed by a certain numerical coefficient that will not influence the understanding of the physics of the phenomenon, i.e., the understanding of the behavior of the spin-wave oscillation branches in the film as a function of the magnetic field.

The energy density of the simplified domain structure in the film can be represented as follows:

$$
\begin{aligned}
F = \sum_{i=1}^{2} \Bigg\{ & \frac{1}{4}K_1\cos^2\vartheta_i \\
& + \frac{1}{8}K_4(\sin^2 2\vartheta_i + \sin^2 2\varphi_i \sin^4\vartheta_i) \\
& - \frac{1}{2}MH\cos\vartheta_i - \frac{1}{2}M(h_x\sin\vartheta_i\cos\varphi_i \\
& + h_y\sin\vartheta_i\sin\varphi_i + h_z\cos\vartheta_i) \\
& + \frac{1}{2}B_1(u_{xx}^{(i)}\sin^2\vartheta_i\cos^2\varphi_i
\end{aligned}
$$

$$\begin{aligned}
&+u_{yy}^{(i)}\sin^2\vartheta_i\sin^2\varphi_i+u_{zz}^{(i)}\cos^2\vartheta_i)\\
&+\frac{1}{2}B_2(u_{xy}^{(i)}\sin^2\vartheta_i\sin2\varphi_i\\
&+u_{xz}^{(i)}\sin2\vartheta_i\cos\varphi_i+u_{yz}^{(i)}\sin2\vartheta_i\sin\varphi_i)\\
&+\frac{1}{4}c_{11}(u_{xx}^{(i)2}+u_{yy}^{(i)2}+u_{zz}^{(i)2})\\
&+\frac{1}{2}c_{12}(u_{xx}^{(i)}u_{yy}^{(i)}+u_{zz}^{(i)}u_{yy}^{i}+u_{xx}^{(i)}u_{zz}^{(i)})\\
&\left.+c_{44}(u_{xy}^{(i)2}+u_{xz}^{(i)2}+u_{zy}^{(i)2})\right\}\\
&+\frac{1}{8}M^2[N_1(\sin\vartheta_1\cos\varphi_1+\sin\vartheta_2\cos\varphi_2)^2\\
&+N_2(\sin\vartheta_1\sin\varphi_1+\sin\vartheta_2\sin\varphi_2)^2\\
&+N_3(\cos\vartheta_1+\cos\vartheta_2)^2]\\
&+\frac{\pi}{2}M^2(\sin\vartheta_1\sin\varphi_1-\sin\vartheta_2\sin\varphi_2)^2.
\end{aligned}\tag{1}$$

Here, $K_1 > 0$ is the uniaxial anisotropy constant in the film plane, induced, for example, by the sample texture; K_4 is the first constant of cubic anisotropy; $\mathbf{h}$ is a microwave magnetic field; B_i are magnetostriction constants; c_{ij} are elasticity moduli; u_{ij} is the strain tensor; and N_i are demagnetization factors. The last two terms in the equation describe the energy of demagnetizing fields of the sample and the energy of domain walls [11].

Consider the ground state in which the polar ϑ_i and azimuthal φ_i angles of magnetization $\mathbf{M}$ are expressed by the formulas (axis z is perpendicular to the film plane and axes x and y lie in the film plane; axis y is perpendicular to the plane of domain walls)

$$\vartheta_{10}=\vartheta_{20}=\vartheta_0,\quad \varphi_{10}=0,\quad \varphi_{20}=\pi,\tag{2}$$

where the angle ϑ_0 is determined from the condition that energy (1) is minimal, which is expressed as

$$\begin{aligned}
MH&=K_1\cos\vartheta_0\\
&-2\tilde{K}_4\cos\vartheta_0\cos2\vartheta_0+M^2N_3\cos\vartheta_0,
\end{aligned}\tag{3}$$

where

$$\tilde{K}_4=K_4+\frac{B_1^2}{c_{11}-c_{12}}-\frac{B_2^2}{2c_{44}}.$$

At the starting point of reorientation, when $H = 0$, we have $\vartheta_0 = \pi/2$, while, at the end point of reorientation, we have $\vartheta_0 = 0$. The value of the magnetic field at which the reorientation is completed is given by

$$H_{\text{tr}}=\frac{K_1}{M}-\frac{2\tilde{K}_4}{M}+MN_3.\tag{4}$$

One can see that, in the absence of texture in the film plane and under suppressed cubic anisotropy, the transition field is determined by the demagnetizing field, i.e., by the shape anisotropy: $H_{\text{tr}} = N_3M$. When the film is infinite in the plane, $N_3 = 4\pi$.

The spectrum of spin oscillations in the simplified domain structure considered is determined from the system of Landau–Lifshits equations

$$\dot{\vartheta}_i=-\frac{g}{M\sin\vartheta_i}\frac{\partial F}{\partial\varphi_i},\quad \dot{\varphi}_i=\frac{g}{M\sin\vartheta_i}\frac{\partial F}{\partial\vartheta_i},\tag{5}$$

where g is the gyromagnetic ratio. When linearized near the equilibrium point (2), Eqs. (5) are rewritten as

$$\begin{aligned}
\dot{\vartheta}_+&=-\frac{g}{M}\sin\vartheta_0A\varphi_+,\quad \dot{\vartheta}_-=-\frac{g}{M}\sin\vartheta_0B\varphi_-,\\
\dot{\varphi}_+&=\frac{g}{M\sin\vartheta_0}C\vartheta_+,\quad \dot{\varphi}_-=\frac{g}{M\sin\vartheta_0}D\vartheta_-,
\end{aligned}\tag{6}$$

where

$$\vartheta_\pm=\frac{\vartheta_1\pm\vartheta_2}{2},\quad \varphi_\pm=\frac{\varphi_1\pm\varphi_2}{2},$$

$$\begin{aligned}
A&=\tilde{K}_4\sin^2\vartheta_0+\frac{B_2^2}{2c_{44}}+2\pi M^2,\\
B&=\tilde{K}_4\sin^2\vartheta_0+\frac{B_2^2}{2c_{44}}+\frac{1}{2}N_2M^2+2\pi M^2\sin^2\vartheta_0,\\
C&=-\frac{1}{2}K_1\cos2\vartheta_0+\tilde{K}_4\cos4\vartheta_0+\frac{1}{2}MH\cos\vartheta_0\\
&-\frac{1}{2}N_3M^2\cos2\vartheta_0+\frac{B_2^2}{2c_{44}}\cos^22\vartheta_0\\
&+\frac{B_1^2}{2(c_{11}-c_{12})}\sin^22\vartheta_0+2\pi M^2\sin^2\vartheta_0,\\
D&=-\frac{1}{2}K_1\cos2\vartheta_0+\tilde{K}_4\cos4\vartheta_0+\frac{1}{2}MH\cos\vartheta_0\\
&-\frac{1}{2}N_3M^2\cos2\vartheta_0+\frac{1}{2}N_1M^2\cos^2\vartheta_0+\frac{B_2^2}{2c_{44}}\cos^22\vartheta_0\\
&+\frac{B_1^2}{c_{11}-c_{12}}\sin^22\vartheta_0+2\pi M^2\sin^2\vartheta_0.
\end{aligned}\tag{7}$$

The last terms in the expressions for B, C, and D are attributed to the interaction between the spin and electromagnetic waves (a dipole–dipole interaction) [1, 2] (these terms arise from the simultaneous solution of Eqs. (5) and the Maxwell equations).

Taking into account (7), one can easily derive from (6) the spectrum of harmonic oscillations of the simplified domain structure under consideration:

$$\omega_1 = \sqrt{AC}, \quad \omega_2 = \sqrt{BD}. \tag{8}$$

The frequency ω_1 corresponds to in-phase oscillations of magnetizations in adjacent domains, while the frequency ω_2 corresponds to their out-of-phase oscillations.

Note that, when deriving expressions (8) for the frequencies of spin waves, we did not take into account the interaction between these spin waves and the relaxation oscillations of magnetization in the domains. It is known that, in contrast to REOFs [1, 2], the consideration of the relaxation oscillations of magnetization in the case of a ferromagnetic material results in a negligible contribution to the activation of the frequency of precession as compared with the contributions of interactions of precession oscillations with other oscillatory subsystems (elastic and dipolar ones) [12].

4. DISCUSSION OF THE RESULTS

Formula (8) allows one to determine the values of spin-wave activations at the starting and end points of reorientation. At the end point of reorientation (4), which corresponds to the OPT-2 from an inhomogeneous domain state with the magnetization in the domains directed at an angle to the normal to the film plane to a homogeneous state with the magnetization perpendicular to the film plane, frequencies (8) can be rewritten as follows:

$$\omega_1 = gM\sqrt{\left(2\pi + \frac{B_2^2}{2M_0^2c_{44}}\right)\frac{B_2^2}{2M_0^2c_{44}}},$$

$$\omega_2 = gM \tag{9}$$

$$\omega_2 = gM\sqrt{\left(\frac{1}{2}N_2 + \frac{B_2^2}{2M_0^2c_{44}}\right)\left(\frac{1}{2}N_1 + \frac{B_2^2}{2M_0^2c_{44}} + 2\pi\right)}.$$

At the starting point of reorientation $H = 0$, which corresponds to the transition from the state with the domain structure such that the domain magnetizations lie in the film plane to the state with the domain structure where the domain magnetizations make an angle with the film plane, frequencies (8) are given by

$$\begin{aligned}\omega_1 &= \frac{g}{M}\sqrt{\left(\tilde{K}_4 + 2\pi M^2 + \frac{B_2^2}{2c_{44}}\right)\left(\frac{1}{2}K_1 + \tilde{K}_4 + \frac{1}{2}N_3M^2 + \frac{B_2^2}{2c_{44}} + 2\pi M^2\right)},\\ \omega_2 &= \frac{g}{M}\sqrt{\left(\tilde{K}_4 + \frac{1}{2}N_2M^2 + \frac{B_2^2}{2c_{44}} + 2\pi M^2\right)\left(\frac{1}{2}K_1 + \tilde{K}_4 + \frac{B_2^2}{2c_{44}}\right)}.\end{aligned} \tag{10}$$

To evaluate the activations of spin waves at the starting and end points of reorientation, we use the following values of the film parameters, which are characteristic of garnet ferrites:

$$g \approx 2\times10^7\ 1/(\mathrm{s\ Oe}), \quad M \approx 100\ \mathrm{Oe}, \quad c_{44} \approx 10^{11}\ \mathrm{erg/cm^3},$$

$$B_2 = 10^7\ \mathrm{erg/cm^3}, \quad \tilde{K}_4 \approx 5\times10^3\ \mathrm{erg/cm^3}.$$

We will neglect the texture in the film plane ($K_1 \approx 0$). The demagnetization factors of the plane are estimated by the formulas for an oblate ellipsoid of rotation [13]:

$$N_1 \approx N_2 \approx 2\times10^{-4}, \quad N_3 \approx 1.$$

At the end point of reorientation, the activations of spin waves are given by

$$\nu_1 = \frac{\omega_1}{2\pi} \approx 180\ \mathrm{MHz}, \quad \nu_2 = \frac{\omega_2}{2\pi} \approx 180\ \mathrm{MHz}, \tag{11}$$

whereas, at the starting point of reorientation, they are given by

$$\nu_1 = \frac{\omega_1}{2\pi} \approx 3000\ \mathrm{MHz}, \quad \nu_2 = \frac{\omega_2}{2\pi} \approx 580\ \mathrm{MHz}. \tag{12}$$

It follows from (11) and (12) that, at the end point of reorientation, the experimental value of the spin-wave frequency ($\nu_0 \approx 200$ MHz; Fig. 2) is in good agreement with the activations ν_1 and ν_2 of both branches of magnetization oscillations in the domains, whereas, at the starting point of reorientation, it agrees with the activation ν_2 of out-of-phase oscillations of magnetizations in the domains. Such a small difference between the activations of in-phase and out-of-phase oscillations of magnetizations in the domains at the end point of reorientation is associated with the small value of the demagnetization factors in the film plane.

Thus, the experimentally observed behavior of spin-wave frequencies in a garnet-ferrite film can be attributed to the softening of out-of-phase and in-phase oscillations of magnetizations in the domains at the starting and end points of reorientation of activations.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Education of the Russian Federation, grant no. E00-3.4-536.

REFERENCES

1. V. D. Buchel'nikov, N. K. Dan'shin, L. T. Tsymbal, and V. G. Shavrov, Usp. Fiz. Nauk **166**, 585 (1996) [Phys. Usp. **39**, 547 (1996)].
2. V. D. Buchel'nikov, N. K. Dan'shin, L. T. Tsymbal, and V. G. Shavrov, Usp. Fiz. Nauk **169**, 1049 (1999).
3. G. S. Krinchik, *Physics of Magnetic Phenomena* (Mosk. Gos. Univ., Moscow, 1985).
4. A. M. Balbashov, F. V. Lisovskiĭ, V. K. Raev, *et al.*, *Elements and Facilities on Bubbles: Handbook*, Ed. by N. N. Evtikhiev and B. N. Naumov (Radio i Svyaz', Moscow, 1987).
5. E. A. Turov, *Physical Properties of Magnetically Ordered Crystals* (Izd. Akad. Nauk SSSR, Moscow, 1963).
6. J. Maartense, J. Appl. Phys. **53**, 2466 (1982).
7. A. M. Balbashov, A. G. Berezin, Yu. M. Gufan, *et al.*, Zh. Éksp. Teor. Fiz. **93**, 302 (1987) [Sov. Phys. JETP **66**, 174 (1987)].
8. E. A. Turov and V. G. Shavrov, Usp. Fiz. Nauk **140**, 429 (1983) [Sov. Phys. Usp. **26**, 593 (1983)].
9. A. M. Balbashov, Yu. M. Gufan, P. Yu. Marchukov, and E. G. Rudashevskiĭ, Zh. Éksp. Teor. Fiz. **94** (4), 305 (1988) [Sov. Phys. JETP **67**, 821 (1988)].
10. J. O. Artman, Phys. Rev. **105**, 62 (1957).
11. A. G. Gurevich and G. A. Melkov, *Magnetic Oscillations and Waves* (Nauka, Moscow, 1994).
12. V. D. Buchel'nikov and V. G. Shavrov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 534 (1994) [JETP Lett. **60**, 548 (1994)].
13. *Physical Encyclopedia* (Bol'shaya Ross. Éntsiklopediya, Moscow, 1994), Vol. 4, p. 242.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 11–25.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 17–32.
Original English Text Copyright © 2002 by Butkevich, Mikheyev.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Cross Section of the Muon-Nuclear Inelastic Interaction¶

A. V. Butkevich* and S. P. Mikheyev**
Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia
**e-mail: butkevic@al20.inr.troitsk.ru*
***e-mail: mikheyev@pcbai10.inr.ruhep.ru*

Received February 6, 2002

Abstract—It is shown that the combination of the structure functions F_2 predicted by the CKMT model at low and moderate values of Q^2 and the MRS99 parton distribution functions at high Q^2 gives a good description of the data over the complete measured region of x and Q^2. Using these structure functions, the main characteristics of the muon-nuclear inelastic scattering are calculated. Nuclear effects and contributions of the neutral current and the γ–Z interference are taken into account. © *2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Muon inelastic scattering off nuclei plays an important role in muon propagation through matter. In this process, the muon can lose a significant part of its energy and can be scattered at large angles. Therefore, the muon-nuclear inelastic scattering is of interest in numerous applications related to the muon transport in matter, in particular, in calculations of the muon intensity at large depth of matter, the muon-induced hadron flux underground, the background produced by atmospheric muons in underground neutrino experiments, etc.

Several models [1–4] have been developed to describe the muon-nuclear inelastic interaction; however, uncertainties of this process are much larger than for purely electromagnetic interactions. The reason is that the bulk of this process is characterized by a low squared four-momentum transfer Q^2. The smallness of Q^2 does not allow us to use the perturbative QCD (pQCD) to calculate of the nuclear structure function, and phenomenological models such as the Regge or General Vector Dominance Model (GVDM) must be used. The parametrization of the nucleon structure functions obtained in these models depends on free parameters that can be determined from a fit of experimental data and can be applied in the range $Q^2 \leq 1$–$3\ \mathrm{GeV}^2$. This range is often referred to as photoproduction. But these models fail to describe deep inelastic scattering (DIS) data at high Q^2. The pQCD (NLO QCD) gives a good description of the structure functions at $Q^2 \geq 3\ \mathrm{GeV}^2$. A model combining various aspects of these approaches is therefore needed to describe the Q^2 behavior of nucleon structure functions over the entire range from photoproduction to DIS.

The widely used approximation [4] of the muon photonuclear cross section was obtained twenty years ago in the GVDM framework. Experimental data for $Q^2 \leq 100\ \mathrm{GeV}^2$ and $x \geq 0.01$ were used to determine the parameters. Recently, precise data [5] on structure functions in wide ranges of Q^2 ($0.045 \leq Q^2 \leq 10^4\ \mathrm{GeV}^2$) and x ($10^{-6} \leq x \leq 0.98$) have been obtained, and new nuclear effects (antishadowing and EMC effects) were observed.

The main goal of this paper is the calculation of the muon-nuclear inelastic cross section based on the modern nucleon structure functions and on the present knowledge of nuclear effects.

The paper is organized as follows. In Section 2, we give the general relations and definitions used in describing neutral current charged lepton–nucleon scattering. The procedure for calculating the nucleon structure function using the CKMT Regge model [6] and the MRS99 parton distribution function [7] is described in Section 3. In Section 4, nuclear effects and their parametrization are described. The total cross section, the muon energy loss, and the angular distributions of scattered muons are given in Section 5. In the Conclusions, we summarize the main results of the paper.

2. NEUTRAL CURRENT CHARGED LEPTON–NUCLEON SCATTERING CROSS SECTION

The cross section of the neutral current charged lepton scattering off a nucleon,

$$l(k) + N(p) \longrightarrow l(k') + X(p'), \qquad (1)$$

is given by the sum of contributions of the processes shown in Fig. 1 Here, $k(E, \mathbf{k})$ and $k'(E', \mathbf{k}')$ are the initial and final lepton four-momenta; $q = k - k'$ is the virtual photon or Z-boson momentum; and p and p' are the initial nucleon momentum and the total momentum of the final hadrons X, respectively. This process can be described by the transferred four-momentum Q^2, the

---

¶This article was submitted by the authors in English.

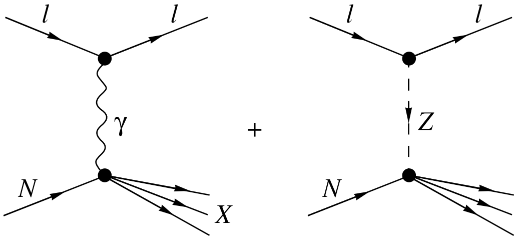


Fig. 1. Schematic diagrams for the neutral current charged lepton scattering off a nucleon.

Bjorken variable x, and the lepton energy loss ν (or inelasticity y) defined as

$$Q^2 = -q^2 = (k-k')^2,$$
$$x = \frac{Q^2}{2pq}, \quad \nu = \frac{qp}{M}, \quad y = \frac{pq}{pk}. \tag{2}$$

In the laboratory system

$$Q^2 = 2(EE' - \mathbf{k}\cdot\mathbf{k}') - 2m^2,$$
$$x = \frac{Q^2}{2M\nu}, \quad \nu = E - E', \quad y = \frac{\nu}{E}, \tag{3}$$

where M and m are the nucleon and lepton masses, respectively.

The general form of the differential cross section for the scattering of a charged nonpolarized lepton on a nonpolarized nucleon, summed over the final lepton polarizations, can be expressed as

$$\frac{d^2\sigma^{l^-,l^+}}{d\nu dQ^2} = \frac{2\pi\alpha^2}{Q^4E^2}$$
$$\times[E^{l^-,l^+}(x,Q^2) - I^{l^-,l^+}(x,Q^2) + Z^{l^-,l^+}(x,Q^2)], \tag{4}$$

where

$$E^{l^-,l^+} = 2xF_1^{\mathrm{el}}(x,Q^2)Y_1 + F_2^{\mathrm{el}}(x,Q^2)Y_2, \tag{5}$$

$$I^{l^-,l^+} = P_Z\{g_V[2xF_1^I(x,Q^2)Y_1 + F_2^I(x,Q^2)Y_2]$$
$$\mp g_A xF_3^I(x,Q^2)Y_3\}, \tag{6}$$

$$Z^{l^-,l^+} = P_Z^2\{(g_V^2+g_A^2)(2xF_1^Z(x,Q^2)Y_1$$
$$+ F_2^Z(x,Q^2)Y_2) \mp 2g_Vg_A xF_3^Z(x,Q^2)Y_3\}, \tag{7}$$

and

$$Y_1 = (Q^2-2m^2)\frac{\nu}{Q^2},$$
$$Y_2 = \left[2E(E-\nu) - \frac{Q^2}{2}\right]\frac{1}{\nu}, \tag{8}$$
$$Y_3 = (E-\nu).$$

Here, the term P_Z accounts for the Z^0 propagator,

$$P_Z = \frac{G}{\sqrt{2}}\frac{Q^2}{2\pi\alpha}\frac{m_Z^2}{Q^2+m_Z^2}, \tag{9}$$

where $G/\sqrt{2}$ is the Fermi constant, $\alpha = 1/137$ is the fine structure constant, and m_Z is Z-boson mass. The lepton weak coupling constants g_V and g_A are

$$g_V = -\frac{1}{2} + 2\sin^2\theta_W, \quad g_A = -\frac{1}{2}, \tag{10}$$

where θ_W is the Weinberg angle.

The functions $F_i^{\mathrm{el},Z}$ are the electromagnetic (γ-exchange) and neutral current (Z-exchange) structure functions, respectively. The functions F_i^I correspond to the decomposition over invariant functions of the tensor (γ–Z interference)

$$W^I \approx \sum\{\langle p'|J_\alpha^{\mathrm{el}}|p\rangle\langle p|J_\beta^Z|p'\rangle$$
$$+\langle p'|J_\alpha^Z|p\rangle\langle p|J_\beta^{\mathrm{el}}|p'\rangle\}\delta(p'-p-q). \tag{11}$$

The upper sign in Eqs. (6) and (7) corresponds to the lepton scattering (e^-, μ^-) and the lower sign is for the antilepton (e^+, μ^+) scattering.

The term proportional to the function F_3^I is due to the interference between the electromagnetic scattering amplitude and the axial-vector current weak-interaction amplitude. The amplitudes have opposite C parities, and the corresponding terms therefore have opposite signs for the lepton and the antilepton scattering. At low Q^2, the γ–Z interference term is much smaller than the γ-exchange one, but it increases linearly with Q^2 (Eq. (9)) and becomes comparable to the γ-exchange term at $Q^2 \approx 10^3$ GeV^2.

In terms of the parton distribution in the LO approximation, the structure functions can be written as

$$F_1^{\mathrm{el}} = \frac{1}{2}\sum e_q^2(f_q + f_{\bar{q}}), \tag{12}$$

$$F_2^{\mathrm{el}} = x\sum e_q^2(f_q + f_{\bar{q}}), \tag{13}$$

$$F_1^Z = \frac{1}{2}\sum(v_q^2 + a_q^2)(f_q + f_{\bar{q}}), \tag{14}$$

$$F_2^Z = x\sum(v_q^2 + a_q^2)(f_q + f_{\bar{q}}), \tag{15}$$

$$F_3^Z = 2\sum v_q a_q(f_q - f_{\bar{q}}), \tag{16}$$

$$F_1^I = \sum e_q v_q(f_q + f_{\bar{q}}), \tag{17}$$

$$F_2^I = 2x\sum e_q v_q(f_q + f_{\bar{q}}), \tag{18}$$

$$F_3^I = 2\sum e_q a_q(f_q - f_{\bar{q}}). \tag{19}$$

Here, f_q and $f_{\bar{q}}$ are parton distribution functions in the proton; e_q, v_q, and a_q are the charge, vector, and axial-vector week couplings of quarks. For the up-quarks (u, c, t), they are given by

$$e_{u,c,t} = \frac{2}{3}, \quad v_{u,c,t} = \frac{1}{2} - \frac{4}{3}\sin^2\theta_W, \quad a_{u,c,t} = \frac{1}{2} \tag{20}$$

and for the down-quarks (d, s, b), they are

$$\begin{gathered} e_{d,s,b} = -\frac{1}{3}, \quad v_{d,s,b} = -\frac{1}{2} + \frac{2}{3}\sin^2\theta_W, \\ a_{d,s,b} = -\frac{1}{2}. \end{gathered} \tag{21}$$

It can be seen from Eqs. (4) and (9) that the main contribution to the total cross section is due to photoproduction (a low-Q^2 process); however, at a fixed outgoing muon energy, the large scattering angle corresponds to high Q^2,

$$\cos\theta = (EE' - Q^2/2 - m^2)/|\mathbf{k}||\mathbf{k}'|. \tag{22}$$

Therefore, the calculation of the muon scattering at large angles requires knowledge of the behavior of the nucleon structure functions in the wide range of $Q^2 \approx$ 0.01–10^6 GeV^2.

3. LOW- AND HIGH-Q^2 APPROXIMATIONS OF THE NUCLEON STRUCTURE FUNCTIONS

At high Q^2, the QCD predictions for the nucleon structure functions are obtained by solving the DGLAP evolution equations at the NLO approximation in the $\overline{\text{MS}}$ or DIS schemes. These equations yield the parton distribution functions at all values of Q^2 provided these functions are given as a function of x at some input scale Q_0^2 = 1.2–5 GeV^2. The latest global fits performed by several groups (MRS99 [7], GRV98 [8], and CTEQ5 [9]) give a good description of the experimental data. At Q^2 below Q_0^2, the perturbative QCD fails to describe data and phenomenological nonperturbative (GVDM or Regge) models are required. A considerable number of nonperturbative models have been developed recently [10–12]. These models predict the correct limit of F_2 at $Q^2 = 0$ and give a good description of the structure functions at low and medium Q^2. Thus, neither the nonperturbative approaches nor pQCD can be expected to describe the Q^2 behavior of the structure functions over the entire range from photoproduction to DIS. A number of models combining QCD and phenomenological approaches have been developed to describe data in the transition region of Q^2 (see review [13]). In this paper, we use the CKMT model [6] at low and moderate Q^2 and the MRS99 fit of the parton distribution function [7] at high Q^2.

The CKMT model proposes the following parametrization of the proton structure function F_2^p:

$$F_2^p(x, Q^2) = F_S^p(x, Q^2) + F_{NS}^p(x, Q^2). \tag{23}$$

The singlet term

$$F_S^p(x, Q^2) = A_S x^{-\Delta(Q^2)}(1-x)^{n(Q^2)+4}\left(\frac{Q^2}{Q^2+a}\right)^{1+\Delta(Q^2)} \tag{24}$$

corresponds to the Pomeron contribution that determines the small-x behavior of sea quarks and gluons. The dependence of the effective intercept of the Pomeron Δ on Q^2 is parameterized as

$$\Delta(Q^2) = \Delta_0\left(1 + \frac{2Q^2}{Q^2+d}\right). \tag{25}$$

The $x \longrightarrow 1$ behavior of $F_S(x, Q^2)$ is determined by the function

$$n(Q^2) = \frac{3}{2}\left(1 + \frac{Q^2}{Q^2+c}\right). \tag{26}$$

The parameterization for the nonsinglet term which corresponds to the secondary (f, A_2) Reggion (valence quark) contribution, is given by

$$F_{NS}^p(x, Q^2) = Bx^{(1-\alpha_R)}(1-x)^{n(Q^2)}\left(\frac{Q^2}{Q^2+b}\right)^{\alpha_R}, \tag{27}$$

where the behavior as $x \longrightarrow 0$ is determined by the secondary Reggeon intercept α_R. The valence quark distribution can be separated into contributions of the u and d valence quarks by replacing

$$F_{NS}^p(x, Q^2) = xU_V(x, Q^2) + xD_V(x, Q^2), \tag{28}$$

where

$$xU_V(x, Q^2) = B_u x^{(1-\alpha_R)}(1-x)^{n(Q^2)}\left(\frac{Q^2}{Q^2+b}\right)^{\alpha_R}, \tag{29}$$

$$xD_V(x, Q^2) = B_d x^{(1-\alpha_R)}(1-x)^{n(Q^2)+1}\left(\frac{Q^2}{Q^2+b}\right)^{\alpha_R}, \tag{30}$$

and B_u and B_d are fixed at $Q^2 = Q_0^2$ by the normalization conditions for valence quarks in a proton:

$$\begin{gathered} \int_0^1 \frac{1}{x}[xU_V(x, Q^2)]dx = 2e_u^2, \\ \int_0^1 \frac{1}{x}[xD_V(x, Q^2)]dx = e_d^2. \end{gathered} \tag{31}$$

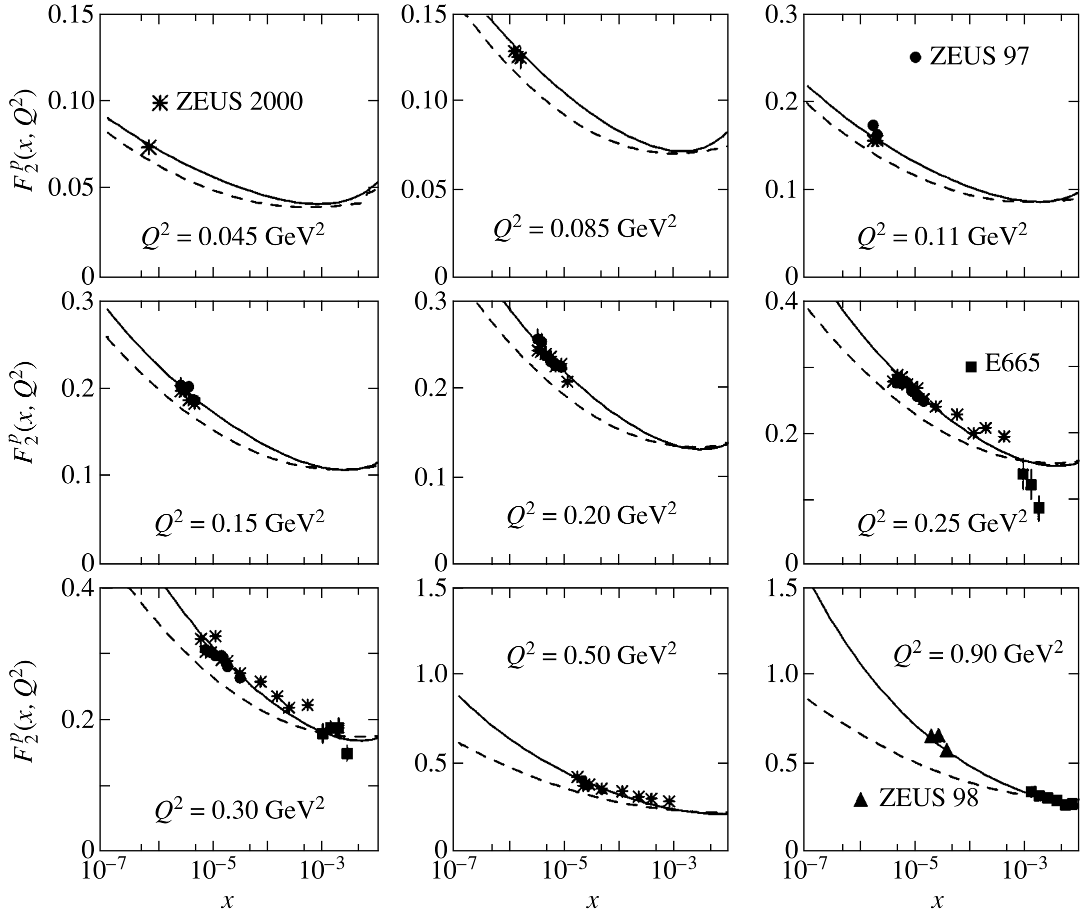


Fig. 2. The proton structure function F_2^p at low Q^2 versus x. Solid curves are our calculations using the CKMT model and dashed curves are the calculations of Bezrukov and Bugaev [4]. Points are experimental results [5, 15].

The limit of $Q^2 = 0$ corresponds to the interaction of real photons. The total cross section for real photons can be written as

$$\sigma^{\mathrm{tot}}_{\gamma p}(\nu) = \left[\frac{4\pi^2\alpha}{Q^2}F_2(x, Q^2)\right]_{Q^2=0}. \qquad (32)$$

We see from Eqs. (23), (24), and (28) that $F_2 \approx Q^2$ as $Q^2 \longrightarrow 0$ for a fixed ν. Thus, the parametrization

$$\sigma^{\mathrm{tot}}_{\gamma p}(\nu) = 4\pi^2\alpha[A_s a^{-1-\Delta_0}(2M\nu)^{\Delta_0} + (B_u + B_d)b^{-\alpha_R}(2M\nu)^{\alpha_R - 1}] \qquad (33)$$

applies in the CKMT model.

In this way, we find parametrizations of both the F_2^p and γp cross sections with seven free parameters: a, b, c, d, Δ_0, α_R, and A_S. To determine the parameters, we have made a joint fit of the $\sigma^{\mathrm{tot}}_{\gamma p}$ data and the NMC, E665, SLAG, ZEUS, and H1 data on the proton structure function F_2 in the regions $0.11 \le Q^2 \le 5.5$ GeV^2 and $10^{-6} \le x \le 0.98$ [5]. As the initial condition for the values of different parameters, we used those obtained in the previous fit in [6]. A global fit results in the following values of the parameters (all dimensional parameters are expressed in GeV^2):

$$a = 0.2513, \quad b = 0.6186, \quad c = 3.0291, \quad d = 1.4817,$$
$$\Delta_0 = 0.0988, \quad \alpha_R = 0.4056, \quad A_S = 0.12.$$

The values of the parameters $B_u = 1.2437$ and $B_d = 0.1853$ were determined from the normalization conditions for valence quarks (at $Q_0^2 = 2$ GeV^2). The quality of the description of all experimental data is very good and χ^2/d.o.f. = 754.8/600, where only the statistical errors have been used. Recently, a modified version of the CKMT model with the new data on F_2^p at low Q^2 was published [14]. The values of the main parameters are in a good agreement with those obtained in the present work.

To calculate F_2 in the entire region of Q^2, we therefore use the CKMT model at $Q^2 \le 5$ GeV^2, the MRS99 parton distribution function at $Q^2 \ge 6$ GeV^2, and a linear

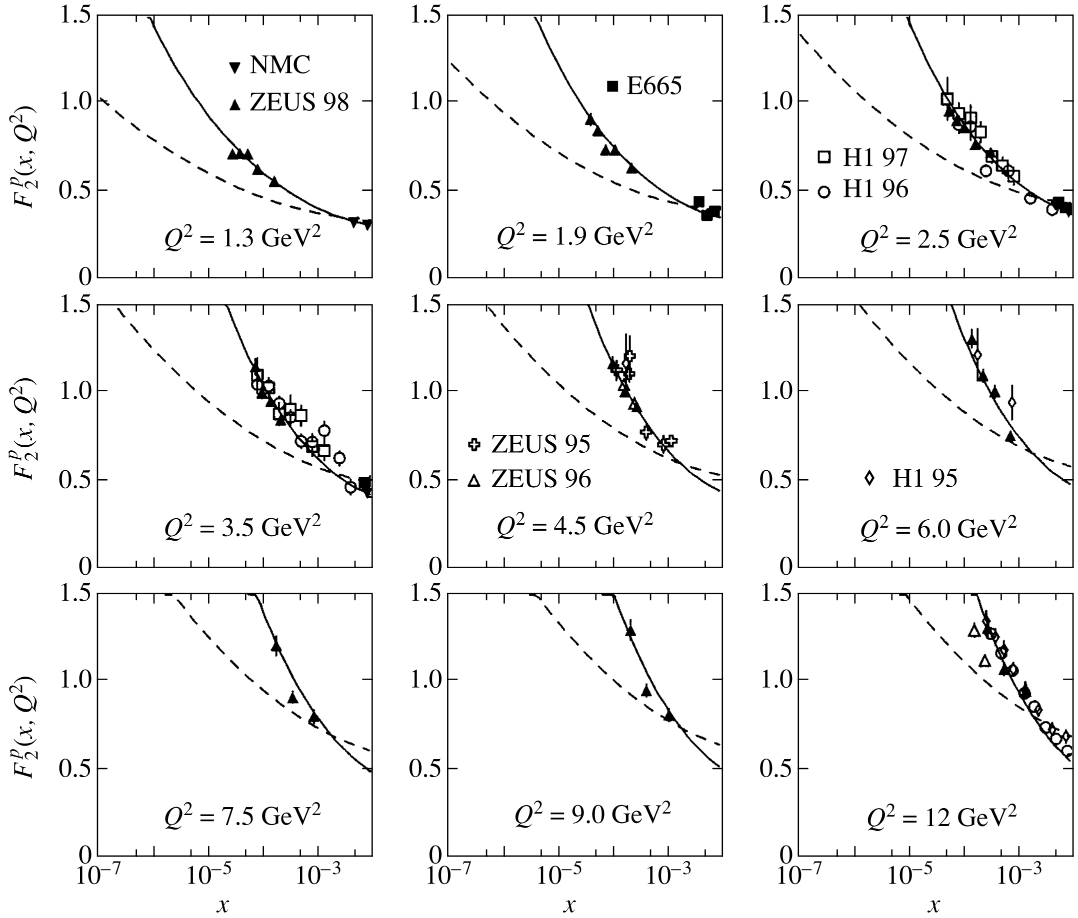


Fig. 3. The same as in Fig. 2, but at moderate Q^2.

fit between F_2^p (CKMT) and F_2^p (MRS99) in the region $5 \le Q^2 \le 6$ GeV^2. The result of the fit of F_2^p and $\sigma_{\gamma p}^{\text{tot}}$ is shown in Figs. 2 and 3 (F_2^p versus x for different values of Q^2) and Fig. 4 (F_2^p versus Q^2 for different values of x) along with the experimental data [5, 15]. The cross section $\sigma_{\gamma p}^{\text{tot}}$ as a function of $W^2 = M^2 + 2M\nu - Q^2$ is shown in Fig. 5 (the data from [16, 17]).

A good description of experimental data is obtained for all x and Q^2 values. We note that

(a) the recent ZEUS BPT97 data [15] were not included in our fit, but are in agreement with the CKMT model prediction at $Q^2 \le 0.1$ GeV^2;

(b) the rise of F_2^p at low x and low Q^2 is well described by the CKMT model with the slope $\Delta_0 = 0.0988$, while the experimental value is 0.102 ± 0.070 [15];

(c) the $\sigma_{\gamma p}^{\text{tot}}$ values found by the ZEUS collaboration are the result of a phenomenologically motivated extrapolation.

In Figs. 2–5, we show the structure function F_2^p and $\sigma_{\gamma p}^{\text{tot}}$ that were obtained by Bezrukov and Bugaev [4] and were used for calculating the muon photonuclear cross section are also shown. At $x < 10^{-3}$, the structure functions rise slower than the present data indicate. On the other hand, in the region $x > 0.01$ and $Q^2 > 5$ GeV^2, the structure functions are overestimated.

The CKMT parameterization gives separate contributions of valence quarks, sea quarks, and gluons. We used this peculiarity for parameterizing of the neutron structure function F_2^n that can be extracted from the deuteron F_2^d and the proton F_2^p data using the relation

$$F_2^d = \frac{1}{2}[F_2^p(x) + F_2^n(x)] \tag{34}$$

and the Gottfried sum rule

$$\begin{aligned} S_G &= \int_0^1 (F_2^p - F_2^n)\frac{dx}{x} \\ &= \frac{1}{3}\int_0^1 (d_V - u_V)dx - \frac{2}{3}\int_0^1 (\bar{d} - \bar{u})dx. \end{aligned} \tag{35}$$

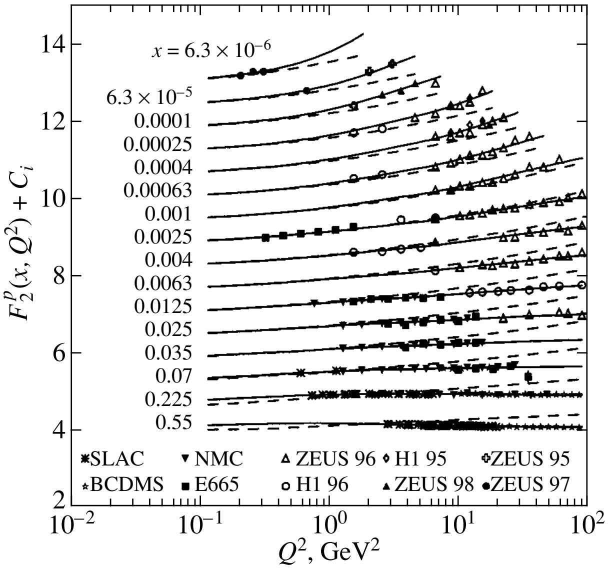


Fig. 4. The proton structure function F_2^p versus Q^2 at fixed values of x. The solid curves were obtained using the CKMT + MRS99 model, and the dashed curves are the results from [4]. Data set is due to experiments [5]. For clarity, the quantity $C_i = 13.6–0.6i$ is added to F_2^p, where $i = 1$ (16) for the lowest (highest) x value.

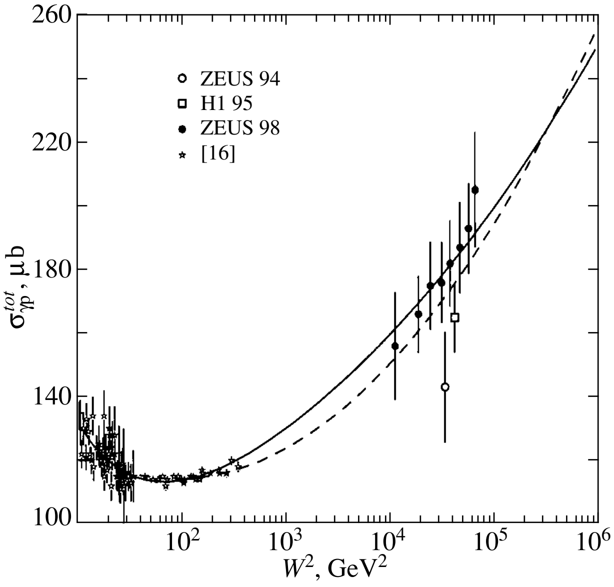


Fig. 5. The total cross section $\sigma_{\gamma p}^{\text{tot}}$ as a function of W^2. The solid curve is the result of our calculations using the combined CKMT + MRS99 model and the dashed curve is taken from [4]. Experimental data are taken from [16] at low energies and from ZEUS and H1 [17] at higher energies.

In the case of the SU(2)-symmetric sea, the parton distribution function of $\bar{d}$ is equal to that of $\bar{u}$, and therefore, $S_G = 1/3$. But the NMC collaboration [5] gives $S_G = 0.235 \pm 0.026$ at $Q^2 = 4$ GeV^2, which is significantly below 1/3 and shows that $F_2^p(x) - F_2^n(x) \longrightarrow 0$ and $F_2^n(x)/F_2^p(x) \longrightarrow 1$ as $x \longrightarrow 0$. Taking these results into account, the singlet term of F_2^n must be modified. Because of the isotopical invariance of the strong interaction, the nonsinglet term F_{NS}^n is

$$F_{NS}^n(x, Q^2) = \frac{1}{4}xU_V(x, Q^2) + 4xD_V(x, Q^2), \qquad (36)$$

where $xU_V(x, Q^2)$ and $xD_V(x, Q^2)$ are given by Eqs. (29) and (30). The singlet term

$$F_S^x(x, Q^2) = A_S x^{-\Delta(Q^2)}(1-x)^{n(Q^2)+\tau}\left(\frac{Q^2}{Q^2+a}\right)^{1+\Delta(Q^2)} \qquad (37)$$

involves an additional free parameter τ. The value of this parameter obtained from the fit of the F_2^d data [5] in the region $Q^2 \le 5$ GeV^2 (with all the other parameters fixed by the fit of F_2^p and $\sigma_{\gamma p}^{\text{tot}}$) is $\tau = 1.8152$. The quality of the data description is good with the value of χ^2/d.o.f. = 611.1/453 for the structure function F_2^d data and χ^2/d.o.f. = 452.8/380 for the F_2^n/F_2^p data, where only statistical errors have been used.

To calculate the structure function F_2^n in the entire region of Q^2, we used the approximation in Eqs. (36) and (37) at $Q^2 \le 5$ GeV^2, the MRS99 parton distribution function at $Q^2 \ge 6$ GeV^2, and a linear fit between F_2^n(CKMT) and F_2^n(MRS) in the transition region $5 < Q^2 < 6$ GeV^2. The structure function F_2^d (Fig. 6) and the ratio F_2^n/F_2^p (Fig. 7) versus Q^2 for different values of x are shown along with experimental data. Figure 8 shows $F_2^p - F_2^n$ versus x at $Q^2 = 4$ GeV^2. The calculations are in agreement with the NMC data [18].

For the calculation of the cross section of lepton–nucleon scattering, the behavior of the structure function $2xF_1$ must be known in a wide range of Q^2 and x. This function can be expressed using the longitudinal structure function

$$F_L = \left(1 + \frac{4M^2x^2}{Q^2}\right)F_2 - 2xF_1. \qquad (38)$$

Then,

$$2xF_1 = \frac{1}{1+R}\left(1 + \frac{4M^2x^2}{Q^2}\right)F_2, \qquad (39)$$

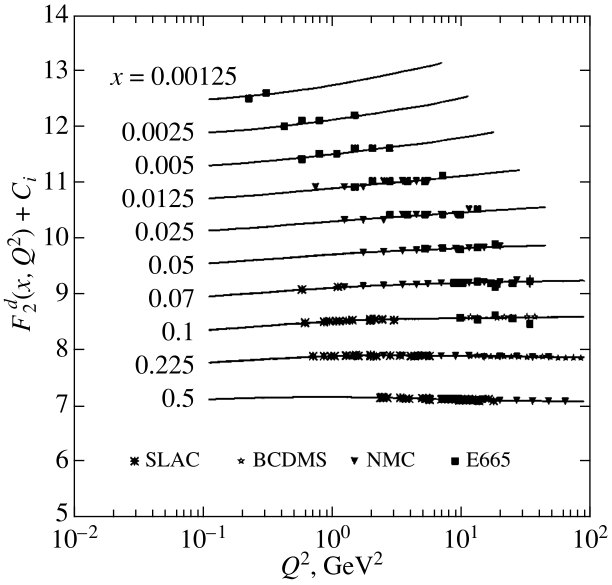


Fig. 6. The deutron structure function F_2^d versus Q^2 at fixed values of x. Data set is due to BCDMS, E665, NMC, and SLAC experiments [5]. For clarity, the quantity $C_i = 13.6–0.6i$ is added to F_2^d, where $i = 1$ (10) for the lowest (highest) x value.

where

$$R \equiv \frac{\sigma_L}{\sigma_T} = \frac{F_L}{(1 + 4M^2x^2/Q^2)F_2 - F_L}. \tag{40}$$

Perturbative QCD describes the available data on the ratio $R(x, Q^2)$ at large values of Q^2 reasonably well, but very little is known about possible extrapolations towards the region of low Q^2. In the limit as $Q^2 \longrightarrow 0$, the structure function F_L must vanish as Q^4 (for fixed ν), and therefore, $R \approx Q^2$. At $x < 0.01$ and $Q^2 < 0.5$ GeV^2, the experimental results are poor. The data show a small value of R at moderate values of x and a possible increase in R as x decreases. The data come from experiments carried out with different targets (nuclei and protons), and differences $R^A - R^p$ are consistent with zero and do not exhibit any significant dependence on x [19].

Data can be fitted by the parametrization R(SLAC98) [20]. But this fit must not be used at $Q^2 < 0.35$ GeV^2. In this region, we used the GVDM asymptotic form of R as $Q^2 \longrightarrow 0$ given by [21]

$$R_{GVDM}(Q^2, x) \approx \frac{Q^2}{Q^2 + m_\rho^2},$$

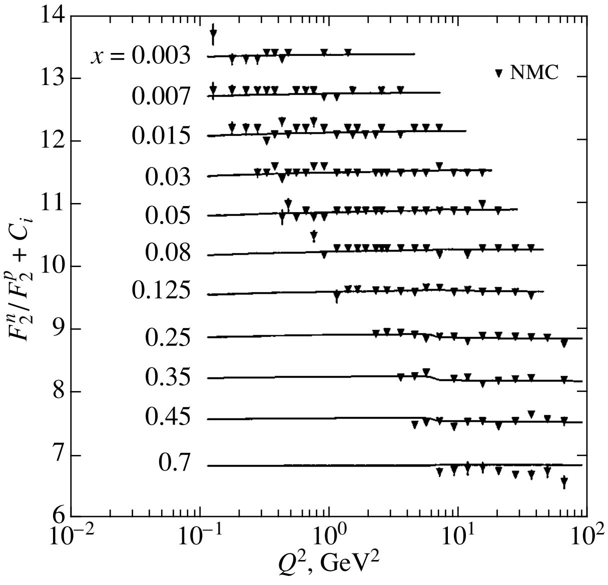


Fig. 7. The ratio of the structure functions F_2^n/F_2^p versus Q^2 at fixed values of x, as compared to NMC data [5]. For clarity, the quantity $C_i = 13.6–0.6i$ is added to F_2^n/F_2^p, where $i = 1$ (11) for the lowest (highest) x value.

where $m_\rho = 0.77$ GeV is the ρ-meson mass. At $Q^2 > Q_0^2 = 1.4$ GeV^2, the function $R(x, Q^2)$ is therefore calculated as

$$R(x, Q^2) = \begin{cases} R(\text{MRS99}) & \text{at} \quad x < 10^{-3}, \\ R(\text{SLAC98}) & \text{at} \quad x \ge 5 \times 10^{-3}. \end{cases} \tag{41}$$

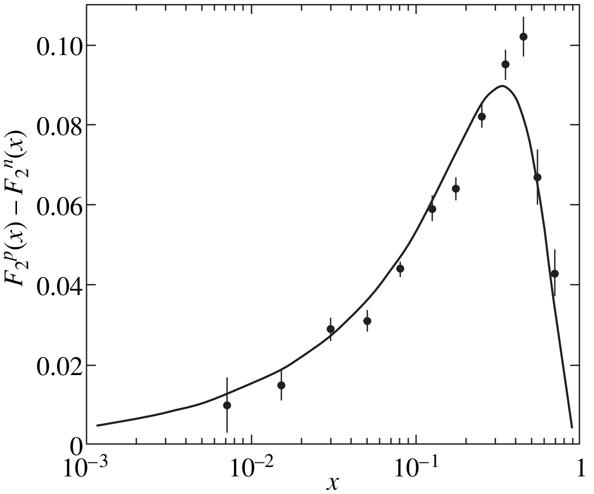


Fig. 8. The difference $F_2^p - F_2^n$ at $Q^2 = 4$ GeV^2 as a function of x, as compared to NMC data [5].

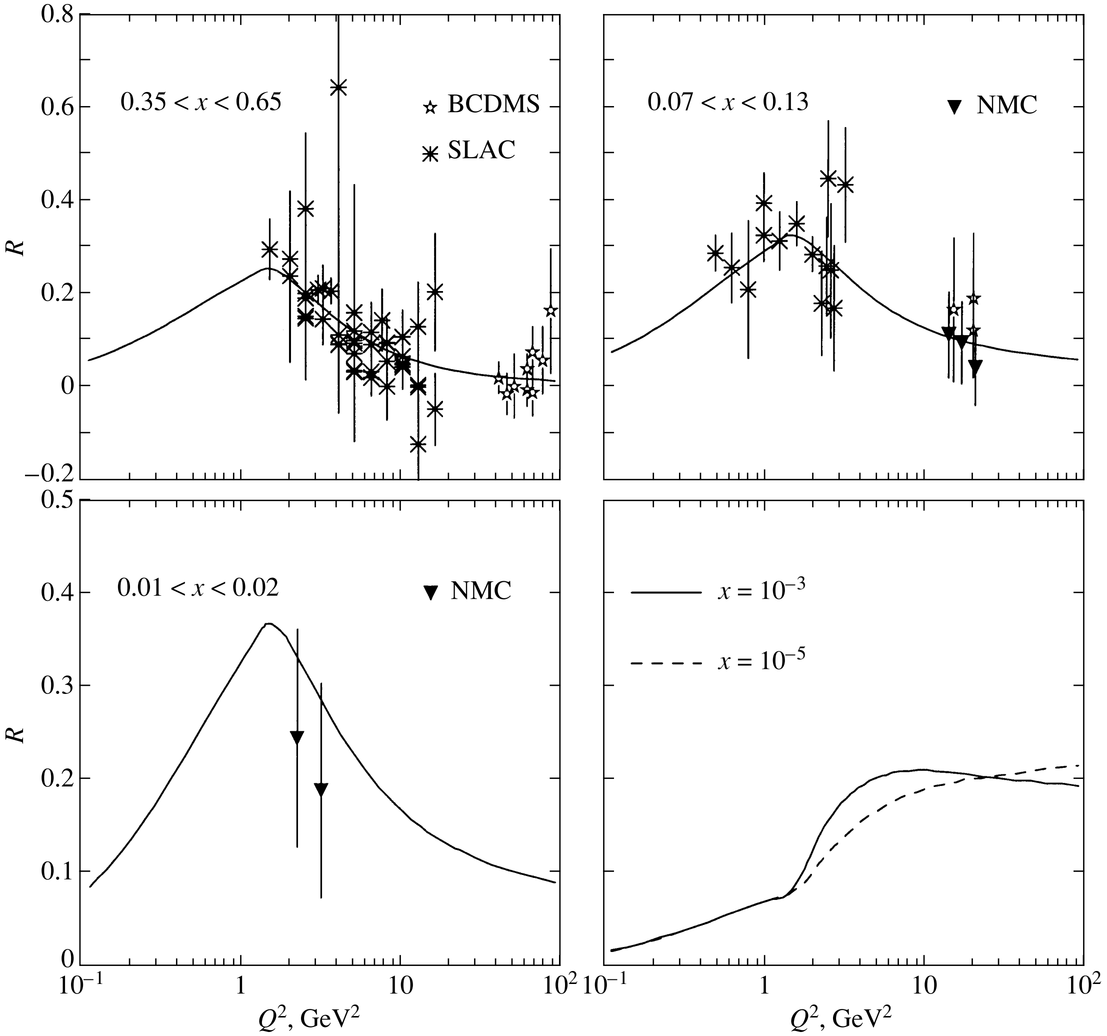


Fig. 9. The ratio $R(x, Q^2)$ as a function of Q^2 at fixed x. Data are from [5, 20].

In the region $10^{-3} < x < 5 \times 10^{-3}$, a linear fit between $R(\text{MRS99}, x = 10^{-3})$ and $R(\text{SLAC98}, x = 5 \times 10^{-3})$ is used. At $Q^2 < Q_0^2$,

$$R(x, Q^2) = R_{GVDM}(x, Q^2) = C(x)\frac{Q^2}{Q^2 + m_\rho^2}, \quad (42)$$

where the function $C(x)$ is determined by the normalization condition at Q_0^2

$$R_{GVDM}(x, Q_0^2) = R(x, Q_0^2), \quad (43)$$

and the function $R(x, Q_0^2)$ is calculated using Eq. (41). Figure 9 shows the experimental values of R as a function of Q^2 in four ranges of x along with the result of the parameterization in Eqs. (41) and (42). In the region of low $Q^2 < Q_0^2$, R decreases with Q^2 at all values of x, but the dependence on x is not strong (Fig. 10). However, the extrapolation of R outside the kinematical range of data (namely, as $Q^2 \longrightarrow 0$ and $x \longrightarrow 0$) based on the presently available data is a rather delicate problem.

In Figs. 11 and 12, we show the results of calculations of the differential cross sections of the neutral current $e^{\pm}p$ scattering $d\sigma/dQ^2$ and $d\sigma/dy$ at high Q^2. The cross section $d\sigma/dQ^2$ decreases by six orders of magnitude between $Q^2 = 400$ and 4000 GeV^2. This decrease is due to the photon propagator. The cross section $d\sigma/dy$ is shown for different Q^2 regions. For $Q^2 > 400$ GeV^2, the bulk of the cross section is concentrated at small values of y. For $Q^2 > 10^4$ GeV^2, the differential cross section is approximately constant in y. The predictions using the MRS99 parton distribution function give a good description of the measured cross sections. The neutral current scattering at high Q^2 is sensitive to the contribution of Z^0. In accordance with Eqs. (4)–(9), the Z^0 contribution reduces the e^+p (e^-p) cross section at $Q^2 > 10^4$ GeV^2 by approximately 25% (12%).

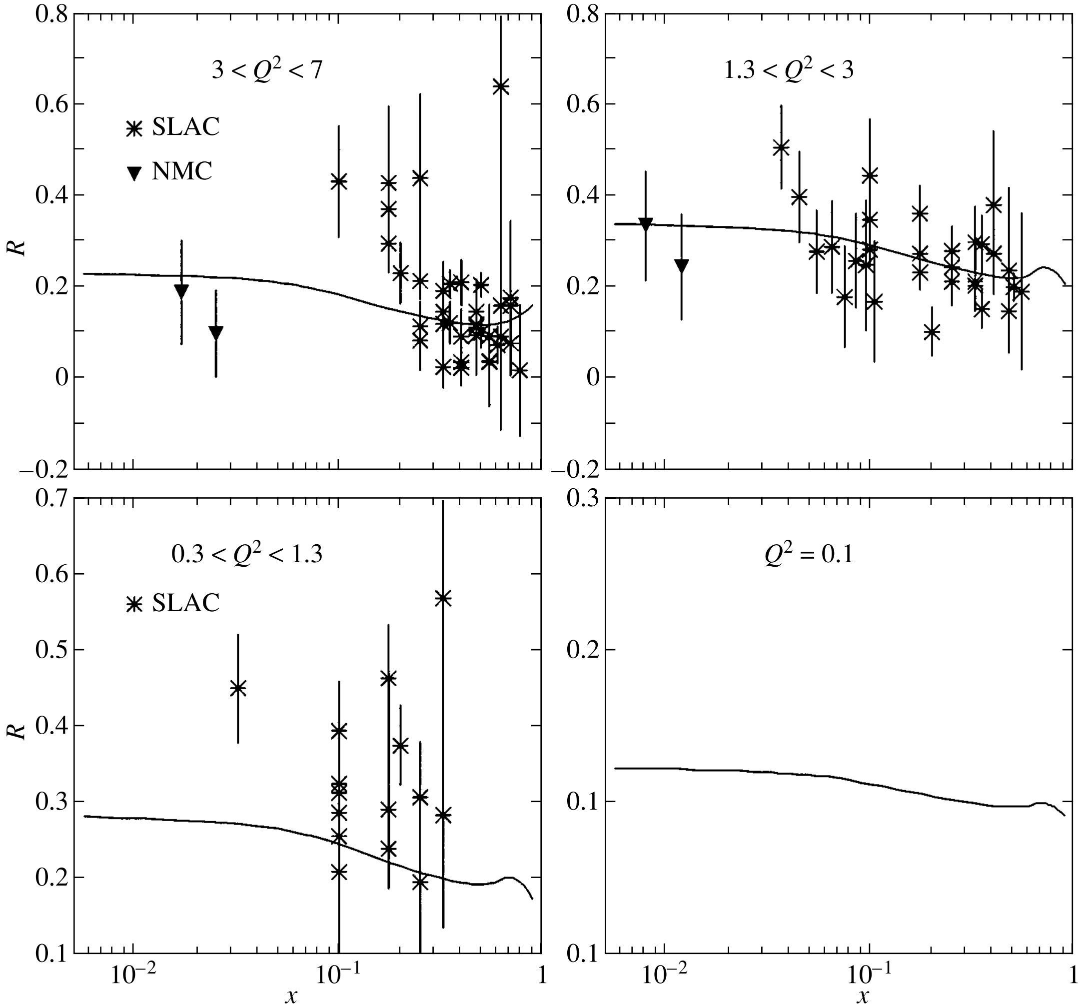


Fig. 10. The ratio $R(x, Q^2)$ as a function of x at fixed Q^2. Data are from [5, 20].

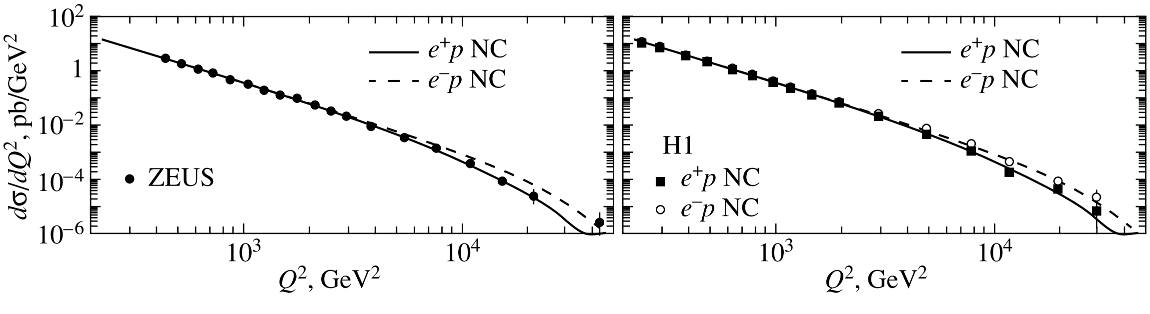


Fig. 11. The differential cross section $d\sigma/dQ^2$ of the neutral current proton scattering off electron (positron) as a function of Q^2. Points are the results of ZEUS [22] and H1 [23] experiments.

4. NUCLEAR STRUCTURE FUNCTIONS

The structure functions measured for different nuclei A are found to differ from the structure functions measured on a deuteron [24, 25]. The modifications are usually observed as a deviation of the ratio $r^{A/d} = F_2^A/F_2^d$ from unity, where F_2^A and F_2^d are the structure functions per nucleon measured in a nucleus and deuteron, respectively. Neglecting nuclear effects in the deuteron, F_2^d can approximately stand for an isospin averaged nucleon structure function, $F_2^N = (F_2^p + F_2^n)/2$. Different nuclear effects are observed in different regions of x.

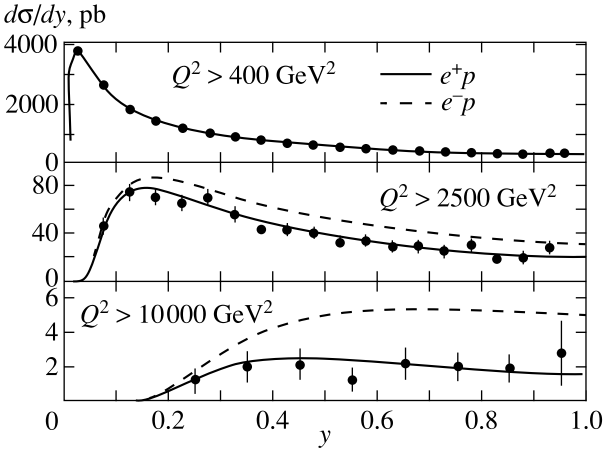


Fig. 12. The differential cross section $d\sigma/dy$ of the neutral current proton scattering off electron (positron) as a function of y. Experimental data are due to ZEUS [22] experiment.

(i) Shadowing at $x < 0.1$. The ratio $r^{A/d}$ is smaller than unity. The experimental data cover the region $x > 10^{-4}$, and $r^{A/d}$ decreases with decreasing x. Shadowing increases with the nuclear mass A and weakly depends on Q^2.

(ii) Antishadowing at $0.1 < x < 0.2$. The NMC data have established a small (several percent) but statistically significant excess over unity. Within the accuracy of the data, no significant Q^2 dependence of this effect has been found.

(iii) The EMC effect at $0.2 < x < 0.8$. The measured ratio $r^{A/d}$ decreases as x rises and has a minimum at $x = 0.6$. The magnitude of this depletion grows slowly with the nuclear mass number. The data imply that a strong Q^2 dependence of $r^{A/d}$ is also excluded in this region.

(iv) The Fermi motion. At $x > 0.8$, the ratio $r^{A/d}$ rises above unity, but experimental information is rather scarce.

Investigations of differences between the longitudinal-to-transverse cross section ratio R for different nuclei showed that $R^{A_i} - R^{A_j}$ is compatible with zero. This implies that nuclear effects influence both structure functions, F_1 and F_2, in a similar way.

At the moment, there is no unique theoretical description of these effects; it is believed that different mechanisms are responsible for them in different kinematical regions. For example, the EMC effect indicates that the averaged momentum carried by valence quarks in nuclei is reduced relative to a free nucleon. It has been shown [26] that the pattern of the function $r^{A/d}(x)$ has a universal shape in the range $10^{-3} < x < 0.96$ and for the mass of the nuclei $A \geq 4$. Namely, the ratio $F_2^A(x)/F_2^d(x)$ can be well approximated with phenomenological functions in different regions of x. At $x > 0.3$,

$$r^{A/d}(x) = 1 - m_b(A)a_{\text{osc}}(x), \tag{44}$$

where the A dependence of m_b can be approximated as

$$m_b(A) = M_b[1 - N_s(A)/A] \quad \text{and} \quad M_b = 0.437 \tag{45}$$

for $A \neq 4$. The number of nucleons $N_s(A)$ at the nuclear surface is given by the Woods–Saxon potential,

$$N_s(A) = 4\pi\rho_0 \int\limits_{r_0(A)}^{\infty} \frac{r^2 dr}{1 + \exp\{[r - r_0(A)]/a\}}, \tag{46}$$

with the values of the parameters $\rho_0 = 0.17$ fm^{-3}, $a = 0.54$ fm, and

$$r_0(A) = 1.12A^{1/3} - 0.86A^{-1/3}. \tag{47}$$

The function $a_{\text{osc}}(x)$ is

$$a_{\text{osc}}(x) = (1 - \lambda x)\left\{\left(\frac{1}{u} - \frac{1}{c}\right) - \mu\left(\frac{1}{u^2} - \frac{1}{c^2}\right)\right\}, \tag{48}$$

where $u = 1 - x$, $c = 1 - x_2$, $x_2 = 0.278$, $\lambda = 0.5$, and $\mu = m_\pi/M$ (m_π is the pion mass). At $10^{-3} \leq x \leq 0.3$, the function is given by

$$r^{A/d}(x) = x^{m_1}(1 + m_2)(1 - m_3 x) \tag{49}$$

with

$$m_i = M_i[1 - N_s(A)/A], \tag{50}$$

where $M_1 = 0.129$, $M_2 = 0.456$, and $M_3 = 0.553$. We used Eq. (49) to calculate $r^{A/d}$ up to $x_0 < 10^{-3}$.

The value of x_0 as a function of A was obtained as follows. The experimental data [26] show that, in the region $5 \times 10^{-3} < x < 0.1$, the ratio $r^{A/d}$ decreases with x. Generally, small x correspond to small Q^2, and the approach of the real photon interaction can therefore be used. Hence, as $x \longrightarrow 0$, $r^{A/d} \longrightarrow \eta^A = \sigma_{\gamma A}/A\sigma_{\gamma N}$, where $\sigma_{\gamma A}$ is the photon-nuclear cross section and $\sigma_{\gamma N}$ is the photon–nucleon cross section averaged over the proton and neutron. The expression for the function η^A has been obtained in [4] using the optical nuclear model,

$$\eta^A = 0.75G(z) + 0.25, \tag{51}$$

where

$$G(z) = \frac{3}{z^2}\left[\frac{z^2}{2} - 1 + e^{-z}(1 + z)\right] \tag{52}$$

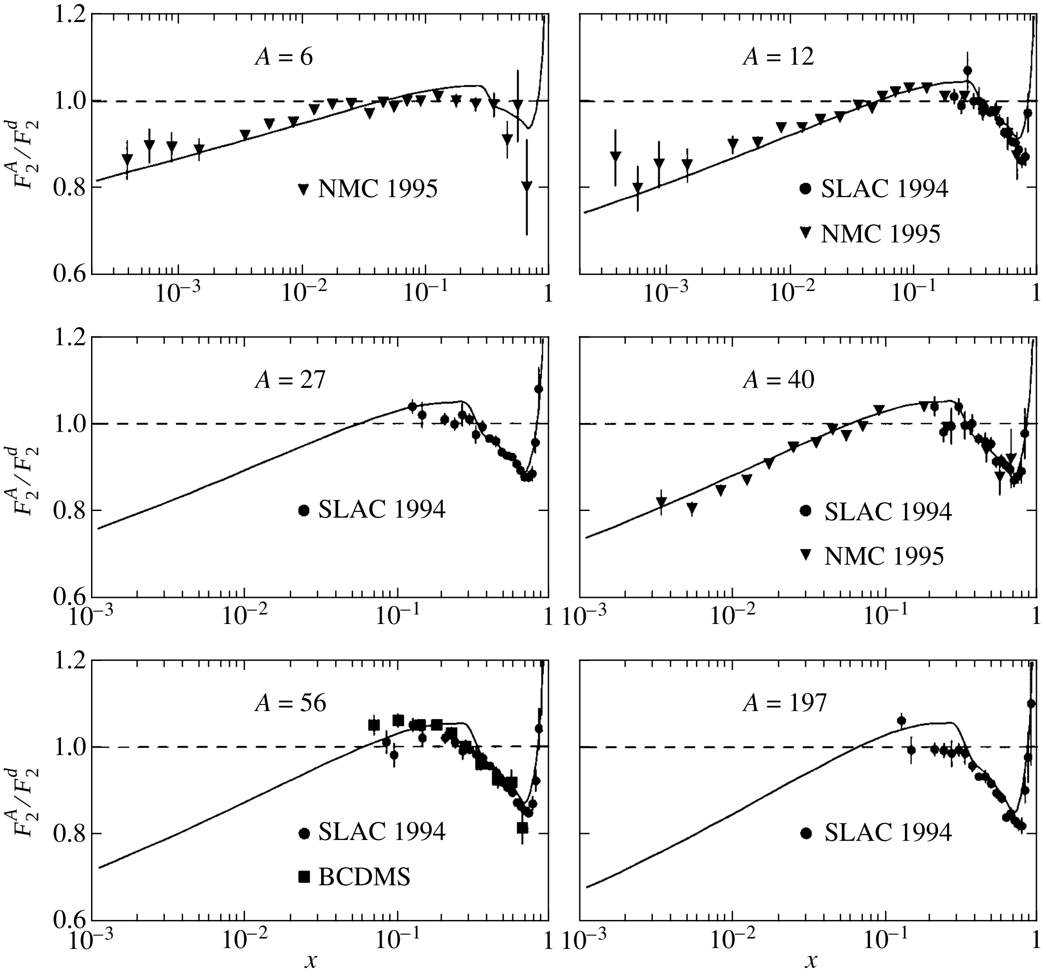


Fig. 13. The comparison of the ratio F_2^A/F_2^d measured by the SLAC, BCDMS, and NMC groups with the approximation given in [26] (solid lines).

and $z = 0.00282A^{1/3}\sigma_{\gamma N}(\nu)$. Using Eq. (34) with the values of parameters obtained in this work, we can write the averaged photon–nucleon cross section as

$$\sigma_{\gamma N} = \frac{1}{2}(\sigma_{\gamma p} + \sigma_{\gamma n}) \\ = 112.2(0.609\nu^{0.0988} + 1.037\nu^{-0.5944}). \quad (53)$$

In the range $x \ll 1$, Eq. (49) is reduced to

$$r^{A/d}(x) = x^{m_1}(1 + m_2). \quad (54)$$

From the asymptotic condition

$$r^{A/d}(x_0) = 0.75G(z) + 0.25, \quad (55)$$

we then obtain the expression for x_0:

$$x_0 = \left[\frac{1}{1 + m_2}(0.75G(z) + 0.25)\right]^{1/m_1}. \quad (56)$$

At $x < x_0$, we assumed that the function $r^{A/d}$ is constant and

$$r^{A/d}(x) = r^{A/d}(x_0). \quad (57)$$

The results of approximating the ratio $r^{A/d}$ are presented in Fig. 13 as functions of x for different nuclear targets and are in good agreement with experimental data.

Taking the nuclear effects into account, the nuclear structure functions $F_i(x, Q^2)$ and total photon-nuclear cross section can be written as

$$F_i^A(x, Q^2) = Ar^{A/d}(x, A)F_i^N(x, Q^2) \quad (58)$$

and

$$\sigma_{\gamma A}(\nu) = A\sigma_{\gamma N}(\nu)[0.75G(z) + 0.25]. \quad (59)$$

The calculated cross sections $\sigma_{\gamma A}(\nu)$ are shown in Fig. 14 as functions of the real photon energy for nuclei C, Cu, and Pb.

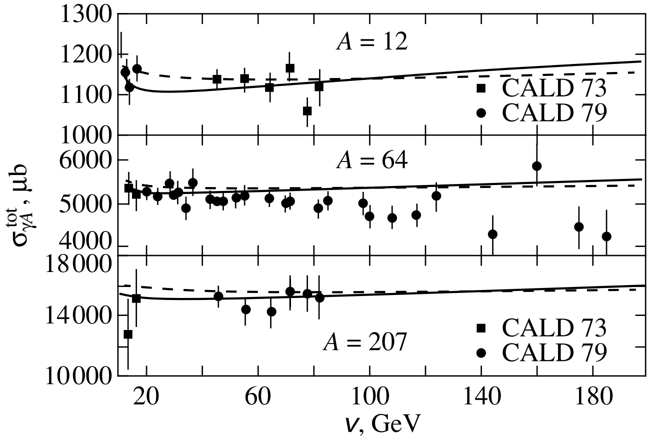


Fig. 14. The total cross section $\sigma_{\gamma A}(\nu)$ for C, Cu, and Pb as a function of the real photon energy. Results of our calculations (solid curves) and calculations by Bezrukov and Bugaev [4] (dashed curves) as compared to experimental data [27].

5. MUON INELASTIC SCATTERING IN THE STANDARD ROCK

We have calculated the main characteristics of the inelastic muon scattering in the standard rock ($A = 22$, $Z = 11$, and $\rho = 2.65$ g/cm^3). The spectra of the muon energy loss in a single interaction,

$$\frac{N_{A\nu}}{A}\nu\frac{d\sigma_{\mu A}}{d\nu} = \frac{N_{A\nu}}{A}\nu\int\limits_{Q^2_{\min}}^{Q^2_{\max}}\frac{d\sigma^2_{\mu A}}{d\nu dQ^2}dQ^2 \qquad (60)$$

(where $N_{A\nu}$ is the Avogadro number), are shown in Fig. 15 as functions of the inelasticity y for different muon energies. The energy dependences of the total cross section,

$$\sigma_{\mu A}(E) = \int\limits_{\nu_{\min}}^{\nu_{\max}}\frac{d\sigma_{\mu A}}{d\nu}d\nu, \qquad (61)$$

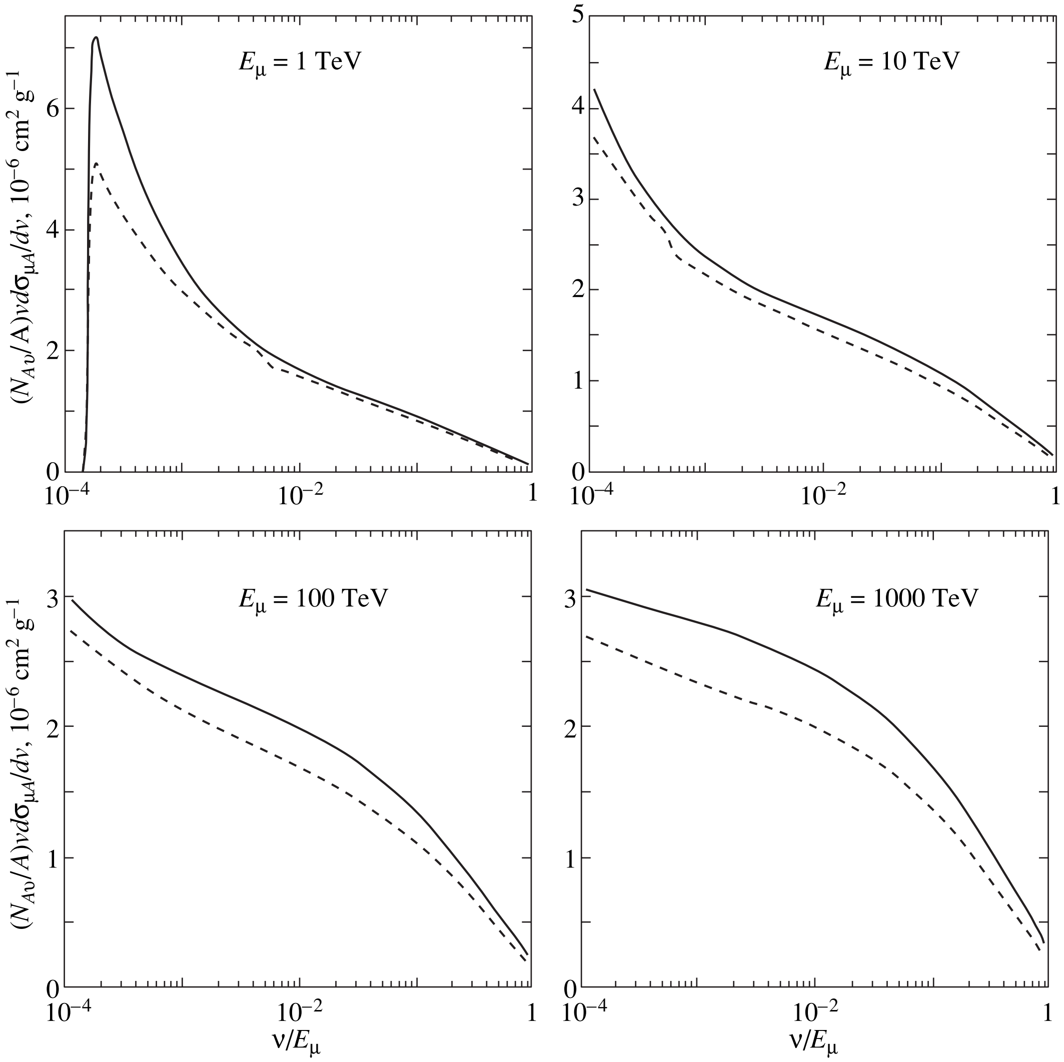


Fig. 15. The spectra of the muon energy losses due to the muon inelastic scattering in the standard rock ($A = 22$, $Z = 11$) as a function of y for fixed muon energies, as compared to calculations by Bezrukov and Bugaev [4] (dashed curve).

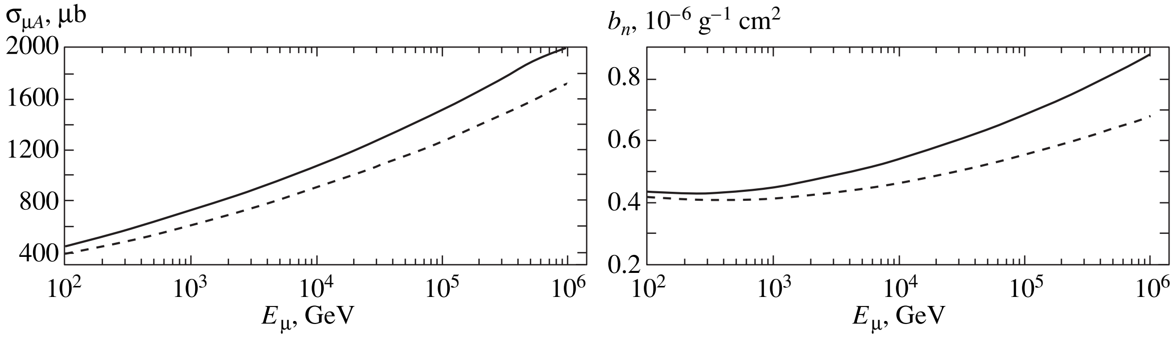


Fig. 16. The total cross section $\sigma_{\mu A}$ and the muon energy loss b_n for the muon inelastic scattering in the standard rock ($A = 22$, $Z = 11$) as a function of the muon energy E_μ, as compared to those in [4] (dashed curve).

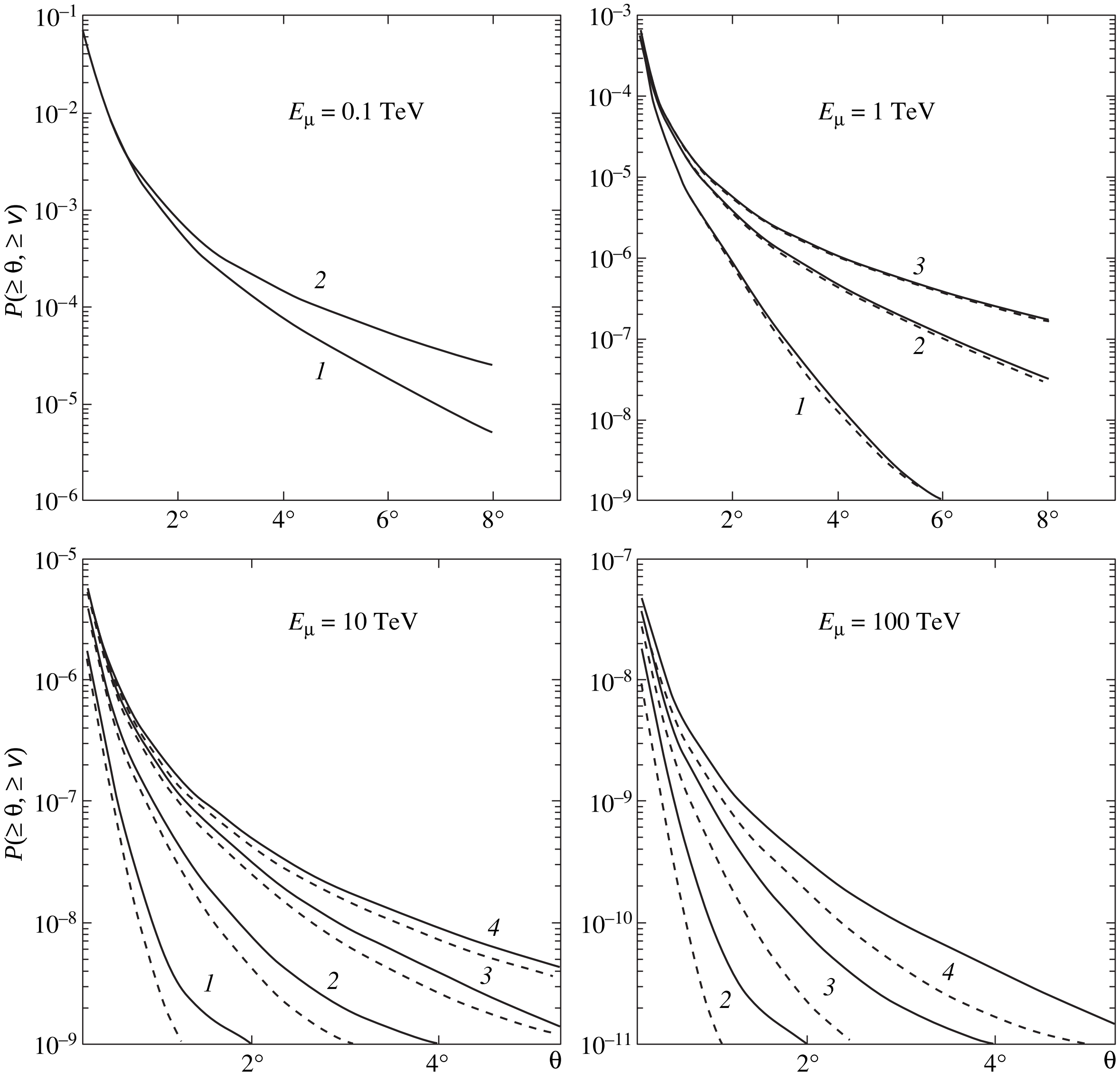


Fig. 17. The probabilities $P(\geq \theta, \geq \nu)$ of the muon scattering in a single interaction in the standard rock ($A = 22$, $Z = 11$) at an angle larger than θ with the outgoing muon energy $E' \geq \nu E$ as a function of θ for fixed values of the incoming muon energies and ν: (*1*) $\nu \geq 10^{-1}$; (*2*) $\nu \geq 10^{-2}$; (*3*) $\nu \geq 10^{-3}$; (*4*) $\nu \geq 10^{-4}$. Solid (dashed) curves are for a positive (negative) muon charge.

and the muon energy loss

$$b_n(E) = \frac{N_{Av}}{A}\int_{\nu_{\min}}^{\nu_{\max}} \nu \frac{d\sigma_{\mu A}}{d\nu} d\nu \quad (62)$$

are shown in Fig. 16. The allowed kinematical region for the variables ν and Q^2 is determined by the equations

$$Q^2 = 2(EE' - |\mathbf{k}||\mathbf{k}'|\cos\theta) - 2m^2 \quad (63)$$

with $\cos\theta = \pm 1$ and

$$Q^2 = 2M\nu + M^2 - W^2. \quad (64)$$

The results in [4] are also given in these figures for comparison. We note that the cross section and the muon energy losses [4] have been calculated taking only the shadowing effect into account. The cross section of the inelastic muon scattering obtained in the present work is larger by a factor of 1.2 and the muon energy loss $b_n(E)$ is also larger by about 8% at $E = 10^3$ GeV and 30% at $E = 10^6$ GeV. As a result, the total energy loss (the sum of bremsstrahlung, pair production, and inelastic muon scattering) increases by about 1% at $E = 10^3$ GeV and 4% at $E = 10^6$ GeV. These differences are mainly due to contributions of small x and small Q^2, where the modern structure functions are larger than that used by Bezrukov and Bugaev [4].

The probabilities $P(\geq \theta, \geq v)$ of the muon scattering in a single interaction at angles larger than θ with the outgoing muon energy $E' \geq vE$ are shown in Fig. 17 as a function of θ for different values of v and primary muon energies. The results are given for μ^- and μ^+ scattering. The main features of inelastic muon scattering are as follows.

(i) At fixed values of θ and E', the probability decreases very rapidly with the initial muon energy E. For example, for $\theta \geq 2°$ and $E' = 10$ GeV, we have $P = 6.3 \times 10^{-4}$ at $E = 10^2$ GeV and $P = 3.8 \times 10^{-6}$ at $E = 10^3$ GeV.

(ii) At fixed values of θ and E, the probability increases with decreasing the outgoing muon energy E'. For $\theta \geq 2°$ and $E = 10^3$ GeV, we have $P = 7.8 \times 10^{-7}$ at $E' \geq 10^2$ GeV and $P = 3.8 \times 10^{-6}$ at $E' \geq 10$ GeV.

(iii) At fixed values of E and E', the mean values of x and Q^2 ($\langle x\rangle$ and $\langle Q^2\rangle$) increase with θ. For the muon energies $E = 10^2$ GeV and $E' \geq 0.1E$, the values of $\langle x\rangle$ and $\langle Q^2\rangle$ increase from $\langle x\rangle = 0.12$ (antishadowing region) and $\langle Q^2\rangle = 0.75$ GeV^2 at $\theta = 0.25°$ up to $\langle x\rangle = 0.25$ (EMC region) and $\langle Q^2\rangle = 34$ GeV^2 at $\theta = 6°$. For the energies $E = 10^3$ GeV and $E' \geq 0.1E$, we have $\langle x\rangle = 0.09$ (EMC region) and $\langle Q^2\rangle = 28$ GeV^2 at $\theta = 0.25°$ and $\langle x\rangle = 0.46$ (antishadowing region) and $\langle Q^2\rangle = 925$ GeV^2 at $\theta = 6°$. The probability of scattering at large angles is therefore suppressed by the EMC effect.

6. CONCLUSIONS

We have studied the inelastic muon scattering off nuclei.

1. It is shown that the combination of the structure function F_2 predicted by the CKMT model at low and moderate values of Q^2 and the MRS99 parton distribution functions at high Q^2 gives a good description of the data over the entire measured region from photoproduction to DIS. In particular, the CKMT model describes the rise of F_2^p well at low x and Q^2 with the slope $\Delta_0 = 0.0988$. Furthermore, the expression for the neutron structure function F_2^n can be obtained in the framework of this model. The result is in good agreement with the F_2^p/F_2^n and $F_2^p - F_2^n$ data.

2. The MRS99 parton distribution functions describe well the differential cross sections $d\sigma/dQ^2$ and $d\sigma/dy$ calculated taking entire not only the electromagnetic current contribution but also the contributions of the neutral current and the γ–Z interference into account. The γ–Z interference contribution is clearly seen at high $Q^2 > 10^3$ GeV^2.

3. The nuclear effects modify the nucleon structure functions in the entire measured region of x and Q^2. The modification depends on Q^2 very slightly and increases with A.

4. The obtained structure functions have been used for calculations of the muon–nucleus scattering cross section, muon energy losses, and muon angular distributions in inelastic interactions. The nuclear effects and contributions of the neutral current and γ–Z interference were taken into account. As a results, the total cross section and energy losses increase with muon energy faster than predicted in [4]. The scattering of high-energy muons ($E > 10^3$ GeV) at large angles is suppressed by both the EMC effect and the γ–Z interference.

ACKNOWLEDGMENTS

We are grateful A.B. Kaidalov and F.V. Tkachev for helpful discussions. This work was supported by the Russian Foundation for Basic Research (project no. 99-02-18374).

REFERENCES

1. K. Kobayakawa, Nuovo Cimento B **47**, 156 (1967).
2. P. Ditsas, B. J. Read, and G. Shaw, Nucl. Phys. B **99**, 85 (1975).
3. P. Ditsas and G. Shaw, Nucl. Phys. B **113**, 246 (1976).
4. L. B. Bezrukov and E. V. Bugaev, Yad. Fiz. **33**, 635 (1981) [Sov. J. Nucl. Phys. **33**, 635 (1981)].
5. http://www-spires.dur.ac.uk/HEPDATA/structure.html
6. A. Capella, A. Kaidalov, C. Merino, and J. Tran Thanh Van, Phys. Lett. B **337**, 358 (1994).

7. A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. C **14**, 133 (2000).
8. M. Gluck, E. Reya, and A. Vogt, Eur. Phys. J. C **5**, 461 (1998).
9. H. L. Lai, J. Huston, S. Kuhlmann, *et al.*, Eur. Phys. J. C **12**, 375 (2000).
10. D. Schildknecht, Acta Phys. Pol. B **28**, 2453 (1997).
11. H. Abramowicz and A. Levy, hep-ph/9712415.
12. K. Abel, F. Barreiro, and F. J. Yandurain, hep-ph/9610380; K. Abel and F. J. Yandurain, Nucl. Phys. B **495**, 221 (1997).
13. A. M. Cooper-Sarkar, R. Devenish, and A. De Roeck, Int. J. Mod. Phys. A **13**, 3385 (1998).
14. C. Merino, A. B. Kaidalov, and D. Pertermann, Eur. Phys. J. C **20**, 301 (2001).
15. J. Breitweg *et al.* (ZEUS Collaboration), Phys. Lett. B **487**, 53 (2000).
16. D. O. Caldwell, J. P. Cumalat, A. M. Eisner, *et al.*, Phys. Rev. Lett. **40**, 1222 (1978).
17. M. Derrick *et al.* (ZEUS Collaboration), Z. Phys. C **63**, 391 (1994); S. Aid *et al.* (H1 Collaboration), Z. Phys. C **69**, 27 (1995); J. Breitweg *et al.* (ZEUS Collaboration), Eur. Phys. J. C **7**, 609 (1999).
18. M. Arneodo *et al.* (New Muon Collaboration), Phys. Rev. D **50**, R1 (1994).
19. M. Arneodo *et al.* (New Muon Collaboration), Nucl. Phys. B **487**, 3 (1997); L. H. Tao, L. Andivahis, P. L. Anthony, *et al.* (E143 Collaboration), Z. Phys. C **70**, 83 (1996).
20. K. Abe, T. Akagi, P. L. Anthony, *et al.* (E143 Collaboration), Phys. Lett. B **452**, 194 (1999).
21. A. D. Martin, M. G. Ryskin, and A. M. Stasto, Eur. Phys. J. C **7**, 643 (1999).
22. J. Breitweg *et al.* (ZEUS Collaboration), Eur. Phys. J. C **11**, 427 (1999).
23. C. Adloff *et al.* (H1 Collaboration), Eur. Phys. J. C **13**, 609 (2000); C. Adloff *et al.*, Eur. Phys. J. C **19**, 269 (2001).
24. M. Arneodo, Phys. Rep. **240**, 301 (1994).
25. G. Piller and W. Weise, Phys. Rep. **330**, 1 (2000).
26. G. I. Smirnov, Phys. Lett. B **364**, 87 (1995); G. I. Smirnov, Eur. Phys. J. C **10**, 239 (1999).
27. D. O. Caldwell, V. B. Elings, W. P. Hesse, *et al.*, Phys. Rev. D **7**, 1362 (1973); D. O. Caldwell, J. P. Cumalat, A. M. Eisner, *et al.*, Phys. Rev. Lett. **42**, 553 (1979).

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 114–122.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 131–139.
Original Russian Text Copyright © 2002 by Elesin.

SOLIDS
Electronic Properties

The Theory of Coherent Optimized-Nanostructure Taking Laser into Account the Electron–Electron Interaction

V. F. Elesin
Moscow Institute of Engineering Physics (Technical University),
Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: VEF@supercon.mephi.ru

Received January 30, 2002

Abstract—An analytic theory of generation of a coherent laser (laser possessing a coherent electronic subsystem) operating on an optimized nanostructure is developed taking into account the electron–electron interaction. This interaction must be included since it may lead to a violation of stringent resonance conditions of coherent lasing of unipolar lasers in view of the fact that the population in such lasers increases with the pumping current. Using the Hartree–Fock approximation, analytic solutions of the Schrödinger equation were obtained for a strong electromagnetic field with open boundary conditions. The expressions derived for polarization current and electron concentration make it possible to determine the power and frequency of generation as well as amplification profile and other characteristics. It is shown that optimal lasing is realized even when electron–electron interactions are taken into account. In this optimal mode with tuning, no population inversion is required (the populations of working levels are identical). The lasing efficiency is equal to unity; the resonance–tunneling coherent pumping is effective since reflection is zero, and the amplification profile is not broadening by the field. Multimode generation stability, good spectral characteristics, and high limiting powers can be expected. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Thirty years ago, Kazarinov and Suris [1] proposed a new type of a unipolar semiconductor laser (cascade laser) in which radiative transitions occur between space-quantization levels (subbands). The main element of a cascade laser is a quantum well with two working resonance levels. The pumping to the upper level and the removal of electrons from the lower level are executed through resonance tunneling.

Cascade lasers were created in 1994 [2] and covered wavelengths from infrared to submillimeter range.

A unique feature of such lasers is the possibility of coherent pumping, i.e., the transportation of electrons from the emitter to the working level due to coherent resonance tunneling. Coherent tunneling takes place if the time τ_{ph} of phase degradation exceeds the electron lifetime $\tau_\Gamma = \hbar/\Gamma$ in a nanostructure (Γ is the resonance level width). Generally speaking, this is possible in quantum wells, in wires, and, especially, in quantum dots. In fact, this condition is necessary, for example, to realize population inversion in a cascade laser [2].

If condition $\tau_{\mathrm{ph}} > \tau_\Gamma$ is satisfied, it is possible in principle that an electron supplied to the upper level experiences a radiative transition to the lower level and leaves it without violating the coherence.

Indeed, it was shown in [3] that lasing may take place without the participation of dissipative processes. For the sake of brevity, we will refer to a laser in which not only the electromagnetic field but also the electron subsystem is coherent as a coherent laser. It was found that a coherent laser differs in principle from conventional lasers in that population inversion in it is not necessarily required. Coherent tunneling ensures more effective pumping since electrons are accumulated in the well due to quantum interference, and reflection tends to zero.

A coherent laser possesses a number of other advantages (see below). However, effective lasing takes place if, according to [3], rather stringent conditions are satisfied. First, the electromagnetic field frequency must be equal to the energy difference $\varepsilon_{2R} - \varepsilon_{1R}$ of the resonance levels. Second, the energy of electrons supplied from the emitter must coincide with the energy of the upper level, renormalized due to the interaction with a strong electromagnetic field.

At the same time, the electron–electron interaction shifting the resonance levels may violate these conditions. The more so that the population of levels in unipolar lasers increases with the pumping current, leading to a strong energy shift. For this reason, it is important to study the effect of electron–electron interactions on coherent lasing.

The present work aims at developing a consistent theory of coherent lasing on a nanostructure optimized as compared to [3] by taking into account the electron–electron interaction.

Using the Hartree–Fock approximation, we succeeded in obtaining analytic solutions to the Schrödinger equation in a strong electromagnetic field with open boundary conditions. The equations derived for polarization currents and electron concentration make it possible to determine the generation power and frequency, amplification profile, and other characteristics.

The mechanism corresponding to the optimal mode with tuning is determined [3], and it is shown that the lasing efficiency is equal to unity, the populations of the working levels are identical, and the reflectivity of the nanostructure is zero. It is proved that the optimal energy of electrons supplied from the emitter coincides with the resonance energy of the structure renormalized due to the interaction of electrons with the electromagnetic field and with one another.

It is shown that the optimal lasing mode is also preserved when the electron–electron interaction is taken into account. The reason for the stability is that the populations of energy levels remain identical upon an increase in the field so that the shifts of the levels are equal.

It is found that the profile of the amplification line of a coherent laser does not experience broadening induced by the field in contrast to conventional lasers ("saturation effect") [4, 5]. It is well known that the "saturation effect" leads to multimode generation and to power limitation [4, 5].

2. DESCRIPTION OF THE MODEL: BASIC EQUATIONS

Following [3], we consider the following model of a coherent laser (see figure). The figure shows a 1D quantum well with barriers in the form of a δ function at points $x = -a$ and $x = 0$. The parameters of the well are chosen so that the lower two levels with energies ε_{2R} and ε_{1R} have a difference approximately equal to the frequency of the electromagnetic field:

$$\hbar\omega = \varepsilon_{2R} - \varepsilon_{1R}.$$

A steady-state flow of electrons with a density proportional to q^2 and with energy ε approximately equal to ε_{2R} is supplied from the left ($x < -a$). The electromagnetic field

$$E_z(z, t) = E(t)\sin kz\cos(\omega t + \varphi(t)) \tag{1}$$

is emitted upon a transition of electrons from the upper level 2 to the lower level 1. The field is polarized perpendicularly to the plane of the well, and the wave vector is directed along the plane (along the z axis). It is assumed that the field is in an optical cavity of length L. We confine our analysis to single-mode lasing. The equations for the amplitude $E(t)$ and phase $\varphi(t)$ of the field have the form (see [3])

$$\frac{\partial E}{\partial t} = -\frac{E}{2\tau_0} - \frac{2\pi}{\kappa}J_c, \tag{2}$$

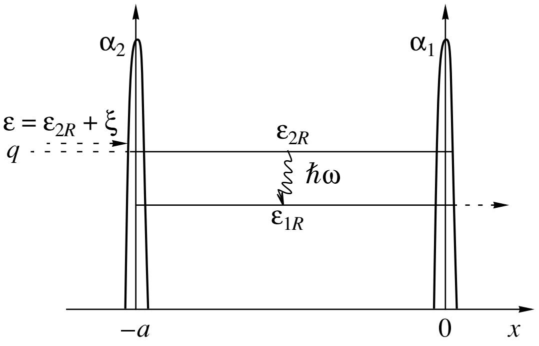


Figure.

$$E\frac{\partial\varphi}{\partial t} + (\omega - \Omega)E = -\frac{2\pi}{\kappa}J_s,$$
$$J_{c,s} = \frac{1}{a}\int_0^a J_{c,x}(x)dx, \tag{3}$$

where $J_{c,s}(x)$ are the polarization currents describing transitions between energy levels, J_c coinciding in phase with the field and J_s being shifted in phase by $\pi/2$.

The wave function $\Psi(x, t)$ of the system satisfies the Schrödinger equation

$$i\frac{\partial\Psi}{\partial t} = -\frac{\partial^2\Psi}{\partial x^2} + U(x)\Psi + g\Psi|\Psi|^2 + \hat{V}(x, t)\Psi. \tag{4}$$

Here, we set $2m^*/\hbar = 1$, $c = 1$,

$$U(x) = \alpha_2\delta(x + a) + \alpha_1\delta(x), \tag{5}$$

α_2 and α_1 being the "intensities" of the emitter and collector barriers, respectively. The cubic term simulates the local electron–electron interaction in the Hartree–Fock approximation (see [6]) with parameter g. The last term in Eq. (4),

$$\hat{V}(x, t) = i2eA_x(t)\frac{\partial\Psi}{\partial x} = V(e^{i\omega t} - e^{-i\omega t})\frac{\partial\Psi}{\partial x},$$
$$V = -\frac{eE}{\omega}, \tag{6}$$

describes the interaction of electrons with the electromagnetic field, and $A_x(t)$ is the vector potential in the Coulomb gauge, which differs from zero in the well. It should be noted that in Eq. (4), as in [3] (see also [5]), the term quadratic in $A(t)$ is omitted, which is justified when

$$V/p = eE/p\omega \ll 1$$

(p is the electron momentum).

Following [3], we seek the steady-state solution to Eq. (4) in the resonance approximation in the form

$$\Psi(x, t) = \psi_2(x)e^{-i\varepsilon t} + \psi_1(x)e^{-i(\varepsilon - \omega)t}. \tag{7}$$

The partial wave functions $\psi_n(x)$ ($n = 1, 2$) describe states with quasi-energies ε and $\varepsilon - \omega$ and satisfy the system of equations

$$\begin{aligned} \varepsilon\psi_2 + \psi_2'' &= V\psi_1' + g\psi_2[|\psi_2|^2 + 2|\psi_1|^2], \\ (\varepsilon - \omega)\psi_1 + \psi_1'' &= -V\psi_2' + g\psi_1[|\psi_1|^2 + 2|\psi_2|^2]. \end{aligned} \tag{8}$$

We consider an optimized structure with several boundary conditions slightly differing from those in [3]:

$$\begin{gathered} \psi_2(-a)\left(1 - \frac{\alpha_2}{ip_2}\right) + \frac{\psi_2'(-a)}{ip_2} = 2\tilde{q}, \\ \tilde{q} = q\exp(-ip_2a), \\ \psi_2(0) = 0, \quad \psi_1(-a) = 0, \quad \varepsilon = p_2^2, \\ p_2^2 - \omega = p_1^2, \\ \psi_1(0)\left(1 - \frac{\alpha_1}{ip_1}\right) - \frac{\psi_1'(0)}{ip_1} = 0. \end{gathered} \tag{9}$$

The boundary conditions describe the electron flux on the left ($x = -\infty$) and the reflection of electrons and their departure from level 1 to the region $x > 0$. In contrast to [3], the departure of electrons from level 2 to the region $x > 0$ and from level 1 to the region $x < -a$ is forbidden. This modification improves the lasing conditions and can be realized in practice (see, for example, [2, 7]).

Currents $J_c(x)$ and $J_s(x)$ can be expressed in terms of functions $\psi_n(x)$:

$$\begin{aligned} J_c(x) &= -ie[(\psi_2^*\psi_1' + \psi_1^*\psi_2') - \text{k.k.}], \\ J_s(x) &= e[(\psi_1^*\psi_2' - \psi_2^*\psi_1') + \text{k.k.}]. \end{aligned} \tag{10}$$

3. WAVE FUNCTIONS FOR A COHERENT LASER

We will solve the system of equations (8), (9) using the methods developed in [3, 6]. In [3], we succeeded in finding exact solutions to Eqs. (8), (9) with $g = 0$ without applying perturbation theory for the field and derived sample expressions for ψ_n and currents using the smallness of parameters V/p and p/α. The inequality $p/\alpha \ll 1$ indicates that the resonance level widths Γ_j are smaller than the resonance energies ε_{nR}. It is precisely in this situation that remarkable properties of quantum wells are manifested. According to [3], the general solution to Eqs. (8) and (9) with $g = 0$ can be sought in the form

$$\psi_n(x) = \sum_{j=1}^{4} A_{nj}\exp(\gamma_j x), \tag{11}$$

where the complex eigenvalues γ_j satisfy the equation

$$\gamma^4 + 2\gamma^2\left(\varepsilon + \frac{V^2 - \omega}{2}\right) + \varepsilon^2 - \varepsilon\omega = 0. \tag{12}$$

Using the smallness of parameter V/p, we can present the roots of Eq. (12) in the form

$$\begin{aligned} \gamma_{1,2} &= \pm ip_1\left(1 - \frac{V^2}{\omega}\right)^{1/2}, \\ \gamma_{3,4} &= \pm ip_2\left(1 + \frac{V^2}{\omega}\right)^{1/2}. \end{aligned} \tag{13}$$

The inclusion of the interaction renders the system of equations (8), (9) nonlinear. In order to find the solution, we can use the approach employed in [6] on account of the following circumstance. It was mentioned above that resonance polarization currents are quite sensitive to the shift of levels due to the electron–electron interaction. Indeed, the energy shift by the value equal to the level width Γ, which is smaller than ε_{nR}, changes the currents considerably, although the electron concentration is relatively low. Consequently, we can assume that the nonlinear term is small and seek the solution to Eq. (8) in the form (11), but with coefficients $A_{nj}(x)$ depending on coordinate x:

$$\psi_n(x) = \sum_{j=1}^{4} A_{nj}(x)\exp(\gamma_j x). \tag{14}$$

In this case, $A_{nj}(x)$ can be regarded as slowly varying functions of x, their rate of variation dA_{nj}/dx being of the same order of smallness as the nonlinear term. Moreover, it is sufficient to assume that the following relations hold in expression (14):

$$\gamma_{1,2} = \pm ip_1, \quad \gamma_{3,4} = \pm ip_2.$$

Substituting solution (14) into system (8) and averaging over the slowly varying functions $\exp(\pm ip_n x)$ (see [8] for details), we arrive at the following equations for $A_{nj}(x)$:

$$\begin{gathered} 2ip_2\frac{dA_{23}(x)}{dx} \\ = gA_{23}(x)[|A_{23}|^2 + 2(|A_{24}|^2 + |A_{11}|^2 + |A_{12}|^2)], \\ -2ip_2\frac{dA_{24}(x)}{dx} \\ = gA_{24}(x)[|A_{24}|^2 + 2(|A_{23}|^2 + |A_{11}|^2 + |A_{12}|^2)], \\ 2ip_1\frac{dA_{11}(x)}{dx} \\ = gA_{11}(x)[|A_{11}|^2 + 2(|A_{23}|^2 + |A_{24}|^2 + |A_{12}|^2)], \end{gathered} \tag{15}$$

$$-2ip_1\frac{dA_{12}(x)}{dx}$$
$$= gA_{12}(x)[|A_{12}|^2 + 2(|A_{23}|^2 + |A_{24}|^2 + |A_{11}|^2)].$$

While deriving Eqs. (15), we have omitted terms of the order of $(V/p)^2$ and Vg/p. It should be borne in mind that, according to [3], coefficients A_{21}, A_{22} and A_{13}, A_{14} are of the order of smallness of V^2/p^2 relative to A_{23}, A_{24} and A_{11}, A_{12}, respectively. Consequently, the inclusion of the coordinate dependence of A_{21}, A_{22} and A_{13}, A_{14} would only result in an excessive accuracy in parameter Vg/p. Equations (8) also lead to the following relations for coefficients A_{1j} and A_{2j}:

$$A_{ij} = \varepsilon_j A_{2j}, \quad \varepsilon_{1,2} = \mp\frac{i\omega}{Vp_1}, \quad \varepsilon_{3,4} = \pm\frac{ip_2V}{\omega}. \tag{16}$$

Using Eqs. (15), we can easily prove that moduli $|A_{nj}|$ are independent of the coordinate. Taking this circumstance into account, we arrive at the solution of system (15):

$$\begin{aligned} A_{23}(x) &= A_{23}\exp(-ip_2xg_3), \\ A_{24}(x) &= A_{24}\exp(ip_2xg_4), \\ A_{11}(x) &= A_{11}\exp(-ip_1xg_1), \\ A_{12}(x) &= A_{12}\exp(ip_1xg_2), \end{aligned} \tag{17}$$

$$\begin{aligned} g_1 &= \frac{g}{2p_1^2}(|A_{12}|^2 + n_1 + 2n_2), \\ g_2 &= \frac{g}{2p_1^2}(|A_{11}|^2 + n_1 + 2n_2), \\ g_3 &= \frac{g}{2p_2^2}(|A_{24}|^2 + n_2 + 2n_1), \\ g_4 &= \frac{g}{2p_2^2}(|A_{23}|^2 + n_2 + 2n_1), \end{aligned} \tag{18}$$

$$n_1 = |A_{11}|^2 + |A_{12}|^2, \quad n_2 = |A_{23}|^2 + |A_{24}|^2. \tag{19}$$

Here, A_{nj} are the constant coefficients that will be determined from the boundary conditions (9). The corresponding equations for the sought coefficients A_{nj} have the form

$$\begin{gathered} A_{21}m_1\exp(-\gamma_1 a) + A_{22}m_2\exp(-\gamma_2 a) \\ + A_{23}m_{3g}\exp(-\gamma_{3g}a) + A_{24}m_{4g}\exp(-\gamma_{4g}a) = 2\tilde{q}, \\ A_{21} + A_{22} + A_{23} + A_{24} = 0, \\ A_{21}\varepsilon_1\exp(-\gamma_{1g}a) + A_{22}\varepsilon_2\exp(-\gamma_{2g}a) \\ + A_{23}\varepsilon_3\exp(-\gamma_3 a) + A_{24}\varepsilon_4\exp(-\gamma_4 a) = 0, \\ A_{21}\varepsilon_1\tilde{m}_{1g} + A_{22}\varepsilon_2\tilde{m}_{2g} + A_{23}\varepsilon_3\tilde{m}_3 + A_{24}\varepsilon_4\tilde{m}_4 = 0. \end{gathered} \tag{20}$$

Here, the following notation has been introduced:

$$\begin{aligned} \gamma_{1g} &= ip_1(1-g_1), \quad \gamma_{2g} = -ip_1(1-g_2), \\ \gamma_{3g} &= ip_2(1-g_3), \quad \gamma_{4g} = -ip_2(1-g_4), \\ m_{jg} &= 1 - \frac{\alpha_2 - \gamma_{jg}}{ip_2}, \quad \tilde{m}_{jg} = 1 - \frac{\alpha_1 + \gamma_{jg}}{ip_1}. \end{aligned} \tag{21}$$

In order to find m_j and $\tilde{m}_j$, we must set $g = 0$ in m_{jg} and $\tilde{m}_{jg}$.

Solving system (20), we obtain the following expression for coefficients to within small terms of the order of $(V/p)^2$ and Vg/p:

$$A_{23} = -A_{24} = \frac{2\tilde{q}}{\Delta(g)}\varepsilon_1\varepsilon_2\tilde{\Delta}_{12}(g), \tag{22}$$

$$\begin{aligned} A_{21} = \frac{4\tilde{q}p_2}{p_1\Delta(g)}\Bigl[&\frac{i\alpha_1}{p_1}(\cos(p_2a) - \exp(-\gamma_{2g}a)) \\ &+ 2\cos(p_2a) - \exp(-\gamma_{2g}a)\Bigr], \end{aligned} \tag{23}$$

$$\begin{gathered} A_{22} = \frac{4\tilde{q}p_2}{p_1\Delta(g)} \\ \times\left[\frac{i\alpha_1}{p_1}(\cos(p_2a) - \exp(-\gamma_{1g}a)) - \exp(-\gamma_{1g}a)\right], \end{gathered} \tag{24}$$

$$\Delta(g) = -\varepsilon_1\varepsilon_2\tilde{\Delta}_{12}(g)\Delta_{34}(g) + \frac{16\alpha_1\alpha_2}{p_1^2}, \tag{25}$$

$$\begin{aligned} \tilde{\Delta}_{12}(g) &= \tilde{m}_{2g}\exp(-\gamma_{1g}a) - \tilde{m}_{1g}\exp(-\gamma_{2g}a), \\ \tilde{\Delta}_{34}(g) &= m_{4g}\exp(-\gamma_{4g}a) - m_{3g}\exp(-\gamma_{3g}a). \end{aligned} \tag{26}$$

If energies ε and ε – ω are close to resonance values, we can write the determinants in the form

$$\begin{aligned} \tilde{\Delta}_{12}(g) &= -\frac{\alpha_1 a}{p_1^2}(\xi_1 + \nu + i\Gamma_1), \\ \Delta_{34}(g) &= -\frac{\alpha_2 a}{p_2^2}(\xi_2 + i\Gamma_2), \end{aligned} \tag{27}$$

$$\begin{aligned} \Delta(g) &= \frac{\varepsilon_1\varepsilon_2\alpha_1\alpha_2}{(p_1p_2)^2}\tilde{\Delta}(g), \\ \tilde{\Delta}(g) &= [\lambda^2 - (\xi_2 + i\Gamma_2)(\xi_1 + \nu + i\Gamma_1)], \end{aligned} \tag{28}$$

$$\begin{gathered} \lambda = \frac{4p_1p_2V}{a\omega}, \quad \xi_1 = \xi - \tilde{g}_1, \quad \xi_2 = \xi - \tilde{g}_2, \\ \xi = \varepsilon - \varepsilon_{2R}, \quad \nu = \omega_{21} - \omega, \quad \omega_{21} = \varepsilon_{2R} - \varepsilon_{1R}, \end{gathered} \tag{29}$$

$$\tilde{g}_1 = g\left(2n_2 + \frac{3}{2}n_1\right), \quad \tilde{g}_2 = g\left(2n_1 + \frac{3}{2}n_2\right),$$
$$\Gamma_j = \frac{2p_j^3}{a\alpha_j^2}, \quad \varepsilon_{nR} = p_{nR}^2, \quad p_{nR} = \frac{\pi n}{a}. \tag{30}$$

In turn, in accordance with (19), (22)–(24), the electron densities n_2 and n_1 at the upper and lower levels, respectively, satisfy the following equations:

$$n_2 = \frac{Q\Gamma_2(\xi_1^2 + \Gamma_1^2)}{a|\tilde{\Delta}(g)|^2}, \tag{31}$$

$$n_1 = \frac{Q\Gamma_2\lambda^2}{a|\tilde{\Delta}(g)|^2}, \quad Q = 4q^2p_2. \tag{32}$$

Thus, the self-consistent system of equations (31), (32) makes it possible to find the wave function for a fixed field λ and the pumping current Q. The results coincide with those obtained in [3] for $g = 0$ and with the results of [6] for $\lambda = 0$.

4. POLARIZATION CURRENTS OF A COHERENT LASER

Substituting functions ψ_n into Eqs. (10), we obtain currents J_c and J_s:

$$J_c = -\frac{EQ\Gamma_1\Gamma_2\eta}{|\tilde{\Delta}(g)|^2}, \tag{33}$$

$$J_s = -\frac{EQ\Gamma_2(\xi_1 + \nu)\eta}{|\tilde{\Delta}(g)|^2}, \quad \eta = \frac{64e^2p_1p_2}{3\omega^2a^2}. \tag{34}$$

We will first analyze the frequency and energy dependences of the polarization current $J_c(\xi, \nu)$, which describe the transitions between energy levels. In addition, we will determine the coefficients of electron reflection from the structure. In this section, field λ and current Q will be regarded as fixed and independent parameters.

The most interesting results can be obtained in the case when resonance levels have identical widths $\Gamma_1 = \Gamma_2 = \Gamma$. This is attained with the following choice of the barrier parameters:

$$\alpha_2 = \alpha_1\left(\frac{p_2}{p_1}\right)^{3/2} \approx 2\sqrt{2}\alpha_1. \tag{35}$$

A generalization to the case of different values of Γ_j can be carried out easily (see below).

It is convenient to introduce the quantity $K(\xi, \nu)$,

$$K(\xi, \nu) = -\frac{J_c}{E\eta} = \frac{Q\Gamma^2}{|\tilde{\Delta}(g)|^2}, \tag{36}$$

having the meaning of the high-frequency nonlinear response describing the amplification (or absorption) of the electromagnetic field. On the other hand, the quantity $KE^2/2\omega$ is proportional to the rate of electron transition from level 2 to level 1.

We will first analyze the situation with noninteracting electrons ($g = 0$) in order to determine subsequently the contribution of the interaction more clearly. In addition, the analysis for $g = 0$ is of independent interest since a structure optimized as compared to that in [3] is considered and new features of coherent resonance tunneling in a strong electromagnetic field are revealed thereby. It should be recalled that fields with $\lambda > \Gamma$ are regarded here as strong.

First of all, we determine the values ξ_0 and ν_0 for which $K(\xi, \nu)$ attains its maximum value proceeding from the equations

$$\left.\frac{\partial K}{\partial \xi}\right|_{\xi = \xi_0} = 0, \quad \left.\frac{\partial K}{\partial \nu}\right|_{\nu = \nu_0} = 0. \tag{37}$$

Since the square of the determinant modulus for $g = 0$ is defined as

$$|\tilde{\Delta}|^2 = (\lambda^2 + \Gamma^2 - \xi^2 - \xi\nu)^2 + \Gamma(2\xi + \nu)^2, \tag{38}$$

we can write Eqs. (37) in the form

$$(2\xi_0 + \nu_0)(\xi_0^2 + \Gamma^2 - \lambda^2 + \xi_0\nu_0) = 0,$$
$$\xi_0(\xi_0^2 + \Gamma^2 - \lambda^2 + \xi_0\nu_0) + \nu_0\Gamma^2 = 0. \tag{39}$$

It can easily be verified that, for any ξ_0, the peak of the gain $K(\xi_0, \nu_0)$ corresponds to $\nu_0 = 0$, i.e., to frequency ω coinciding with the resonance value ω_{21}. Setting $\nu_0 = 0$ in Eqs. (39), we arrive at the following equation for the optimal energy tuning ξ_0:

$$\xi_0(\xi_0^2 + \Gamma^2 - \lambda^2) = 0. \tag{40}$$

The first solution, $\xi_0 = 0$, corresponds to the maximum value of $K(\xi_0, 0)$ for $\lambda < \Gamma$ and to its minimum value for $\lambda > \Gamma$. The second solution,

$$\xi_0^2 = \lambda^2 - \Gamma^2, \quad \lambda > \Gamma, \quad \xi_0 = \pm\sqrt{\lambda^2 - \Gamma^2}, \tag{41}$$

gives the maximum gain for a strong field $\lambda > \Gamma$. In this case, $K(\xi_0, 0)$ assumes the following values for $\xi_0 = 0$ and $\xi_0^2 = \lambda^2 - \Gamma^2$:

$$K(0, 0) = \frac{4Q\Gamma^2}{(\lambda^2 + \Gamma^2)^2}, \tag{42}$$

$$K(\xi_0, 0) = \frac{Q}{\lambda^2}, \quad \xi_0^2 = \lambda^2 - \Gamma^2. \tag{43}$$

In order to clarify the physical meaning of ξ_0, we calculate the coefficient of reflection from the structure, using Eqs. (11) and (22):

$$R = \frac{(\xi^2 + \Gamma^2 - \lambda^2)^2}{(\lambda^2 + \Gamma^2 - \xi^2)^2 + 4\xi^2\Gamma^2}. \tag{44}$$

The substitution of ξ_0 from expression (41) nullifies the reflectivity for any $\lambda \geq \Gamma$ and Q.

It follows hence that the electron energy $\varepsilon_0 = \varepsilon_{2R} + \xi_0$ coincides with the resonance energy of the structure in which the resonance levels of the well split due to the action of electromagnetic field [9].

Thus, each electron supplied to level 2 from the emitter experiences a radiative transition and is removed from level 1 to the collector; i.e., the efficiency of the structure is equal to unity. Indeed, if we calculate the number of transitions from level 2 to level 1 per unit time,

$$\frac{J_c E}{2\omega} = Q, \quad \omega = \varepsilon_{2R} - \varepsilon_{1R} \approx 3p_1^2, \tag{45}$$

it is found to be exactly equal to the number Q of supplied electrons.

Let us now analyze the frequency dependence of the gain $K(\xi, \nu)$ (the so-called amplification profile [4]). In the general case of arbitrary ξ, the gain $K(\xi, \nu)$ is described by formula (36) with $\tilde{\Delta}$ from Eq. (38).

In the resonance situation, when ξ_0 is defined by formula (41), we obtain

$$K(\xi_0, \nu) = \frac{Q\Gamma^2}{\lambda^2[\nu^2 + 4\Gamma^2]}. \tag{46}$$

It follows from this expression that the broadening is determined only by Γ and is independent of λ. Such a line profile differs fundamentally from that for ordinary lasers. For example, for gas lasers [4] and semiconductor lasers, the quantity $K(\nu)$ has the form

$$K(\nu) = \frac{Q}{\nu^2 + \Gamma^2 + \lambda^2}. \tag{47}$$

It can be seen from this expression that the line width increases with the field. The broadening is associated with the "effect of saturation" of the level population due to interlevel electron transitions in a strong field. The "saturation effect" leads to burning of a "hole" in the electron distribution [4, 5], the emergence of additional modes, and limitation of the generation power [5].

In the nonresonance case, when $\xi_0 = 0$, we have

$$K(0, \nu) = \frac{Q}{\nu^2 + (\lambda^2 + \Gamma^2)^2/\Gamma^2}, \tag{48}$$

and the broadening ($\lambda^4/\Gamma^2 \gg \lambda^2$) is much stronger than in coherent lasers.

The reason for the independence of broadening from the field is that electrons supplied from the emitter undergo a radiative transition and emerge in the collector. This phenomenon resembles the "self-transparency" effect [10]. However, in contrast to this effect, coherent tunneling controlling the supply and removal of electrons in a certain phase plays a decisive role here.

Let us find the dependence of $K(\xi, 0)$ on ξ in the vicinity of ξ_0. After calculations, we obtain

$$K(\xi, 0) = \frac{Q\Gamma^2}{4\lambda^2[\Gamma^2 + (|\xi| - |\xi_0|)^2]}. \tag{49}$$

It follows hence that the peak of the gain near ξ_0 is very narrow and has a width independent of the field λ and equal to Γ.

Let us now take into account the interaction between electrons. The gain $K_g(\xi, \nu)$ is defined by formula (36) with $|\tilde{\Delta}(g)|^2$ such that

$$\begin{aligned} |\tilde{\Delta}(g)|^2 &= (\lambda^2 + \Gamma^2 - \xi_1\xi_2 - \nu\xi_2)^2 \\ &\quad + \Gamma^2(\xi_1 + \xi_2 + \nu)^2. \end{aligned} \tag{50}$$

The interaction leads to a level shift and, generally, to a violation of resonances. This effect becomes significant for electron concentrations at which the energy shift starts exceeding the resonance level width.

In this work, we confine our analysis to the most interesting solution of system (31), (32), for which $K_g(\xi, \nu)$ attains its maximum value. It can be verified directly that the following solution of system (31), (32), (36) satisfies this condition:

$$\begin{gathered} n_1 = n_2 = n = \frac{Q}{4\Gamma a}, \\ \xi_1^2 = \xi_2^2 = (\tilde{\xi}_0 - \tilde{g})^2 = \lambda^2 - \Gamma^2, \\ \tilde{g}_1 = \tilde{g}_2 = \tilde{g} = \frac{7g}{2}n, \quad \omega = \omega_{21}, \end{gathered} \tag{51}$$

$$K_g(\tilde{\xi}_0, 0) = \frac{Q}{4\lambda^2} \equiv K(\xi_0, 0). \tag{52}$$

In this case, the energy of supplied electrons is given by the formula

$$\tilde{\xi}_0 = \pm\sqrt{\lambda^2 - \Gamma^2} + \frac{7gQ}{8\Gamma a}. \tag{53}$$

It follows from expression (52) that the gain preserves its form (43) if $\tilde{\xi}_0$ is defined by formula (53). The resonance form of $K_g(\tilde{\xi}_0, 0)$ is preserved due to the equality of populations $n_1 = n_2$ (51). Indeed, since the populations are identical, the energy gap $\varepsilon_{2R} - \varepsilon_{1R}$ does not change, and the energy ξ_0 (53) of supplied electrons coincides with the resonance energy.

However, in contrast to ξ_0 defined by formula (43), a correction associated with the level shift due to the interaction also appears and $K(\tilde{\xi}_0, 0)$ becomes an asymmetric function of ξ (see relation (56) below). It should be noted that the second solution in (53) gives a value of ξ_0 modulo decreasing with Q, which vanishes for a certain value of λ_0:

$$\tilde{\xi}_0 = -\sqrt{\lambda_0^2 - \Gamma^2} + \frac{7gQ}{8a\Gamma} = 0. \tag{54}$$

This means that there exists an optimal mode in the absence of tuning due to the compensation of level shifts. It can be proved that the reflection coefficient becomes equal to zero as before if the value of ξ_0 is chosen in accordance with relation (53). The shape of the amplification line profile

$$K_g(\tilde{\xi}_0, \nu) = \frac{4Q\Gamma^2}{\lambda^2[\nu^2 + 4\Gamma^2]} \tag{55}$$

is preserved as well as the form of the dependence of K on $\tilde{\xi}$:

$$K(\tilde{\xi}, 0) = \frac{Q\Gamma^2}{\lambda^2[\Gamma^2 + (|\tilde{\xi}| - |\tilde{\xi}_0|^2)]}. \tag{56}$$

Thus, the optimal self-tuning mode predicted in [3] is also preserved when the electron–electron interaction is taken into account.

Concluding the section, we note that the maximum efficiency of the structure is attained only for $\Gamma_1 = \Gamma_2$. If $\Gamma_1 \neq \Gamma_2$, we have the following expression instead of (44) for $\lambda \gg \Gamma$:

$$R = \frac{(\Gamma_2 - \Gamma_1)^2}{(\Gamma_2 + \Gamma_1)^2}. \tag{57}$$

The equality $\Gamma_1 = \Gamma_2$ can be attained with the optimized structure considered by us here under the condition (35).

5. SINGLE-MODE COHERENT LASING

The amplitude and frequency of steady-state coherent lasing can be determined from the equations

$$1 = \tilde{Q}\frac{\Gamma^2}{|\tilde{\Delta}(g)|^2}, \tag{58}$$

$$\omega - \Omega = \tilde{Q}\frac{\Gamma(\varepsilon - \omega - \varepsilon_{1R}(g))}{\tau_0|\tilde{\Delta}(g)|^2}, \tag{59}$$

$$\tilde{Q} = 4\pi\tau_0 Q\eta/k,$$

which are derived by substituting currents J_c and J_s into Eqs. (2) and (3).

It can easily be seen that the frequency is defined by the expression

$$\omega = \frac{S(\varepsilon - \varepsilon_{1R}(g)) + \Omega}{1 + S}, \quad S = 1/\Gamma\tau_0, \tag{60}$$

which is valid for any $\tilde{Q}$ and λ.

The stabilization factor S may vary over a wide range. For example, it we take the typical values for a cascade laser [2] $\Gamma = 10^{12}$ and $\tau_0 = 10^{-11}$ s, we obtain $S \approx 0.1$. In this case, we have

$$\omega \approx \Omega, \quad S \ll 1; \tag{61}$$

i.e., the frequency is determined by the resonance frequency. Assuming that it can be varied, we consider ω to be a free parameter. The opposite case of $S \gg 1$ will be considered below.

We will first study lasing for noninteracting electrons for the same reason as in the previous section. Using the results obtained in Section 4, where it was proved that the current J_c attains its maximum value for $\nu = 0$, we set $\omega = \Omega = \omega_{21}$ in relations (58) and (60). Equation (58) implies that the lasing power is

$$P_\omega = \lambda^2 = \xi^2 + \Gamma(\sqrt{\tilde{Q} - 4\xi^2} - \Gamma). \tag{62}$$

The threshold current is obtained by setting $\lambda = 0$ in expression (62):

$$\tilde{Q}_{\text{th}} = (\xi^2 + \Gamma^2)^2/\Gamma^2. \tag{63}$$

It was proved in [3] that, for a certain tuning of the energy ξ_0, power P_ω attains its maximum value. The equation for ξ_0,

$$\frac{dP_\omega}{d\xi} = 2\xi_0\left(1 - \frac{2\Gamma}{\sqrt{\tilde{Q} - 4\xi_0^2}}\right) = 0, \tag{64}$$

has two solutions:

$$\xi_0 = 0, \quad \xi_0^2 = \tilde{Q}/4 - \Gamma^2. \tag{65}$$

The first solution corresponds to the peak power in the pumping interval $1 < \tilde{Q}/\Gamma^2 < 4$, where the power

$$P_\omega = \Gamma(\sqrt{\tilde{Q}} - \Gamma) \tag{66}$$

is a root function of $\tilde{Q}$.

The second solution is realized for $\tilde{Q} > 4\Gamma^2$; in this case, the power is a linear function of the pumping current:

$$P_\omega = \tilde{Q}/4, \tag{67}$$

which corresponds to the maximum effectiveness (the efficiency is equal to unity). The conditions for the emergence of the second solution taking into account relation (67) and the expression for ξ_0^2 can be written in the form

$$\lambda > \Gamma, \quad \xi_0^2 = \lambda^2 - \Gamma^2. \tag{68}$$

These expressions exactly coincide with relations (41) defining the peak of current J_c and the zero of the reflection coefficient. Consequently, we may conclude that the energy $\varepsilon_0 = \varepsilon_{2R} + \xi_0$ of the supplied electrons exactly coincides with the resonance energy of the structure of a coherent laser, so that each electron makes a contribution to lasing. It is for this reason that the power increases linearly with $\tilde{Q}$ (67). If, however, $\xi_0 = 0$, the energy of electrons begins to differ from the resonance energy upon an increase in $\tilde{Q}$. This leads to an increase in reflection (see relation (44)) and a decrease in the lasing efficiency, i.e., to a root dependence (66) of the power.

It should be noted that the lasing power P_ω (67) does not depend on Γ, while $\tilde{Q}_{\text{th}} = \Gamma^2$ decreases with Γ. Consequently, one can strive, in principle, to reduce the width Γ. The lower limit is set only by the coherence condition $\Gamma > \tau_{\text{ph}}^{-1}$. In addition, it should be emphasized that the maximum efficiency is attained in an optimized structure with symmetric barriers for $\Gamma_1 = \Gamma_2$.

It was indicated in [3] that a distinguishing feature of a coherent laser is the absence of a requirement of population inversion. We will prove this by calculating explicitly the difference between populations n_1 and n_2 defined by formulas (31) and (32). In the absence of tuning ($\xi_0 = 0$), the difference between populations,

$$n_1 - n_2 = \frac{Q\Gamma}{a|\tilde{\Delta}|^2}(\Gamma^2 - \lambda^2), \tag{69}$$

changes upon an increase in λ. For $\lambda > \Gamma$, the population n_1 of the lower level exceeds the population n_2 of the upper level. In the case of optimal tuning ξ_0 (41), the populations are identical:

$$n_1 = n_2 = \frac{Q}{4\Gamma a}. \tag{70}$$

It can be seen from this relation that the populations increase with the pumping current Q, which is typical of unipolar lasers, both of the semiconductor (e.g., cascade [2]) type and of gas lasers [4]. The increase in the electron concentration leads to an increase in the electron–electron interaction, which, generally speaking, may violate the stringent conditions of resonance tuning (41).

For this reason, we must consider the effect of the electron–electron interaction on coherent lasing proceeding from the system of equations (58), (59), (31), and (32). We can easily verify that the optimal lasing is preserved if the energy of electrons supplied from the emitter is given by

$$\varepsilon = \varepsilon_{2R} + \tilde{\xi}_0, \tag{71}$$

while $\tilde{\xi}_0$ is defined by formula (53). In this case, the solution of the system (58), (59), (31), and (32) has the form

$$\lambda^2 = \tilde{Q}/4, \quad n_1 = n_2 = Q/4\Gamma a, \quad \omega = \omega_{21}. \tag{72}$$

The reasons for the stability of resonance tuning to the electron–electron interaction are similar to those considered in Section 4.

The results of Section 4 concerning the shape of the coherent lasing line (46), which does not depend on field induced broadening, are also preserved. We can expect that the latter circumstance would lead to a higher stability of the single-mode operation of a coherent laser and to higher limiting powers as compared to a conventional laser. Finally, we consider the case of large values of the stabilization factor $S \gg 1$, for which, in accordance with relations (60), we have

$$\omega \approx \varepsilon - \varepsilon_{1R}(g). \tag{73}$$

In this case, there is no resonance tuning mode since

$$|\tilde{\Delta}(g)|^2 = (\lambda^2 + \Gamma^2)^2 + \Gamma^2\xi_2^2. \tag{74}$$

6. CONCLUSIONS

It has been proved that a coherent laser has certain advantages such as the absence of the requirement of population inversion, high effectiveness (efficiency is equal to unity), low thresholds, the absence of the amplification profile broadening above the threshold, and the stability of parameters to the electron–electron interaction. The properties of the amplification profile broadening raise hopes of the multimode lasing stability, high spectral parameters of the laser field, and high limiting values of power. The latter is due to the fact that, according to [5], field induced broadening sets a limit on lasing power.

The remarkable properties of coherent lasers are due to the fact that electrons are supplied from the emitter to the optimal energy interval with optimal phases.

At the same time, effective lasing requires the fulfillment of coherent generation conditions ($\tau_{\text{ph}} > \Gamma^{-1}$) and a small width of the energy distribution of electrons supplied from the emitter. The coherence conditions can be satisfied in principle, especially in quantum dots (see, for example, [2, 7]). The publication [10] on superradiation, where coherence was observed in a system of eight quantum wells, is also worth mentioning.

As regards the requirement of small width of the electron energy distribution, it can be satisfied using filters on the same quantum wells (dots) (see, for example, [2, 7]). In addition, it was shown in [3] that many properties of coherent lasing are also preserved for a broad source.

Naturally, experimental realization requires mode detailed calculations for specific structures by taking into account a number of additional factors.

ACKNOWLEDGMENTS

The author thanks Yu.V. Kopaev for fruitful discussions.

This work was carried out in the framework of the Special Federal Program "Integration" (project no. A0133) under support from the Ministry of Industry and Science of the Russian Federation, program "Physics of Solid Nanostructures" (project no. 99-1140), and the project "Construction of the Theory of Interaction of Strong High-Frequency Electromagnetic Fields with the Electron System of Resonance–Tunnel Diodes and Lasers."

REFERENCES

1. A. Kazarinov and R. Suris, Fiz. Tekh. Poluprovodn. (Leningrad) **5**, 242 (1971) [Sov. Phys. Semicond. **5**, 207 (1971)].
2. J. Faist, F. Capasso, D. Sivco, *et al.*, Science **264**, 553 (1994).
3. V. F. Elesin, Zh. Éksp. Teor. Fiz. **112**, 483 (1997) [JETP **85**, 264 (1997)].
4. W. E. Lamb, Phys. Rev. **134**, 1429 (1964).
5. V. M. Galitskiĭ and V. F. Elesin, Zh. Éksp. Teor. Fiz. **68**, 216 (1975) [Sov. Phys. JETP **41**, 104 (1975)].
6. V. F. Elesin, Zh. Éksp. Teor. Fiz. **119**, 816 (2001) [JETP **92**, 710 (2001)].
7. J. Faist, F. Capasso, D. Sivco, *et al.*, Appl. Phys. Lett. **66**, 538 (1995).
8. N. N. Bogolyubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Fizmatgiz, Moscow, 1963; Gordon and Breach, New York, 1962).
9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1977).
10. S. L. McCall and E. L. Haha, Phys. Rev. **183**, 457 (1969).
11. S. Haas, T. Stroucken, M. Hübuer, *et al.*, Phys. Rev. B **57**, 14860 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 123–131.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 140–149.
Original Russian Text Copyright © 2002 by Demishev, Pronin, Glushkov, Sluchanko, Samarin, Kondrin, Lyapin, Brazhkin, Varfolomeeva, Popova.

SOLIDS
Electronic Properties

Hopping Conductivity of Carbynes Modified under High Pressures and Temperatures: Galvanomagnetic and Thermoelectric Properties

S. V. Demishev[a, *], A. A. Pronin[a, **], V. V. Glushkov[a], N. E. Sluchanko[a, *], N. A. Samarin[a], M. V. Kondrin[b], A. G. Lyapin[b], V. V. Brazhkin[b], T. D. Varfolomeeva[b], and S. V. Popova[b]**

[a]*Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia*
[b]*Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
**e-mail: demis@lt.gpi.ru*
***e-mail: rrr@lt.gpi.ru*
****e-mail: nes@lt.gpi.ru*

Received November 9, 2001

Abstract—The conductivity, thermopower, and magnetoresistance of carbynes structurally modified by heating under a high pressure are investigated in the temperature range 1.8–300 K in a magnetic field up to 70 kOe. It is shown that an increase in the synthesis temperature under pressure leads to a transition from 1D hopping conductivity to 2D and then to 3D hopping conductivity. An analysis of transport data at $T \leq 40$ K makes it possible to determine the localization radius $a \sim (56–140)$ Å of the wave function and to estimate the density of localized states $g(E_F)$ for various dimensions d of space: $g(E_F) \approx 5.8 \times 10^7$ eV^{-1} cm^{-1} ($d = 1$), $g(E_F) \approx 5 \times 10^{14}$ eV^{-1} cm^{-2} ($d = 2$), and $g(E_F) \approx 1.1 \times 10^{21}$ eV^{-1} cm^{-3} ($d = 3$). A model for hopping conductivity and structure of carbynes is proposed on the basis of clusterization of sp^2 bonds in the carbyne matrix on the nanometer scale. *© 2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Various nontraditional materials based on various allotropic modifications of carbon have been intensely studied in recent years. In particular, of considerable interest are carbynes, which are linear polymer chains of *sp* bonds [1, 2]. However, in contrast to other allotropic modifications like graphite or diamond, for which perfect single crystals can be obtained, carbyne samples are found to be disordered from the beginning, which is apparently due to the instability of large linear carbon clusters [2]. It is assumed that linear polymeric *sp* fragments in the carbyne structure alternate with sp^2 centers, which, first, lead to a kink of linear carbon chains and, second, create free bonds for adjoining neighboring chains (in the absence of sp^2 centers, chains are coupled with one another only through Van der Waals forces) [1, 2]. In such a situation, carbyne samples acquire a quasi-amorphous structure which is difficult to determine exactly by using traditional X-ray diffraction methods.

New possibilities of studying carbynes are associated with the samples synthesized under a high pressure [3–5]. Simultaneous effects of high pressure and temperature makes it possible to vary smoothly the fraction of sp^2 bonds in carbynes and to change the magnitude and type of carbyne conductivity. It was proved by us earlier [4, 5] that the conductivity of samples synthesized under pressure is of the hopping type at low temperatures. Since the hopping conductivity is sensitive to the morphology of samples, the study of transport phenomena in carbynes at low temperatures might shed light on the structure of these materials.

It follows from the results obtained in [5] that the conductivity of carbynes in the temperature range $1.8 \text{ K} \leq T \leq 40$ K follows Mott's law [6]

$$\sigma(T) = \sigma_0 \exp[-(T_0/T)^n]. \quad (1)$$

In this case, an increase in the fraction of sp^2 bonds in a sample leads to a decrease in the exponent n of hopping conductivity from $n = 1/2$ to $n = 1/4$ [5]. According to [5], such a behavior can be attributed to a change in the effective dimension d of the space,

$$n = \frac{1}{1+d}, \quad (2)$$

and carbyne is consequently a unique model object on which a transition from 1D ($n = 1/2$) to 2D ($n = 1/3$) and then to 3D hopping conductivity ($n = 1/4$) can be carried out experimentally. It should be noted that the existence of 1D hopping conductivity in carbynes is independently confirmed by the study of dynamic conductivity. It was shown in [5] that the frequency dependence

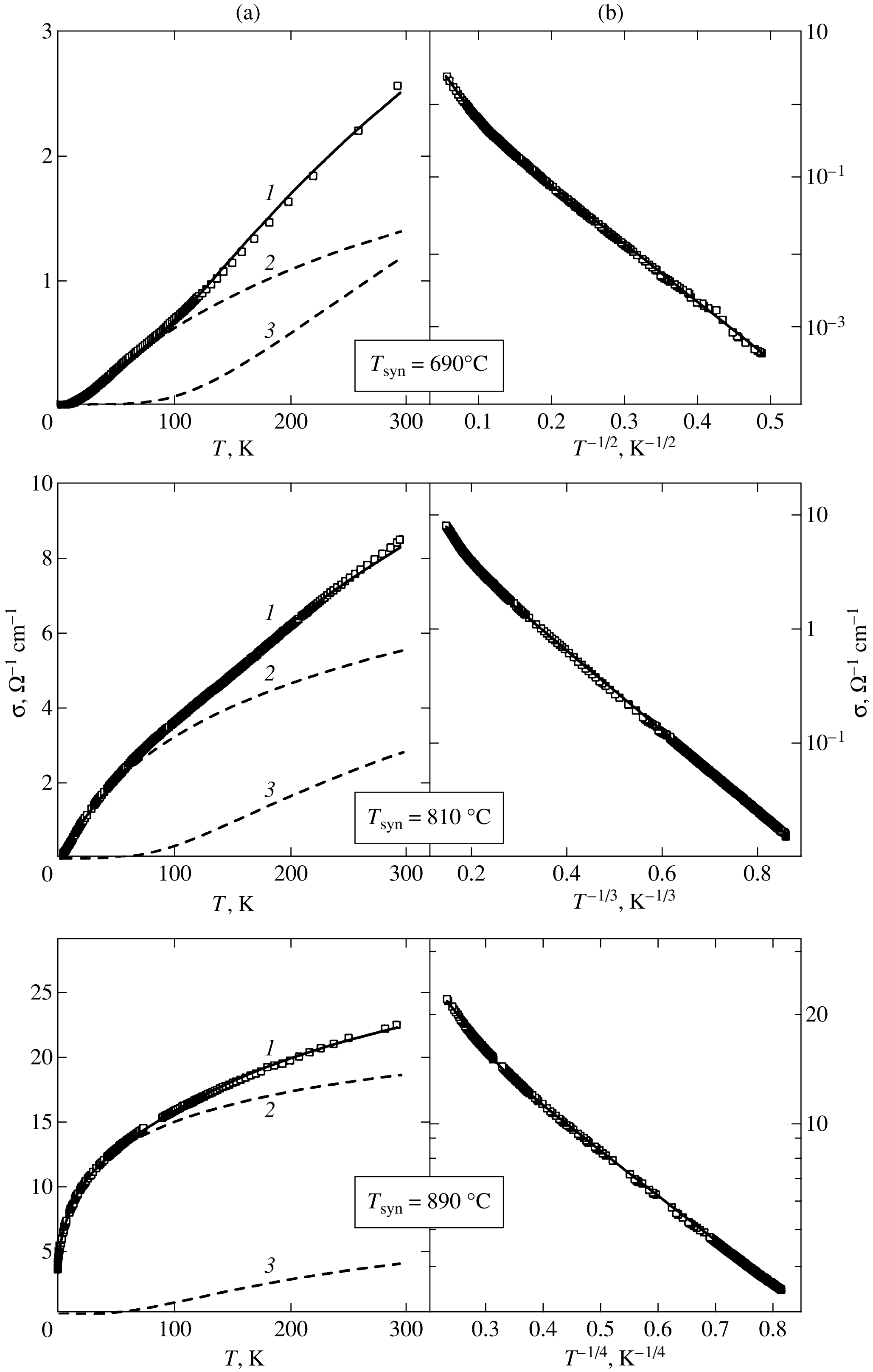


Fig. 1. Conductivity of carbyne samples synthesized at different temperatures under pressure: (a) total conductivity $\sigma(T)$ (curve *1*); symbols correspond to experimental results, and the curve is approximation by formula (4); hopping contribution $\sigma_h(T)$ (curve *2*), and activation contribution $\sigma_a(T)$ (curve *3*); (b) conductivity $\sigma(T)$ in linearizing coordinates $\log\sigma(T)$ vs. $f(T^{-n})$ for hopping conductivity.

Parameters of conductivity and characteristics of localized states of carbynes

T_{syn}, °C	n	T_0, K	E_a, meV	a, Å	$g(E_F)$, cm^{-d} eV^{-1}
690	1/2	360 ± 5	35 ± 3	56 ± 4	5.8×10^{7}
810	1/3	570 ± 10	28 ± 2	75 ± 5	5×10^{14}
890	1/4	70 ± 3	21 ± 2	140 ± 10	1.1×10^{21}

dence of conductivity $\sigma(\omega) \propto \omega^S$ for samples with $n = 1/2$ acquires a strong temperature dependence of the exponent $s(T)$, corresponding to theoretical predictions for the 1D case (index s for 2D and 3D dynamic conductivities is independent of temperature, including the case when Coulomb correlations are taken into consideration [5]).

However, the values of $n > 1/4$ can also be explained by the formation of a Coulomb gap in the density of localized states. For example, the value $n = 1/2$ is typical not only for 1D hopping conductivity, but also in the case of formation of a Coulomb gap irrespective of the dimension of space [7]. For this reason, it would be interesting to verify independently the model [5] which can be obtained by analyzing the thermopower in the hopping region. Indeed, the Seebeck coefficient $S(T)$ is proportional to the derivative of the density of states $g(E_F)$ at the Fermi level [8],

$$S \propto \frac{dg(E_F)}{dE_F} T^{1-2n}, \tag{3}$$

and $S(T)$ must vanish in the case of symmetric correlation gap.

Another important problem in the study of hopping conductivity of carbynes is the determination of the parameters of localized states, such as the localization radius a and the density of states (DoS) $g(E_F)$, since such information is not available in the literature on these materials.

The present study aims at refining the mechanism of hopping conductivity of carbynes and at verifying the results obtain in [5], as well as determining the quantitative parameters describing localized states in the carbyne matrix.

In order to solve this problem, we studied the temperature dependence of conductivity and thermopower as well as field dependences of the magnetoresistance for carbyne samples obtained under a pressure of 7.7 kbar and different temperatures of synthesis in the interval 690°C $\le T_{\text{syn}} \le$ 890°C. The experiments were performed in the temperature range 1.8–300 K in a magnetic field up to 70 kOe. The scheme of the synthesis of the samples was the same as in [3, 4]; the methods of studying transport properties are described in [9]. The structure of carbynes was determined by X-ray analysis [4].

2. CONDUCTIVITY AND THERMOPOWER OF CARBYNES SYNTHESIZED UNDER PRESSURE

The temperature dependences of conductivity $\sigma(T)$ of carbyne samples obtained at various values of T_{syn} are presented in Fig. 1. It can be seen from the data presented in Fig. 1a that, in the entire temperature range 1.8 K $\le T <$ 300 K under investigation, $\sigma(T)$ can be approximated to a high degree of accuracy by the sum of two contributions, viz., hopping σ_h and activation σ_a:

$$\sigma(T) = \sigma_h(T) + \sigma_a(T). \tag{4}$$

Here, $\sigma_h(T)$ is defined by formula (1) and $\sigma_a \propto \exp(-E_a/k_BT)$. In the temperature range $T \le T^* \sim 40$ K, the inequality $\sigma_h \gg \sigma_a$ holds, and clearly manifested linear segments corresponding to variable range hopping conductivity are observed in the $\log\sigma$ vs. $f(T^{-n})$ coordinates (Fig. 1b). According to the results obtained in [5], index n increases upon a decrease in the synthesis temperature from $n = 1/4$ to $n = 1/2$ (see table). For $T > T^*$, the integrated curve $\sigma(T)$ contains a noticeable activation component σ_a, although the ratio σ_h/σ_a is greater than unity up to room temperatures (see Fig. 1). The value of the activation energy for the samples under investigation varies from 20 to 35 meV and decreases upon an increase in the synthesis temperature (see table).

In order to verify the hypothesis [5] according to which values of $n > 1/4$ in carbynes are due to a decrease in the effective dimension of the system, we studied the temperature dependences of the thermopower for samples with $n = 1/3$ and $n = 1/2$. It was found that, in the temperature range $T \ge 20$ K, the Seebeck coefficient decreases with temperature, the sign reversal of thermopower taking place at $T_{\text{inv}} \sim 50$–70 K (Fig. 2a). Thus, the $S(T)$ curves (Fig. 2), as well as the $\sigma(T)$ curves (Fig. 1), indicate the presence of several contributions to the kinetic coefficients of the samples under investigation. The relation between $S(T)$ and $\sigma(T)$ will be discussed in greater detail in Section 4, where the model of conductivity of carbynes is described.

Since the condition $\sigma_h \gg \sigma_a$ holds for $T \le T^* \sim 40$ K (see Fig. 1), the thermopower in this temperature range must be determined by electron jumps. However, the value of T^* is found to be quite close to the thermopower inversion temperature T_{inv}, which may lead to a deviation of the observed dependence $S(T)$ from the

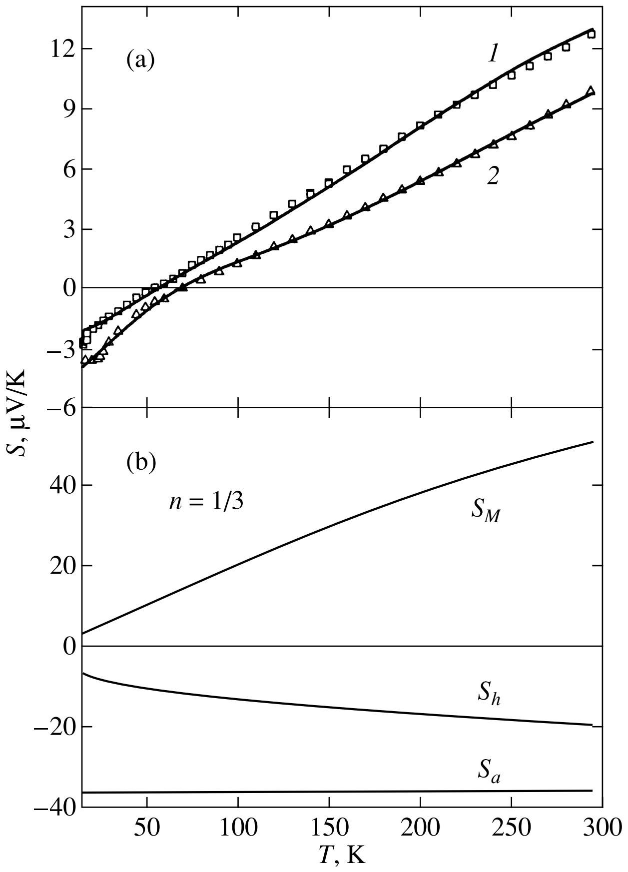


Fig. 2. (a) Temperature dependences of thermopower for a quasi-two-dimensional sample (curve *1*) and for a quasi-one-dimensional sample (curve *2*). Symbols correspond to experiment, and the curves to model computations by formulas (8)–(11). (b) Example of separating contributions for the thermopower of a sample with $n = 1/3$ (see text).

theoretical dependence (formula (3)). For this reason, the analysis of the hopping thermopower $S_h(T)$ was carried out for the temperature range $T \leq 20$ K, in which the function $S(T)$ tends to a power asymptotic behavior (Fig. 3).

It is worth noting that the temperature dependence $S(T)$ for samples with $n > 1/4$ in the temperature range $6\text{ K} \leq T \leq 20$ K is in accord with the theoretical dependence: the thermopower is independent of temperature ($S(T) \approx$ const) for $n = 1/2$ (curve *1* in Fig. 3), while, for $n = 1/3$, the value of $S(T)$ decreases according to the power law $|S| \propto T^{1-2n} = T^{1/3}$ (curve *2* in Fig. 3). In both cases, the value of thermopower in the hopping conductivity region remains finite, indicating, in accordance with formula (3), the asymmetry in the density of states; consequently, the interpretation of values of $n > 1/4$ in the model of correlation gap is ruled out.

Thus, the study of temperature dependences of thermopower confirms the conclusions drawn by us earlier [5], according to which the values of $n > 1/4$ for carbyne samples are due to jumps in the space with a lowered dimensionality.

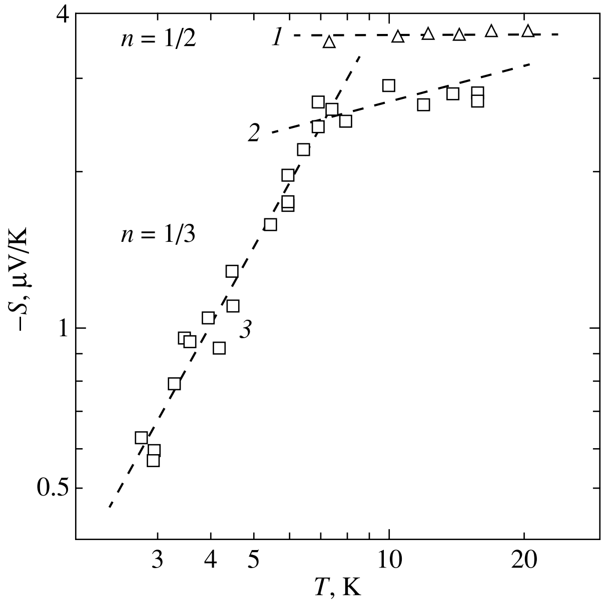


Fig. 3. Temperature dependences of thermopower in the hopping conductivity region for samples with $n = 1/2$ and $n = 1/3$. Asymptotic forms: $S(T) =$ const (*1*), $S(T) \propto T^{1/3}$ (*2*), and $S(T) \propto T^{1.5}$ (*3*).

It is interesting to note that for a sample with $n = 1/3$, for $T \leq 6$ K, the temperature dependence of the hopping type, $|S| \propto T^{1/3}$, is transformed into a stronger dependence $|S| \propto T^{1.5}$ (curve *3* in Fig. 3). Since the exponent $1 - 2n$ in formula (3) for the model of hopping conductivity does not exceed unity, such a behavior cannot be explained if we confine our analysis to the hopping-type conductivity only. We can assume that the stronger decrease in the Seebeck coefficient for $T \longrightarrow 0$ reflects the process of gradual opening of the Coulomb gap and the symmetry in the DoS at the Fermi level caused by this effect (since the thermopower remains finite down to $T \sim 3$K, complete opening of the Coulomb gap does not take place in the temperature range under investigation). In this case, we can expect that a 2D sample with $n = 1/3$ at low temperatures experiences the transition to a temperature dependence of conductivity (1) with exponent $n = 1/2$. Similar effects of enhancement of Coulomb correlations at low temperatures may apparently take place for a sample with 1D conductivity also; however, methodological limitations associated with the measurement of the thermopower for high-resistance samples did not allow us to obtain the data on $S(T)$ at $T \leq 6$ K (curve *1* in Fig. 3). A verification of the assumptions made above requires an analysis of transport properties of carbynes at $T \leq 2$ K (including ultralow temperatures), which may form the subject of an independent study and is beyond the scope of the present paper.

3. MAGNETORESISTANCE AND PARAMETERS OF LOCALIZED STATES OF CARBYNES

In order to determine the localization radius and density of states at the Fermi level, it is necessary to measure simultaneously the temperature dependence of hopping conductivity $\sigma_h(T)$ and the field dependence of the positive contribution to magnetoresistance $\rho_s(H)$ due to the compression of the wave function in the magnetic field [9]. Indeed, for a space of dimension d, parameter T_0 in formula (1) is defined by the formula [6–12]

$$T_0 = \frac{A_d}{g(E_F)a^d k_B}, \tag{5}$$

while $\rho_s(H)$ in the limit of weak magnetic fields has the form [7, 10]

$$\ln\frac{\rho(H)}{\rho(0)} = \frac{t_d e^2 a^4 H^2}{c^2\hbar^2}\left(\frac{T_0}{T}\right)^{3n}. \tag{6}$$

Having determined parameter T_0 from $\sigma_h(T)$, we can use formula (6) to calculate the localization radius a from the derivative $\partial\ln[\rho(H)/\rho(0)]/\partial(H^2)$ and then determine $g(E_F)$ from formula (5). Such a procedure of Mott spectroscopy of localized states requires the knowledge of the values of coefficients A_d and t_d for various dimensions of space. For $d = 3$, the values of A_3 and t_3 are, respectively, 17.6 [8] and 5/2016 [7], while, in the 2D case ($d = 2$), we have $A_2 = 13.8$ [7, 10] and $t_2 = 1/360$ [10]. In the 1D case ($d = 1$), $A_1 \approx 1$ [11, 12], while the value of t_1 is not known. Considering that, for $d = 3$ and $d = 2$, we have $t_3 \approx 2.5 \times 10^{-3} \sim t_2 \approx 2.8 \times 10^{-3}$, we can assume that parameter t_d depends on the dimension of space only slightly. In addition, this parameter appears in the expression for a in the form $a \propto t_d^{1/4}$, and, hence, the equality $t_1 = t_2$ was used for estimating the localization radius in the 1D case. Apparently, we can expect that such a procedure for $d = 1$ gives at least correct orders of magnitude for quantities $g(E_F)$ and a.

The experimental curves of magnetoresistance $\rho(H)$ at $T = 4.2$ K, plotted in $\ln[\rho(H)/\rho(0)]$ vs. H^2 coordinates, are shown in Fig. 4. It can be seen that, in the entire range of magnetic fields under investigation, magnetoresistance is positive and the function $\rho(H)$ in the interval $0 \le H \le 30$ kOe is quadratic in the magnetic field in complete agreement with the theoretical dependence (6). In the range of strong fields ($H > 30$ kOe), a slight downward deviation of the curve $\ln[\rho(H)/\rho(0)] = f(H^2)$ from a linear dependence is observed (see Fig. 4). Such a behavior can be attributed to a transition from the asymptotic behavior of weak fields to the asymptotic behavior of strong fields since, in the latter case, dependence (6) must go over to $\ln[\rho(H)/\rho(0)] \propto H^{1/3}$ dependence [7], and the quadratic dependence of magnetoresistance due to compression of the wave function is suppressed.

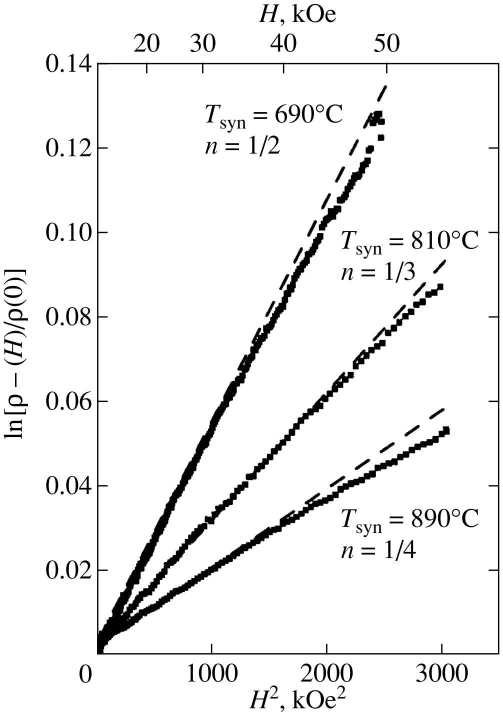


Fig. 4. Magnetoresistance of carbynes synthesized at different temperatures under pressure at $T = 4.2$ K.

It should be noted that for carbyne samples with different dimensions, the curves $\rho(H)$ are not only of the same type, but the amplitudes of magnetoresistance are also close in order of magnitude in spite of radical changes in the effective dimension of the space in which hopping conductivity takes place. This result is in accord with the above assumption that the value of t_d weakly depends on the dimensionality of space. Thus, the results presented in Fig. 4 lead to the conclusion that $\rho(H) \approx \rho_s(H)$ for carbyne samples and the slope of the initial linear segments in the $\ln[\rho(H)/\rho(0)]$ vs. H^2 coordinates can be used for calculating the localization radius.

It was found by using the above approach that the value of a increases with T_{syn} from $a \approx 56$ Å in the 1D case to $a \approx 140$ Å in the 3D case (see table). The corresponding values of the parameter T_0 determined from the data on $\sigma_h(T)$ (see Fig. 1) are given in the table together the DoS calculated for various values of d by formula (5).

It is interesting to note that the carbyne sample under investigation do not exhibit a negative magnetoresistance in the hopping conductivity region. At the same time, the existence of a negative magnetoresistance in weak magnetic fields is typical of the variable range hopping conductivity and was observed repeatedly in experiments with systems having various dimensionalities [9, 10]. Theoretically, the reason for the emergence of a negative magnetoresistance may be associated with the realization of one of two mechanisms: coherent and incoherent [10].

In the coherent case, the probability of a jump should be determined by taking into account interference emerging as a result of scattering from different chains of centers, connecting the initial and final centers [10]. The application of a magnetic field leads to the formation of additional phase shifts between scattered waves, thus "detuning" the interference, increasing the probability of a jump, and, hence, ensuring the emergence of a negative magnetoresistance [10].

In the case of incoherent mechanism, a decrease in the overlap integral between the centers, which is associated with the compression of the wave function in a magnetic field, narrows the function $g(E)$ and, hence, increases $g(E_F)$ and reduces parameter T_0 [10]. Such a renormalization of the density of localized states apparently leads to the emergence of a negative contribution to magnetoresistance.

It follows from a theoretical analysis that a negative magnetoresistance in the hopping conductivity region is of fundamental nature and must be observed along with the positive magnetoresistance described by formula (6). From this point of view, the magnetoresistance of the samples under investigation is anomalous and the absence of a negative magnetoresistance (see Fig. 4) must be associated with considerable peculiarities in the hopping conductivity of carbynes.

4. DISCUSSION

In order to explain the origin of hopping conductivity in carbynes, we must first of all analyze the results of X-ray structural analysis. The X-ray spectra of the samples obtained at different temperatures T_{syn} were used to determine the value of the correlation length $L_{\mathrm{cor}} \sim 10$–12 Å, which defines the characteristic size of an ordered carbon chain …=C=C=C=… of the cumulene type [3, 4]. Since the role of the main structural defects responsible for carbon chain bending in the carbyne matrix is obviously played by carbon atoms in the sp^2 state [2, 4, 5], the concentration of sp^2 centers can be estimated from the relation $N_{sp^2} \sim L_{\mathrm{cor}}^2 \sim 10^{21}$ cm^{-3}.

In addition, it was found that parameter L_{cor} weakly depends on the synthesis temperature in the range $690°\mathrm{C} \le T_{\mathrm{syn}} \le 890°\mathrm{C}$. At the same time, it follows from the results of previous studies that an increase in the value of T_{syn} increases the concentration of sp^2 centers and, as a result, leads to the $sp \longrightarrow sp^2$ transition, i.e., a transition from the quasi-one-dimensional chain structure of carbon atoms to a disordered graphite-like lattice [3–5].

Thus, at first sight, the value of L_{cor} should decrease with increasing concentration of sp^2 bonds (and, hence, with increasing T_{syn}), which contradicts the experimental data. The concept of the $sp \longrightarrow sp^2$ transition can be associated with the constancy of L_{cor} by assuming that new sp^2 centers are formed during synthesis under pressure in the vicinity of the existing defects, e.g., as a result of a lower value of the activation energy for creating another sp^2 center near an existing center as compared to the activation energy of a single sp^2 center in an sp chain. Pairs of adjacent sp^2 centers in this case will lead not only to a kink in individual chains, but also to the formation of closed ring structures which may effectively localize electrons (Fig. 5a).

Indeed, in view of the condition $a > L_{\mathrm{cor}}$ (see table), the volume bounded by the localization radius contains $(4/3)\pi(a/L_{\mathrm{cor}})^3 \sim 10^3$–$10^4$ centers, and the above hypothesis seems to be quite probable.

Another significant feature of localized states in carbynes is that these states are not only multicentered but also multielectron by nature. The carrier concentration N in carbynes can be estimated by assuming that (i) the activation energy E_a is the characteristic scale of variation of the density of localized states, and (ii) in the cases when $d = 1$ and $d = 2$, the corresponding structures ensuring 1D and 2D conductivity are separated by a distance not smaller than the localization radius:

$$N \sim g(E_F)a^{d-3}E_a. \tag{7}$$

An estimate obtained by using this formula and taking into account the data from the table gives values (1–2) × 10^{19} cm^{-3}. This value is close to the volume-averaged electron concentration n_H which can be determined from the Hall effect: $n_H \sim 7 \times 10^{19}$ cm^{-3} for a sample with $T_{\mathrm{syn}} = 890°\mathrm{C}$. It can be seen that both characteristic concentrations N and n_H are 1–2 orders of magnitude lower than the concentration $N_{sp^2} \sim L_{\mathrm{cor}}^{-3}$ of the sp^2 centers. If we attribute the emergence of carriers in the carbyne matrix to carbon atoms in the sp^2 state, which have free bonds, such a discrepancy can be explained by the presence of impurity atoms in considerable concentrations, which weaken the free bonds of the sp^2 centers. It should be noted that the existing scheme of the synthesis of the initial carbyne makes the presence of impurity atoms in this material quite probable [1].

The fulfillment of the inequality $(4/3)\pi a^3 N \gg 1$ for an ordinary doped semiconductor would correspond to the metal side of the metal–insulator transition, for which hopping conductivity is not typical. Consequently, the experimental observation of hopping conductivity in carbynes under the condition $(4/3)\pi a^3 N \gg 1$ indicates that charge carriers are localized in the quasi-metallic region with a characteristic size ~a and form a multielectron state, while jumps take place between such "metallic" regions (Fig. 5a).

The above model of the carbyne structure, which follows from an analysis of the transport properties and structural data, requires a refinement of transition from

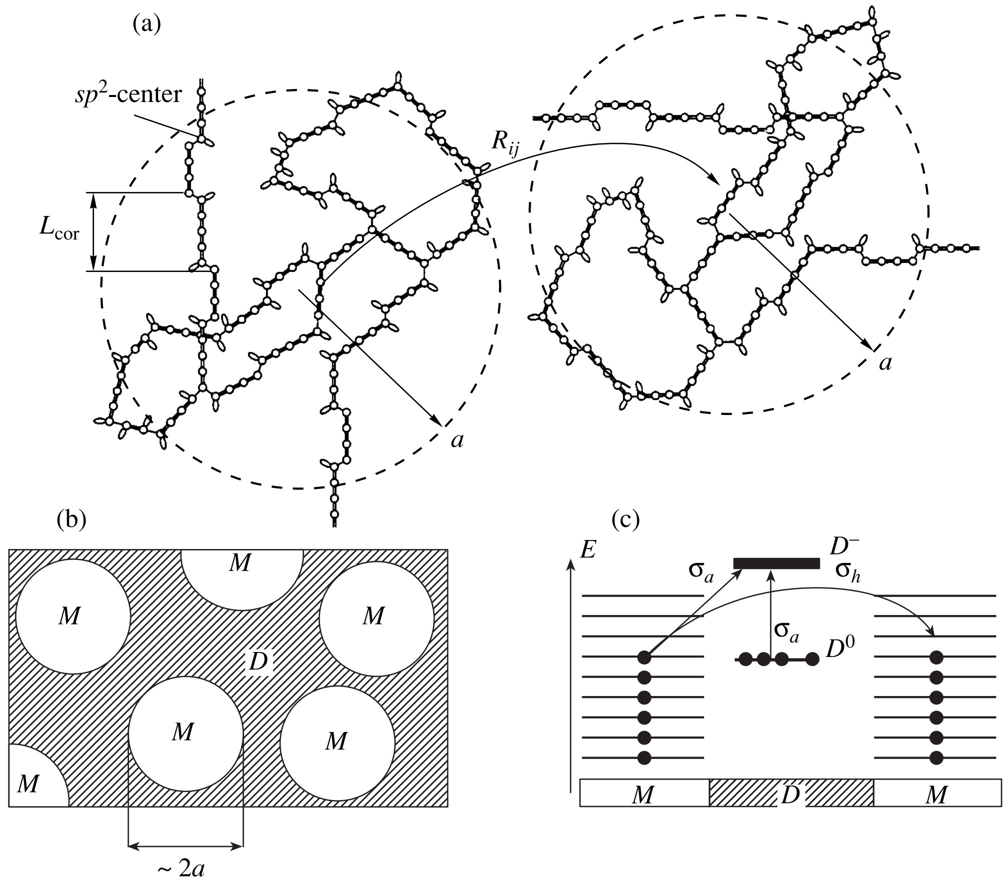


Fig. 5. (a) Model of the carbyne structure; R_{ij} denotes the length of the jump between localized states of radius a. (b) Model of the effective medium used for describing thermopower. (c) Energy level diagram illustrating mechanisms of conductivity in carbynes; D^0 and D^- denote the lower and upper Hubbard bands, respectively.

1D to 3D conductivity as a result of the $sp \longrightarrow sp^2$ transition [5]. A fundamentally new point is the hypothesis on a considerable nonuniformity in the distribution of sp^2 centers on the nanometer scale; the regions with an elevated concentration of carbon atoms in the sp^2 state are apparently responsible for the localization of electron states. If the concentration of sp^2 bonds in the sample is not high (low synthesis temperatures), an analysis of the nonuniform distribution of sp^2 bonds readily suggests a situation in which quasi-linear chains of carbon atoms (containing a small number of sp^2 centers in the sp state) connect regions with a more complex topology characterized by an increased fraction of sp^2 centers (sp^2 "beads" threaded on an sp "string"). The conductivity mechanism in such a system is associated with jumps between clusters with an elevated concentration of sp^2 centers (Fig. 5a). If the interaction between such inhomogeneous chains can be disregarded, the conductivity of the system is of a quasi-one-dimensional type. Obviously, an increase in the fraction of sp^2 bonds upon an increase in T_{syn} leads to the emergence of interaction between the chains and to an increase in the effective dimensionality of conductivity to 2D and then to 3D. Thus, in the proposed modification of the model formulated in [5], the change in the effective dimensionality of conductivity reflects a change in the topology of sp^2 regions in the sp carbyne matrix.

This approach can be used to carry out a quantitative analysis of the temperature dependence of thermopower. It can be seen from Fig. 2*a* that the $S(T)$ value decreases almost linearly with temperature down to $T \sim 70$ K. Such a "quasimetallic" temperature dependence $S(T)$ is typical of various experimental systems exhibiting hopping conductivity [14] and can be interpreted in the framework of the model of an effective medium [15] in which a sample with hopping conductivity is treated as a mixture of a metal and a dielectric (Fig. 5b). The role of the "metal" is played by the regions with spatial localization of electrons, while the "dielectric" is the poorly conducting medium separating the localization regions. The conductivity of such a medium is obviously determined by the regions with the highest resistivity, to which we must ascribe the temperature dependence of the conductivity of the sample to match it with the experimental dependence (see

Fig. 1). However, in contrast to conductivity, the contribution to thermopower comes not only from "dielectric" but also from "metallic" regions. If X is the volume fraction of the metal, the thermopower of such a medium is defined as [13]

$$S = S_M(T)X + S_D(T)(1-X), \tag{8}$$

where $S_M(T)$ and $S_D(T)$ are the thermopowers of the "metal" and the "dielectric." For $S_M(T)$, we can use the standard Mott formula [6] modified with allowance for high-temperature corrections [14]:

$$S_M(T) = aT + bT^3, \tag{9}$$

while $S_D(T)$ is given by the expression taking into account the two contributions (hopping and activation) [8]:

$$S_D(T) = \frac{S_h(T)\sigma_h(T) + S_a(T)\sigma_a(T)}{\sigma_h(T) + \sigma_a(T)}. \tag{10}$$

In this formula, $S_h(T)$ is the hopping contribution and is described by relation (3), while $S_a(T)$ corresponds to the activation mechanism of conductivity. Depending on the nature of activation, $S_a(T)$ may have different functional forms. For example, in the case of activation to the mobility threshold or to the conductivity band, we have

$$S_a(T) \propto \frac{k_B}{e}\frac{E_a}{k_B T},$$

while, in the case of activation to the upper Hubbard band, the generalized Heickes formula is valid [16],

$$S_a(T) = -\frac{k_B}{|e|}\ln\left(\frac{2-\nu}{\nu}\right), \tag{11}$$

where ν is the relative occupancy of the localized state band. It should be noted that $S_a(T)$ = const in the latter case.

It is found that the model of thermopower (8)–(10) enables us to correctly describe the experimental data on $S(T)$ only in the case when $S_a(T)$ = const (see Fig. 2a). Taking into account the proposed structure of carbynes synthesized under pressure (Fig. 5a), we can expect that $X \sim 0.5$ in the samples under investigation. Using this value, we calculated the temperature dependences of the contributions $S_M(T)$, $S_h(T)$, and S_a to the integrated curve $S(T)$ (an example of separation of contributions for a quasi-two-dimensional sample with $n = 1/3$ and $X = 0.5$ is shown in Fig. 2b). It turned out that, in all the samples under investigation, the thermopower of the "metal" is positive ($S_M(T) > 0$), while the thermopower of the "dielectric" is negative (see Fig. 2b), which causes the sign reversal on the dependence $S(T)$ (see Fig. 2a).

The hole-type conductivity of the "metal" is a natural consequence of the model considered above (see Fig. 5a) since sp^2 centers and, hence, charge carriers are concentrated in the localization region. In this case, almost all energy levels in the volume bounded by the localization radius are filled and conductivity becomes of the hole type (see Fig. 5c). At the same time, σ_a and S_a are of the electron type since these quantities describe the activation of electrons to the upper (delocalized) Hubbard band both from the lower Hubbard band in the "dielectric" and from the metallic region (see Fig. 5c).

It should be noted that the energy level diagram shown in Fig. 5c assumes that $\nu \sim 1$ and $S_a < 0$ in accordance with formula (11). Obviously, the activation energy E_a (see table) in this model specifies the distance between the lower and upper Hubbard bands (see Fig. 5c). Electron jumps between localized states also correspond to the electron transport, and the inequality $S_h < 0$ holds under the additional constraint $dg(E_F)/dE > 0$. Thus, the model of hopping conductivity and carbyne structure proposed by us here (see Fig. 5) makes it possible to interpret experimental data on $\sigma(T)$ and $S(T)$ (see Figs. 1 and 2).

The concepts developed by us here provide a qualitative explanation of the anomalous absence of negative magnetoresistance (see Fig. 4). Indeed, the cylindrical volume connecting the localized states i and j separated by the length of the jump R_{ij} contains $\pi a^2 R_{ij}/L_{\text{cor}}^3 \sim \pi a^3 (T_0/T)^n/L_{\text{cor}}^3$ centers at which quantum interference may take place. At $T = 4.2$ K, the number of such scattering centers lies between 1.6×10^3 and 1.7×10^4 (see the data from the table) and is too large to ensure the coherent mechanism since the phases of the waves emerging as a result of reflections from a large number of randomly distributed centers are also random. Note that the number of centers at which interference may take place in conventional doped semiconductors does not exceed a few dozen [10]. At the same time, in the case of carbynes synthesized under pressure, the incoherent mechanism can also be discarded since it is realized only in systems with a narrow band of width ≤ 1 meV [10], while disordered quasi-amorphous structures (see Fig. 5) are characterized by broad singularities in the DoS. Consequently, in the model of an effective medium presented in Fig. 5, we can expect a considerable suppression of negative magnetoresistance, which is indeed observed in experiments (see Fig. 4).

5. CONCLUSIONS

Thus, an analysis of conductivity, thermopower, and magnetoresistance of carbynes taking into account the results of X-ray structural measurements enabled us to determine the parameters of localized states and confirmed the results obtained by us earlier [5] and indicating the effect of the conditions of synthesis on the effective dimension of the space in which hopping takes place. Our results can be explained in the framework of the model of carbyne structure, assuming a considerable nonuniformity of the distribution of carbon atoms in the sp^2 state on the nanometer scale. This leads to a multicentered and multielectron form of localization of electron states, between which hopping takes place. The advantages of the proposed model include the possibility of unified quantitative interpretation of the results on conductivity and thermopower (Figs. 1–3) as well as of fine kinetic features such as the suppression of negative magnetoresistance in the hopping region (Fig. 4). The experimentally observed change in the effective dimensionality of conductivity reflects the change in the topology of sp^2 regions in the sp matrix of carbynes.

It should be emphasized that the hypothesis formulated above is mainly based on an analysis of the structure-sensitive transport characteristics and requires confirmation by the data on the sample structure. A detailed analysis of the carbyne structure will be the subject of future studies.

ACKNOWLEDGMENTS

We are grateful to the late Yu.P. Kudryavtsev who supplied the powder of initial carbynes for our experiments.

This work was financed by the Russian Foundation for Basic Research (project nos. 00-02-16403, 01-02-17543, 01-02-06193, 01-02-06093-MAS); INTAS grant no. 00-807, and YSF00112; the Ministry of Industry, Science, and Technology under the program “Fullerenes and Atomic Clusters,” Physics of Microwaves,” and “Fundamental Spectroscopy”; and also, in part, by the program “Low-Dimensional Quantum Structures,” Russian Academy of Sciences.

REFERENCES

1. Yu. P. Kudryavtsev, S. E. Evsyukov, M. V. Guseva, *et al.*, Izv. Akad. Nauk, Ser. Khim., No. 3, 450 (1993).
2. B. M. Bulychev and I. A. Udod, Ross. Khim. Zh. **39**, 9 (1995).
3. T. D. Varfolomeeva, S. V. Popova, A. G. Lyapin, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **66**, 237 (1997) [JETP Lett. **66**, 255 (1997)].
4. A. G. Lyapin, V. V. Brazhkin, S. G. Lyapin, *et al.*, Phys. Status Solidi B **211**, 401 (1999).
5. S. V. Demishev, A. A. Pronin, N. E. Sluchanko, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **72**, 547 (2000) [JETP Lett. **72**, 381 (2000)].
6. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1979; Mir, Moscow, 1982), Vol. 1.
7. B. I. Shklovskiĭ and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
8. I. P. Zvyagin, *Kinetic Phenomena in Disordered Semiconductors* (Mosk. Gos. Univ., Moscow, 1984).
9. S. V. Demishev, D. G. Lunts, A. G. Lyapin, *et al.*, Zh. Éksp. Teor. Fiz. **110**, 334 (1996) [JETP **83**, 180 (1996)].
10. M. E. Raikh, J. Czingon, Qiu-yi Ye, *et al.*, Phys. Rev. B **45**, 6015 (1992).
11. A. Hunt, Solid State Commun. **86**, 765 (1993).
12. A. Hunt, in *Hopping and Related Phenomena 5: Proceedings of the 5th International Conference on Hopping and Related Phenomena,* Ed. by C. J. Adkins, A. R. Long, and J. A. McInnes (World Sci., Singapore, 1994), p. 65.
13. S. V. Demishev, Yu. V. Kosichkin, N. E. Sluchanko, and A. G. Lyapin, Usp. Fiz. Nauk **164**, 195 (1994) [Phys. Usp. **37**, 185 (1994)].
14. S. V. Demishev, M. V. Kondrin, A. A. Pronin, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **68**, 801 (1998) [JETP Lett. **68**, 842 (1998)].
15. S. V. Demishev, A. A. Pronin, M. V. Kondrin, *et al.*, Phys. Status Solidi B **218**, 67 (2000).
16. P. M. Chaikin and G. Beni, Phys. Rev. B **13** (2), 647 (1976).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 132–144.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 150–163.
Original English Text Copyright © 2002 by Burmistrov.

SOLIDS
Electronic Properties

Two-Dimensional Electron Liquid with Disorder in a Weak Magnetic Field[¶]

I. S. Burmistrov
Landau Institute for Theoretical Physics, Russian Academy of Sciences, ul. Kosygina 2, Moscow, 117940 Russia
E-mail: burmi@itp.ac.ru

Received January 24, 2002

Abstract—We present the effective theory for the low-energy dynamics of two-dimensional interacting electrons in the presence of a weak short-range disorder and a weak perpendicular magnetic field, with the filling factor $\nu \gg 1$. We investigate the exchange enhancement of the g factor, the effective mass, and the decay rate of the simplest spin wave excitations at $\nu = 2N + 1$. We obtain the enhancement of the field-induced gap in the tunneling density of states and the dependence of the tunneling conductivity on the applied bias. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A two-dimensional electron gas in a perpendicular magnetic field has attracted much attention from both theoretical and experimental standpoints. The effects in a strong magnetic field when only the lowest Landau level is occupied have been investigated since the discovery of the quantum Hall effect [1]. Several efforts [2] are made in order to involve larger filling factors $\nu > 1$ in the problem discussed. However, the existence of a small parameter, the ratio of the Coulomb energy at the magnetic field length to the cyclotron energy, has been assumed. In a weak magnetic field, the Coulomb energy, at the magnetic field length actually exceeds the cyclotron energy and some attempts [3] have been undertaken to investigate the case of the large filling factor $\nu > 1$.

Experimental investigations of the tunneling density of states for the system under consideration were performed at small ($\nu < 1$) [4] and large ($\nu > 1$) [5] filling factors. In the case of a weak magnetic field ($\nu \gg 1$), the gap in the tunneling density of states was obtained in the framework of the hydrodynamical approach [6]. Progress was made by Aleiner and Glazman [7], who developed the effective theory for low-energy excitations on a partially filled Landau level at large filling factors $\nu \gg 1$.

Recently, after the prediction that the unidirectional charge-density wave state occurs at half-filled high Landau levels within the framework of the Hartree–Fock theory [8] and the experimental discovery of compressible states with anisotropic magnetotransport properties in high-mobility systems near the half-filling of the high Landau levels [9], the two-dimensional electron liquid in a weak magnetic field was intensely studied [10].

In this paper, we develop the low-energy effective theory for electrons at the partially filled Landau level with a large filling factor in the presence of disorder (Section 2). As an example, the effect of disorder on the exchange enhancement of the effective g factor and the simplest spin-wave excitations are discussed in Section 4. Electron tunneling into the electron liquid is considered in Section 5. Conclusions are given in Section 6.

2. DERIVATION OF THE EFFECTIVE ACTION

2.1. Introduction

We consider the system of a two-dimensional electrons with the Coulomb interaction in the presence of disorder in a perpendicular magnetic field H. The system possesses a partially filled high Landau level with the level index $N \gg 1$ equal to the integer part of half the filling factor ν, $N = [\nu/2]$. The presence of a random potential, which is considered to be short-range, results in a broadening of the Landau levels. We assume that the elastic collision time satisfies the condition

$$\tau_0 \gg \omega_c^{-1},$$

where $\omega_c = eH/m$ is the cyclotron frequency with the electron charge e and the electron mass m. In this case, the broadening of Landau levels, which is of the order $\sqrt{\omega_c\tau_0}/\tau_0$, is much less than the distance between them.

The conventional parameter characterizing the coupling strength of the Coulomb interaction is

$$r_s = \sqrt{2}e^2/v_F,$$

where v_F is the Fermi velocity. We assume that electrons are weakly interacting, i.e., $r_s < 1$. In this case, we can treat the problem in the random phase approximation. We also assume that the number N is sufficiently

---

[¶] This article was submitted by the author in English.

large, and the condition $Nr_s \gg 1$ is therefore satisfied. This means that the cyclotron radius $R_c = \sqrt{\nu/m\omega_c}$ is supposed to be much larger than the Bohr radius $a_B = 1/me^2$,

$$R_c \gg a_B,$$

2.2. The Formalism

The system is described by the grand canonical partition function in the path-integral representation,

$$\begin{aligned} Z = &\int \mathcal{D}[\bar{\psi}, \psi] \int \mathcal{D}[V_{\text{dis}}] \mathcal{P}[V_{\text{dis}}(\mathbf{r})] \\ &\times \exp\{S[\bar{\psi}, \psi, V_{\text{dis}}]\}, \end{aligned} \tag{1}$$

$$\begin{aligned} S = &\sum_{\alpha=1}^{N_r} \int_0^{1/T} d\tau \\ &\times \int d\mathbf{r} \Big[\bar{\psi}^{\alpha,\sigma}(\mathbf{r},\tau)(-\partial_\tau + \mu - \mathcal{H} - V_{\text{dis}}(\mathbf{r}))\psi^{\alpha,\sigma}(\mathbf{r},\tau) \\ &- \frac{1}{2}\int d\mathbf{r}_1 \bar{\psi}^{\alpha,\sigma}(\mathbf{r},\tau)\psi^{\alpha,\sigma}(\mathbf{r},\tau)U_0(\mathbf{r},\mathbf{r}_1) \\ &\times \bar{\psi}^{\alpha,\sigma_1}(\mathbf{r}_1,\tau)\psi^{\alpha,\sigma_1}(\mathbf{r}_1,\tau) \Big], \end{aligned} \tag{2}$$

where the Grassmann variables $\psi^{\alpha,\sigma}$ and $\bar{\psi}^{\alpha,\sigma}$ are defined on the imaginary time interval $\tau \in [0, 1/T]$ with the antiperiodic condition $\psi(\mathbf{r}, 1/T) = -\psi(\mathbf{r}, 0)$. The symbol T stands for the temperature, μ is the chemical potential of the system, and $\sigma, \sigma_1 = \pm 1$ are spin indices. The Hamiltonian

$$\mathcal{H} = \frac{(-i\nabla - e\mathbf{A})^2}{2m}$$

describes the electron with mass m propagating in the two-dimensional space in the perpendicular magnetic field

$$H = \epsilon_{ab}\partial_a A_b.$$

The random potential $V_{\text{dis}}(\mathbf{r})$ is chosen to have the Gaussian distribution function

$$\mathcal{P}[V_{\text{dis}}(\mathbf{r})] = \sqrt{\rho\tau_0}\exp\left(-\pi\rho\tau_0\int d\mathbf{r} V_{\text{dis}}^2(\mathbf{r})\right), \tag{3}$$

where ρ denotes the thermodynamical density of states.

To average $\ln Z$ over disorder, N_r replicated copies of the system are introduced; we let $\alpha = 1, \ldots, N_r$ be the replica indices.

The Matsubara representation seems to be more convenient for the above problem. We therefore use the Fourier transform from the imaginary time τ to the Matsubara frequencies. Because the fermionic fields are antiperiodic within the interval [0, 1/T], the frequencies permitted for $\bar{\psi}$ and ψ are

$$\omega_n = \pi T(2n+1),$$

where n is an integer. The Fourier-transformed fields are defined as

$$\begin{aligned} \bar{\psi}^\alpha(\tau) &= \sqrt{T}\sum_{n=-\infty}^{\infty} \bar{\psi}_n^\alpha e^{i\omega_n\tau}, \\ \psi^\alpha(\tau) &= \sqrt{T}\sum_{n=-\infty}^{\infty} \psi_n^\alpha e^{-i\omega_n\tau}. \end{aligned} \tag{4}$$

In what follows, we omit the limits in the frequency and replica series for brevity.

In the Matsubara representation, action (2) becomes

$$\begin{aligned} S = &\int d\mathbf{r} \\ &\times \sum_{\alpha,n} \Big[\bar{\psi}_n^{\alpha,\sigma}(\mathbf{r})(i\omega_n + \mu - \mathcal{H} - V_{\text{dis}}(\mathbf{r}))\psi_n^{\alpha,\sigma}(\mathbf{r}) \\ &- \frac{T}{2}\sum_{l,m}\int d\mathbf{r}_1 \bar{\psi}_m^{\alpha,\sigma}(\mathbf{r})\psi_{m-n}^{\alpha,\sigma}(\mathbf{r})U_0(\mathbf{r},\mathbf{r}_1) \\ &\times \bar{\psi}_l^{\alpha,\sigma_1}(\mathbf{r}_1)\psi_{l+n}^{\alpha,\sigma_1}(\mathbf{r}_1) \Big]. \end{aligned} \tag{5}$$

The Zeeman term in action (2) is neglected because the g factor is small. In fact, the condition $g \ll 1$ is usually satisfied. The Zeeman term can nevertheless, be included in the effective action after performing integration over fast degrees of freedom. To simplify the notation, the spin indices are associated with the replica ones whenever convenient.

2.3. The Plasmon Field and the Average over Disorder

The Coulomb term entering action (5) is quartic in the fermionic fields. This quartic term can be eliminated by the Hubbard–Stratonovich transformation, introducing an extra path integration over bosonic fields $\lambda_n^\alpha(\mathbf{r})$. With the help of the so-called plasmon field, the Coulomb term can be presented as

$$\begin{aligned} \int \mathcal{D}[\lambda]\exp\Big[&-\frac{T}{2}\iint d\mathbf{r}\,d\mathbf{r}_1 \lambda^\dagger(\mathbf{r})U_0^{-1}(\mathbf{r},\mathbf{r}_1)\lambda(\mathbf{r}_1) \\ &+ iT\int d\mathbf{r}\,\psi^\dagger(\mathbf{r})\hat{\lambda}(\mathbf{r})\psi(\mathbf{r}) \Big], \end{aligned} \tag{6}$$

where U_0^{-1} stands for the inverse operator to U_0. The matrix notation is used for the combined replica and frequency indices,

$$\begin{aligned}\psi^\dagger(\dots)\psi &= \sum_{n,m}^{\alpha,\beta}\bar{\psi}_n^\alpha(\dots)_{nm}^{\alpha\beta}\psi_m^\beta,\\ \lambda^\dagger\lambda &= \sum_n^\alpha\lambda_{-n}^\alpha\lambda_n^\alpha.\end{aligned}\tag{7}$$

The quantities with the hat are defined as

$$\hat{z} = \sum_{\alpha,n} z_n^\alpha I_n^\alpha$$

with the matrix

$$(I_n^\alpha)_{kl}^{\beta\gamma} = \delta^{\alpha\beta}\delta^{\alpha\gamma}\delta_{k-l,n}.$$

The matrices I_n^α represent the diagonals shifted in the frequency space; they are the generators of the $U(1)$ gauge transformations in general. The measure of the path integral over the plasmon field λ is introduced such that integral (6) equals unity for the vanishing fermionic fields $\psi^\dagger$ and ψ.

In order to perform the averaging over disorder in partition function (1), we must integrate over the random potential $V_{\mathrm{dis}}(\mathbf{r})$. This leads to the quartic term

$$\frac{1}{4\pi\rho\tau}\int d\mathbf{r}\sum_{n,m}^{\alpha,\beta}\bar{\psi}_n^\alpha(\mathbf{r})\psi_n^\alpha(\mathbf{r})\bar{\psi}_m^\beta(\mathbf{r})\psi_m^\beta(\mathbf{r})\tag{8}$$

in the action. This term can be decoupled by the Hubbard–Stratonovich transformation. An extra path integration over the Hermitian matrix field variables $Q_{nm}^{\alpha\beta}(\mathbf{r})$ can be introduced [11, 12],

$$\begin{aligned}&\int\mathcal{D}[Q]\\ &\times\exp\int d\mathbf{r}[-\pi\rho\tau_0\mathrm{tr}Q^2(\mathbf{r}) + i\psi^\dagger(\mathbf{r})Q(\mathbf{r})\psi(\mathbf{r})],\end{aligned}\tag{9}$$

where the symbol "tr" denotes the matrix trace over the Matsubara, replica, and spin spaces. The measure of the path integral over the matrix field Q is defined in the same way as for the plasmon field; i.e., integral (9) equals unity for vanishing fermionic fields $\psi^\dagger$ and ψ.

After the above calculations, the partition function becomes

$$Z = \int\mathcal{D}[\bar{\psi},\psi,\lambda,Q]\exp\{S[\bar{\psi},\psi,\lambda,Q]\},\tag{10}$$

$$\begin{aligned}S &= -\pi\rho\tau_0\int d\mathbf{r}\,\mathrm{tr}Q^2\\ &-\frac{T}{2}\iint d\mathbf{r}d\mathbf{r}_1\lambda^\dagger(\mathbf{r})U_0^{-1}(\mathbf{r},\mathbf{r}_1)\lambda(\mathbf{r}_1)\\ &+\int d\mathbf{r}\psi^\dagger(\mathbf{r})(i\omega+\mu-\hat{\mathcal{H}}+iT\hat{\lambda}+iQ)\psi(\mathbf{r}),\end{aligned}\tag{11}$$

where ω is the unit matrix in the replica space, while, in the Matsubara space, it is a matrix containing the frequencies ω_n on the diagonal,

$$(\omega)_{nm}^{\alpha\beta} = \omega_n\delta_{nm}\delta^{\alpha\beta}.$$

2.4. Elimination of the Nth Landau Level

The fermionic fields $\psi^\dagger$ and ψ refer to all Landau levels. In order to integrate over all fermionic degrees of freedom not belonging to the partially filled Nth Landau level, we separate the fermionic fields into two kinds. The first field refers to the Nth Landau level,

$$\begin{aligned}\Psi(\mathbf{r}) &= \sum_k\psi_{Nk}\varphi_{Nk}(\mathbf{r}),\\ \Psi^\dagger(\mathbf{r}) &= \sum_k\psi_{Nk}^\dagger\varphi_{Nk}(\mathbf{r}).\end{aligned}\tag{12}$$

The second one involves the other levels,

$$\begin{aligned}\Phi(\mathbf{r}) &= \sum_{p\neq N,k}\psi_{pk}\varphi_{pk}(\mathbf{r}),\\ \Phi^\dagger(\mathbf{r}) &= \sum_{p\neq N,k}\psi_{pk}^\dagger\varphi_{pk}(\mathbf{r}),\end{aligned}\tag{13}$$

where $\varphi_{pk}(\mathbf{r})$ are the eigenfunctions of the Hamiltonian $\mathcal{H}$ and $p = 0, 1, \dots, N, \dots$ labels Landau levels with the energies $\epsilon_p = \omega_c(p + 1/2)$. In addition, we introduce two types of Green's functions. One is for the Nth Landau level,

$$G(\mathbf{r},\mathbf{r}_1;Q,\lambda) = \sum_{k,k'}\varphi_{Nk}^*(\mathbf{r})G_{Nk,Nk'}(Q,\lambda)\varphi_{Nk'}(\mathbf{r}_1),\tag{14}$$

and the other is for the other levels,

$$\begin{aligned}&\tilde{G}(\mathbf{r},\mathbf{r}_1;Q,\lambda)\\ &= \sum_{p,p'\neq N}\sum_{k,k'}\varphi_{pk}^*(\mathbf{r})G_{pk,p'k'}(Q,\lambda)\varphi_{p'k'}(\mathbf{r}_1),\end{aligned}\tag{15}$$

where the inverse of the Green's function for the ψ_{pk} and $\psi_{p'k'}^\dagger$ operators is given by

$$\begin{aligned}(G^{-1})_{pk,p'k'} &= (i\omega+\mu-\epsilon_p)\delta_{pp'}\delta_{kk'}\\ &+ iT\hat{\lambda}_{pk,p'k'} + iQ_{pk,p'k'}\end{aligned}\tag{16}$$

with the matrix elements

$$f_{pk,p'k'} = \int d\mathbf{r}\varphi_{p'k'}^*(\mathbf{r})f(\mathbf{r})\varphi_{pk}(\mathbf{r}).\tag{17}$$

The action (11) is bilinear in the fermionic fields $\psi^\dagger$ and ψ, and obviously, also in the fermionic fields $\Phi^\dagger$

and Φ. We can therefore integrate over the fermionic fields $\Phi^\dagger$ and Φ; this gives

$$\begin{aligned} S = & -\int \operatorname{tr}\ln\tilde{G} - \pi\rho\tau_0\int \operatorname{tr}Q^2 \\ & + \int \Psi^\dagger[i\omega + \mu - \hat{\mathcal{H}} + iT\hat{\lambda} + iQ]\Psi \\ & - \frac{T}{2}\iint \lambda^\dagger U_0^{-1}\lambda \\ & + \iint [\Psi^\dagger Q\tilde{G}Q\Psi + 2T\Psi^\dagger\hat{\lambda}\tilde{G}Q\Psi + T^2\Psi^\dagger\hat{\lambda}\tilde{G}\hat{\lambda}\Psi]. \end{aligned} \tag{18}$$

Hereafter, the space indices are omitted. It should be noted that the last term arises in action (18) due to the interaction between electrons belonging to the partially filled Nth Landau level and the other electrons.

2.5. Integration over the Q Field

The Q matrix field must be divided into the transverse V and the longitudinal P components as $Q = V^{-1}PV$. Here, the longitudinal component P has a block-diagonal structure in the Matsubara space, i.e., $P^{\alpha\beta}_{nm} \propto \Theta(nm)$, where $\Theta(x)$ is the Heaviside step function. The transverse component V corresponds to a unitary rotation (see [13, 14] for a review).

The change of variables from Q to P and V is motivated by the saddle-point structure of action (18) in the absence of the plasmon field λ and at zero temperature, i.e., as $\omega_n \longrightarrow 0$. This saddle-point solution can be written as

$$Q_{sp} = V^{-1}P_{sp}V,$$

where the matrix P_{sp} obeys the equation

$$2\pi\rho\tau_0 P_{sp} = i[G_0(\mathbf{r},\mathbf{r}) + \tilde{G}_0(\mathbf{r},\mathbf{r})] \tag{19}$$

that coincides with the self-consistent Born approximation equation [15]. Here, the Green's function G_0 is a special case of G, namely,

$$G_0(\mathbf{r},\mathbf{r}_1) = G(\mathbf{r},\mathbf{r}_1; P_{sp}, 0),$$

and similarly for $\tilde{G}_0$.

In the case of small disorder, $\omega_c\tau_0 \gg 1$, the solution to Eq. (19) is given by

$$(P_{sp})^{\alpha\beta}_{nm} = \frac{\operatorname{sgn}n}{2\tau}\delta_{nm}\delta^{\alpha\beta}, \quad \tau = \pi\sqrt{\frac{\rho}{m}\frac{\tau_0}{\sqrt{\omega_c\tau_0}}}. \tag{20}$$

The presence of the plasmon field λ results in a shift of the saddle-point value (20) of the P field; this shift can be found by expanding action (18) to the second order in both λ and $\delta P = P - P_{sp}$. We thus obtain

$$S = S_0 + S_1[\delta P, \lambda] + S_2[\overline{\Psi}, \Psi, \delta P, \lambda], \tag{21}$$

$$\begin{aligned} S_0 = & \int(-\operatorname{tr}\ln\tilde{G}_0 - \pi\rho\tau_0\operatorname{tr}Q^2_{sp} \\ & + \Psi^\dagger[i\omega + \mu - \hat{\mathcal{H}} + iT\hat{\lambda} + iQ_{sp}]\Psi), \end{aligned} \tag{22}$$

$$\begin{aligned} S_1 = & \; iT\int \operatorname{tr}\tilde{G}_0\hat{\lambda} - \pi\rho\tau_0\int \operatorname{tr}(\delta P)^2 - \frac{T}{2}\iint \lambda^\dagger U_0^{-1}\lambda \\ & + \frac{1}{2}\iint \operatorname{tr}[T\hat{\lambda} + \delta P]\pi_0[T\hat{\lambda} + \delta P], \end{aligned} \tag{23}$$

$$\begin{aligned} S_2 = & \iint(\Psi^\dagger[T\hat{\lambda} + \delta P]\tilde{G}_0[T\hat{\lambda} + \delta P]\Psi \\ & - 2\operatorname{tr}[T\hat{\lambda} + \delta P]\tilde{G}_0[T\hat{\lambda} + \delta P]G_0), \end{aligned} \tag{24}$$

where the bare polarization operator π_0 is understood to be a matrix in accordance with the rule

$$\begin{aligned} & \operatorname{tr}A\pi_0 B \\ & = \sum_{n,m}^{\alpha,\beta} A^{\alpha\beta}_{m+n,m}(\mathbf{r})\pi_0^m(n; \mathbf{r},\mathbf{r}_1)B^{\beta\alpha}_{m,m+n}(\mathbf{r}_1) \end{aligned} \tag{25}$$

and is defined by

$$\begin{aligned} \pi_0^m(n; \mathbf{r},\mathbf{r}_1) = & -2(\tilde{G}_0^{m+n}(\mathbf{r},\mathbf{r}_1)\tilde{G}_0^m(\mathbf{r}_1,\mathbf{r}) \\ & + \tilde{G}_0^{m+n}(\mathbf{r},\mathbf{r}_1)G_0^m(\mathbf{r}_1,\mathbf{r}) + G_0^{m+n}(\mathbf{r},\mathbf{r}_1)\tilde{G}_0^m(\mathbf{r}_1,\mathbf{r})). \end{aligned} \tag{26}$$

After decomposing the matrix field Q into the block-diagonal Hermitian matrix field P and the unitary matrix field V, the measure of the functional integral in (21) becomes

$$\mathcal{D}[Q] = \mathcal{D}[V]\mathcal{D}[\delta P]I[\delta P],$$

where [13]

$$\begin{aligned} \ln I[\delta P] = & -\frac{1}{(\pi\rho)^2} \\ & \times\int\sum_{n,m}^{\alpha,\beta}[1 - \Theta(nm)]\delta P^{\alpha\alpha}_{nn}\delta P^{\beta\beta}_{mm}. \end{aligned} \tag{27}$$

The terms that are quadratic in δP in the part S_1 of action (21) together with the contribution of measure (27) determine the propagator of the δP fields,

$$\begin{aligned} & \langle\delta P^{\alpha\beta}_{m_1m_2}(\mathbf{q})\delta P^{\gamma\delta}_{m_3m_4}(-\mathbf{q})\rangle \\ & = \frac{\delta_{m_1m_4}\delta_{m_2m_3}\delta^{\alpha\delta}\delta^{\beta\gamma}\dfrac{\Theta(m_1m_3)}{2\pi\rho\tau_0}}{1 + \dfrac{\pi_0^{m_1}(m_3 - m_1; \mathbf{q})}{2\pi\rho\tau_0}} - \frac{2[1 - \Theta(m_1m_3)]}{(2\pi^2\rho^2\tau_0)^2} \\ & \times\frac{\delta_{m_1m_2}\delta^{\alpha\beta}}{1 + \dfrac{\pi_0^{m_1}(0; \mathbf{q})}{2\pi\rho\tau_0}}\frac{\delta_{m_3m_4}\delta^{\delta\gamma}}{1 + \dfrac{\pi_0^{m_3}(0; \mathbf{q})}{2\pi\rho\tau_0}}. \end{aligned} \tag{28}$$

We note that the propagator of the longitudinal fluctuations (28) proves to be analogous to that previously obtained in the problem of the behavior of a free electron gas in the perpendicular magnetic field [13].

Using expression (28) for the propagator of the δP fields, we can integrate action (21) over the longitudinal fluctuations in the quadratic approximation. This gives

$$S = S_0 + S_\lambda + S_\mu, \tag{29}$$

where S_0 given by Eq. (22) describes the electrons at the partially filled Nth Landau level coupled to the plasmon and Q_{sp} fields. The term S_λ corresponds to the screening of the Coulomb interaction due to the influence of electrons from the other Landau levels and is given by

$$\begin{aligned} S_\lambda &= iT\int d\mathbf{r}\,\mathrm{tr}\,\tilde{G}_0(\mathbf{r},\mathbf{r})\hat{\lambda}(\mathbf{r}) \\ &- \frac{T}{2}\int\frac{d\mathbf{q}}{(2\pi)^2}\sum_n^\alpha \lambda^\alpha_{-n}(\mathbf{q})U_0^{-1}(q)\varepsilon(n,q)\lambda^\alpha_n(-\mathbf{q}), \end{aligned} \tag{30}$$

where the dielectric function is given by

$$\varepsilon(n,q) = 1 + U_0(q)\Pi(n,q)$$

with the polarization operator[1]

$$\begin{aligned} \Pi(n,q) &= T\sum_m \pi_0^m(n,q)\left[1 - \frac{\Theta(n(n+m))\dfrac{\pi_0^m(n,q)}{2\pi\rho\tau_0}}{1+\dfrac{\pi_0^m(n,q)}{2\pi\rho\tau_0}}\right] \\ &+ T\frac{\delta_{n,0}}{(\pi^2\rho^2\tau)^2}\sum_{k,m}\frac{[1-\Theta(km)]\pi_0^m(0,q)}{1+\dfrac{\pi_0^m(0,q)}{2\pi\rho\tau_0}} \\ &\times\frac{\pi_0^k(0,q)}{1+\dfrac{\pi_0^k(0,q)}{2\pi\rho\tau_0}}. \end{aligned} \tag{31}$$

The third term S_μ in action (29) contains the terms that affect the chemical and thermodynamic potentials of the system (see Appendix A).

It is worthwhile to mention that the saddle-point approximation in which the integration over the Q field is performed is valid because the condition

$$\mu\tau = N\omega_c\tau \gg 1$$

is satisfied.

2.6. Integration over the Plasmon Field

As a final step of the procedure, action (29) must be integrated over the plasmon field λ. The integration can be performed in the quadratic approximation in the λ fields. The corresponding propagator is determined by the second term in Eq. (30). After that, we obtain the effective action for electrons on the partially filled Landau level,

$$\begin{aligned} S_{\mathrm{eff}} &= -\frac{\Omega}{T} + \int d\mathbf{r}\Psi^\dagger(\mathbf{r})[i\omega + \tilde{\mu} - \hat{\mathcal{H}} + iQ_{sp}]\Psi(\mathbf{r}) \\ &- \pi\rho\tau_0\int d\mathbf{r}\,\mathrm{tr}\,Q_{sp}^2(\mathbf{r}) - \frac{T}{2}\iint d\mathbf{r}\,d\mathbf{r}_1\sum_{n,m,k}\bar{\Psi}_m^{\alpha,\sigma}(\mathbf{r}) \\ &\times\Psi^{\alpha,\sigma}_{m+n}(\mathbf{r})U_{\mathrm{eff}}(\mathbf{r}-\mathbf{r}_1)\bar{\Psi}_k^{\alpha,\sigma_1}(\mathbf{r}_1)\Psi^{\alpha,\sigma_1}_{k-n}(\mathbf{r}_1) \\ &+ \frac{g\omega_c}{2}\int d\mathbf{r}\sum_n^\alpha \sigma\bar{\Psi}_n^{\alpha,\sigma}(\mathbf{r})\Psi_n^{\alpha,\sigma}(\mathbf{r}), \end{aligned} \tag{32}$$

which is the main result of the paper.

We have incorporated the Zeeman term into the effective action. The Fourier transform of the effective interaction potential

$$U_{\mathrm{eff}}(q) = \frac{U_0(q)}{\varepsilon(q)}$$

is determined by the static dielectric function $\varepsilon(q) \equiv \varepsilon(0,q)$. In general, the low-energy properties of the system under consideration can be described with the help of the retarded interaction alone (see action (29)). However, the description within the framework of the effective action with the instantaneous interaction seems to be a rather good approximation in this problem [7]. This is because transitions between the Landau levels have a characteristic time scale about ω_c^{-1}, while the typical energy scale in the effective theory is of the order of the exchange energy $\Delta_{\mathrm{ex}} \ll \omega_c$ (see Section 4).

The existence of the other Landau levels except the partially filled Nth Landau level affects both the thermodynamic and the chemical potentials. The thermodynamic potential Ω in action (32) can be represented as

$$\Omega = \Omega_0 + \Delta\Omega, \tag{33}$$

where

$$\Omega_0 = T\int d\mathbf{r}\,\mathrm{tr}\ln\tilde{G}_0(\mathbf{r},\mathbf{r}) \tag{34}$$

is the thermodynamic potential of the system of noninteracting electrons for the completely filled Landau levels in the presence of disorder and the quantity $\Delta\Omega$ is analogous to the first-order exchange and correlation corrections equivalent to the sum of ring diagrams contributing to the ground state energy of a clean electron liquid [17],

$$\Delta\Omega = \frac{T}{2}\int d\mathbf{r}\sum_n\int\frac{d\mathbf{q}}{(2\pi)^2}\ln\varepsilon(n,\mathbf{q}). \tag{35}$$

---

[1] A similar form of the polarization operator but with a different bare $\pi_0^m(n,q)$ was first derived by Baranov and Pruisken [16].

The chemical potential $\tilde{\mu}$ in action (32) can be written as

$$\tilde{\mu} = \mu + \delta\mu, \tag{36}$$

where the shift of the chemical potential

$$\delta\mu = 2\pi l^2 T \sum_n \int d\mathbf{r} \tilde{G}_0^n(0, \mathbf{r}) P_N(0, \mathbf{r}) U_{\text{eff}}(n, \mathbf{r}) \tag{37}$$

involves corrections similar to the exchange and correlation ones in a clean electron liquid. Here, $l = 1/\sqrt{m\omega_c}$ is the magnetic field length. The quantity $U_{\text{eff}}(n, \mathbf{r})$ is the Fourier transform of $U_0(q)/\varepsilon(n, \mathbf{q})$, and

$$P_N(\mathbf{r}_1, \mathbf{r}_2) = \sum_k \varphi^*_{Nk}(\mathbf{r}_2)\varphi_{Nk}(\mathbf{r}_1) \tag{38}$$

is the projection operator onto the partially filled Nth Landau level.

We note that corrections to the thermodynamic and chemical potentials contain additional terms except those presented above. They are neglected in the limit of a weak disorder $\omega_c\tau \gg 1$ (see Appendix A).

The integration over the plasmon field is performed in the Gaussian approximation. This can be justified if the fluctuations of the plasmon field are small. The long- and short-range fluctuations are different physically. In the case of a large length scale $r \gg R_c$, only the dipole transitions between the adjacent Landau levels are induced. The long-range fluctuations are small if the condition $Nr_s \gg 1$ is satisfied [7]. Physically, this condition means that the characteristic length scale R_c^2/a_B of the long-range fluctuations must be much greater than the cyclotron radius R_c. The short-range fluctuations correspond to the case of a small length scale $r \ll R_c$. Transitions between distant Landau levels are possible in this case. The condition $r_s \ll 1$ of the smallness of short-range fluctuations is just the criterion of perturbation theory applicability to Coulomb interaction.

3. EFFECTIVE INTERACTION, THE THERMODYNAMIC AND CHEMICAL POTENTIALS

The results of the previous section allow us to find effective action (32) for the electrons on the partially filled Nth Landau level. The main physical quantity that affects the dynamics of the electrons is the effective electron–electron interaction. It is completely determined by the static dielectric function $\varepsilon(q)$. The other two interesting quantities in effective action (32) are the thermodynamic and chemical potentials.

3.1. The Effective Interaction

The most pronounced effect of electrons on the completely filled Landau levels is the screening of the electron–electron interaction on a partially filled Landau level. This screening is determined by the static dielectric function $\varepsilon(q)$.

In accordance with Eq. (31) for the polarization operator $\Pi(n, q)$, the dielectric function can be obtained for arbitrary values of the disorder parameter $\omega_c\tau$. However, the situation of a small Landau level broadening due to disorder is most interesting from the physical standpoint. In this case, the expression for the static dielectric function can be simplified drastically,

$$\varepsilon(q) = 1 + \frac{2\pi e^2}{q} T \sum_n \pi_0^n(0, q), \quad \omega_c\tau \gg 1. \tag{39}$$

The evaluation of the static dielectric function is presented in Appendix B. The result can be written as

$$\varepsilon(q) = 1 + \frac{2}{qa_B}\left(1 - \mathcal{J}_0^2(qR_c) - \frac{\pi}{3\omega_c\tau}\eta(qR_c)\right), \tag{40}$$

where the function $\eta(x)$ can be found analytically only in the asymptotic regions of small and large values of x (see Appendix B),

$$\eta(x) = \begin{cases} \dfrac{\pi^2 x^2}{12}, & x \ll 1 \\[2ex] \dfrac{\ln^2(3.57x)}{\pi x}, & x \gg 1. \end{cases} \tag{41}$$

Here, $\mathcal{J}_0(x)$ is the Bessel function of the first kind, Expression (40) or the static dielectric function is the main result of the paper.

It is worthwhile to note that the asymptotic expressions (in the $qR_c \ll 1$ and $qR_c \gg 1$ domains) for the static dielectric function $\varepsilon(q)$ in a clean system ($\tau^{-1} = 0$) were obtained earlier by Kukushkin, Meshkov, and Timofeev [18]. The general expression for the static dielectric function in a clean system was derived by Aleiner and Glazman [7].

We mention that the asymptotic expressions for the static dielectric function in a clean system can be obtained from a clear physical picture [18, 7]. The behavior of the static dielectric function in the region $qR_c \ll 1$ can be explained by dipole transitions between the adjacent Landau levels. The result for the static dielectric function in the region $qR_c \gg 1$ is explained by the standard Thomas–Fermi screening. But there is no clear physical picture in the case of a weakly dirty system. We have no other opportunity to obtain the dielectric function except the derivation of the effective action for electrons on the partially filled Landau level.

It follows from (40) that in the domain $qR_c \ll 1$, the static dielectric function is given by

$$\varepsilon(q) = 1 + \left(1 - \frac{\pi}{6\omega_c\tau}\right)\frac{R_c^2 q}{a_B}. \quad (42)$$

This shows that the disorder suppresses the effect of the screening. We can expect that the screening decreases as disorder increases. We can estimate the disorder threshold τ^*, i.e., the point of vanishing screening, as $\omega_c\tau^* \sim 1/2\pi$.

From Eq. (40), we can obtain the expression for the static dielectric function in the domain $qR_c \gg 1$,

$$\varepsilon(q) = 1 + \frac{2}{qa_B}\left(1 - \frac{\ln^2(3.57qR_c)}{3\omega_c\tau qR_c}\right). \quad (43)$$

The disorder also suppresses the screening in the region of large wave vectors $qR_c \gg 1$.

Equations (40) and (41) allow us to obtain the asymptotic behavior of the effective interaction $U_{\text{eff}}(r)$ to the coordinate space. The polarization is insignificant for the very large length scale $r \gg R_c^2/a_B$, and the effective interaction coincides with the bare Coulomb interaction

$$U_{\text{eff}}(r) = \frac{e^2}{r}\left(1 - \frac{R_c^4}{a_B^2 r^2}\left[1 - \frac{\pi}{3\omega_c\tau}\right]\right). \quad (44)$$

At the intermediate scale $R_c^2/a_B \gg r \gg R_c$, the polarization becomes important and the effective interaction is given by

$$U_{\text{eff}}(r) = \frac{\omega_c}{2N\left(1 - \frac{\pi}{6\omega_c\tau}\right)} \times \ln\left(1 + \frac{R_c^2\left(1 - \frac{\pi}{6\omega_c\tau}\right)}{a_B r}\right). \quad (45)$$

We note that, while disorder increases, the effective interaction tends to the bare Coulomb interaction. For the small scale $R_c \gg r \gg a_B$, the Thomas–Fermi screening occurs and the effective interaction is given by

$$U_{\text{eff}}(r) = \frac{e^2 a_B^2}{r^3} + \frac{\pi}{3\omega_c\tau}\frac{e^2 a_B}{rR_c}\left[\ln^2\frac{1.31R_c}{r} + 1\right]. \quad (46)$$

We emphasize that disorder in the system most strongly affects the electron–electron interaction within the intermediate length scale $R_c^2/a_B \gg r \gg R_c$. Physically, this is the case where the dipole transitions between the adjacent Landau levels are possible.

3.2. The Thermodynamic and Chemical Potentials

The thermodynamic and chemical potentials in Eqs. (34)–(37) can be evaluated in the leading orders in $1/N$. The detailed calculations are presented in Appendix C.

The thermodynamic potential for the system of noninteracting electrons in the presence of disorder for the completely filled Landau levels is given by

$$\Omega_0 = \frac{L_x L_y}{\pi l^2}\left[\frac{N(N-1)}{2}\omega_c - \mu - \frac{\ln(2\omega_c\tau) - 1}{\pi\tau}\right], \quad (47)$$

where L_x and L_y are the sizes of the system. The first-order exchange correction to the thermodynamic potential is given by

$$\Delta\Omega = -\frac{L_x L_y}{\pi l^2}\frac{e^2}{\pi l}(2N)^{3/2}\left[\frac{2}{3} + \frac{2\ln 2}{\pi\omega_c\tau}\frac{1}{2N}\right]. \quad (48)$$

The presence of disorder changes the dependence of $\Delta\Omega$ on the magnetic field, i.e., on N. For the dirty system, the second term in brackets in Eq. (48) is proportional to $1/N$. This is in contrast to the clean system, where the correction is much smaller and is proportional to $1/N^2$ [7].

The shift of the chemical potential due to the exchange correction can be written as

$$\delta\mu = \frac{2e^2}{\pi l}(2N)^{1/2}\left[1 - \frac{\ln N}{8N} + \frac{1}{\pi\omega_c\tau}\frac{1}{2N}\right]. \quad (49)$$

We note that $\delta\mu$ contains only the exchange correction and does not involve the correlation correction due to normal ordering of the $\Psi^\dagger$ and Ψ fields (see [7]).

4. SPIN EXCITATIONS

In the previous section, we analyzed the renormalization of the electron–electron interaction on the partially filled Nth Landau level due to the existence of the other levels. In this section, we investigate the enhancement of the g factor and the simplest spin excitations at the filling factor $\nu = 2N + 1$.

The electrons on the partially filled Nth Landau level at the filling factor $\nu = 2N + 1$ possess the maximum spin in the ground state, because the ground state does not contain skyrmions at large ν [19]. This ground state is obviously fully spin-polarized and is described by the wave function

$$|N_{el} = N_\Phi, S_z = N_\Phi/2\rangle,$$

where N_{el} is the number of electrons on the partially filled Nth Landau level and

$$N_\Phi = \frac{L_x L_y}{2\pi l^2}$$

is the number of states on the Landau level. The simplest excitations are described by the state with the energy $E_\uparrow$ with an extra hole and the state with the energy $E_\downarrow$ with an extra electron. The width of the spin gap Δ_s is related to the energies of the excited states and to the energy E_0 of the ground state [20, 21, 7] as

$$\Delta_s = E_\uparrow + E_\downarrow - 2E_0.$$

We can find that the width of the spin gap equals

$$\Delta_s = \Delta_{\text{ex}} + g\omega_c,$$

where the shift of the chemical potential Δ_{ex} due to the exchange interaction [21, 22] is determined by

$$\Delta_{\text{ex}} = 2\pi l^2 \int d\mathbf{r} U_{\text{eff}}(\mathbf{r}) P_N(0, \mathbf{r}) P_N(\mathbf{r}, 0). \tag{50}$$

Using expression (38) for the projection operator P_N, we can evaluate the effective g factor. It is defined as $g_{\text{eff}} = \Delta_s/\omega_c$ and is given by

$$\begin{aligned} g_{\text{eff}} = g &+ \frac{r_s}{\pi\sqrt{2}} \ln\frac{2\sqrt{2}}{r_s}\left(1 - \frac{\pi}{6\omega_c\tau}\right) + \frac{E_h}{\omega_c} \\ &+ \frac{r_s}{\pi\omega_c\tau} \frac{\ln^2(14.28N)}{4\sqrt{2}\pi^2 N}, \end{aligned} \tag{51}$$

where the "hydrodynamic" term is

$$\begin{aligned} E_h = &\frac{\omega_c}{2N\left(1 - \frac{\pi}{6\omega_c\tau}\right)} \\ &\times \ln\left[1 + \sqrt{2} r_s N\left(1 - \frac{\pi}{6\omega_c\tau}\right)\right]. \end{aligned} \tag{52}$$

The disorder in the system results in the appearance of a strong dependence of the effective g factor on the magnitude of the magnetic field as $\ln^2 N/N$.

We now discuss the neutral excitations, spin waves [21, 23] at the filling factor $\nu = 2N + 1$. They are described by the wave function

$$\sum_q \exp(ik_x q l^2) \bar{\Psi}_{N, q, \downarrow} \Psi_{N, q - k_y, \uparrow} \left| N_\Phi, \frac{N_\Phi}{2} \right\rangle. \tag{53}$$

Following [21], we must take three contributions into account. They are the difference of the exchange self-energy of an electron in the excited Landau level and the self-energy in the level from which the electron was removed, the direct Coulomb interaction, and the exchange energy. We then obtain the equation that determines the spectrum of the spin wave excitations,

$$\begin{aligned} \omega = g\omega_c &+ \int \frac{d\mathbf{q}}{(2\pi)^2} \frac{U_0(q)}{\varepsilon(q, \omega)} \left[L_N\left(\frac{q^2 l^2}{2}\right)\right]^2 \\ &\times \exp(-q^2 l^2/2)(1 - \exp[i(\mathbf{k}\cdot\mathbf{q}) l^2]), \end{aligned} \tag{54}$$

where $L_N(x)$ is the Laguerre polynomial. The dielectric function $\varepsilon(q, \omega)$ contains the imaginary part (see Eq. (B.4)), which is of order $1/\omega_c\tau$. It results in the decay rate of the spin-wave excitations. Physically, the spin-wave excitations decay because of the scattering on impurities. We mention that the decay rate also appears in the magnetoplasmon spectrum.

The energy of the spin-wave excitations is much less than ω_c: $\omega(k) \ll \omega_c$. We can therefore calculate the real part $E_{SW}(k)$ and the imaginary part $\Gamma_{SW}(k)$ of the spin-wave energy separately. We set $\omega = 0$ on the right-hand side of Eq. (54). The evaluation of Eq. (54) then leads to a quadratic dispersion relation for the small wave vectors $kR_c \ll 1$,

$$\begin{aligned} E_{SW}(k) = g\omega_c &+ \frac{r_s\omega_c}{\pi\sqrt{2}} \\ \times \left[1 + \frac{r_s}{\sqrt{2}}\left(1 - \frac{1}{3\omega_c\tau} \frac{\ln^2(14.28N)}{4}\right)\right]^{-1} &(kR_c)^2. \end{aligned} \tag{55}$$

An additional dependence of the effective mass of the spin-wave excitations on the magnetic field appears because of the presence of disorder in the system. The disorder suppresses the effective mass of the spin-wave excitations. For sufficiently large wave vectors

$$1 \ll kR_c \ll R_c^2/l^2,$$

the energy of the spin wave is given by

$$\begin{aligned} E_{SW}(k) = \Delta_{\text{ex}} - E_h &- \frac{r_s\omega_c}{\pi\sqrt{2}} \\ \times \Bigg[\ln\left(1 + \frac{(\sqrt{2} r_s k R_c)^{-1}}{1 - \frac{1}{3\omega_c\tau} \frac{\ln^2(7.14N/kR_c)}{2N/kR_c}}\right) \\ + \frac{\sin 2kR_c}{2kR_c}\left(1 + \frac{r_s}{\sqrt{2}}\left(1 - \frac{1}{3\omega_c\tau} \frac{\ln^2(14.28N)}{4N}\right)\right)\Bigg]. \end{aligned} \tag{56}$$

To obtain the decay rate of spin-wave excitations, we take into account that the imaginary part ε'' of the dielectric function is small. We then obtain

$$\begin{aligned} \Gamma_{SW}(k) = -\int &\frac{d\mathbf{q}}{(2\pi)^2} \frac{U_0(q)\varepsilon''(q, E_{SW})}{\varepsilon_0^2(q, E_{SW})} \left[L_N\left(\frac{q^2 l^2}{2}\right)\right]^2 \\ &\times \exp(-q^2 l^2/2)(1 - \exp(i(\mathbf{k}\cdot\mathbf{q}) l^2)). \end{aligned} \tag{57}$$

The evaluation of Eq. (57) for small wave vectors $kR_c \ll 1$ yields

$$\Gamma_{SW} = -\frac{\arctan(2\omega_c\tau g)}{6\omega_c\tau}\frac{e^2}{a_B}(kR_c)^2 \times \frac{1}{(1+l^2/a_BR_c)^2}\frac{2-\sin(4N)}{(4N)^2} \tag{58}$$

and for large wave vectors $kR_c \gg 1$,

$$\Gamma_{SW} = -\frac{\arctan(2\omega_c\tau g_{\text{eff}})}{\pi\omega_c\tau}\frac{e^2}{a_B} \times \left[\left(\frac{a_B}{R_c}\right)^2\ln\frac{R_c}{a_B} + \frac{\operatorname{arccosh}(2kR_c)}{2(4N)^2}\right]. \tag{59}$$

We note that the decay rate Γ_{SW} is of the same order as the corrections to the real part of the spin-wave energy E_{SW} due to the presence of disorder.

5. ZERO-BIAS ANOMALY

In this section, we consider the electron tunneling into a two-dimensional electron liquid with disorder in a weak magnetic field. We investigate suppression of the tunneling conductivity near zero bias, the so-called zero-bias anomaly. The properties of the electron tunneling into an electron system are usually described by a dependence of the tunneling conductivity $G(V)$ on the bias V. Recently, the effective action approach to the zero-bias problem was developed by Levitov and Shytov [24]. The effective action describes spreading of the tunneling electron within the electron system in imaginary time ζ.

Following [24], the action of a spreading charge for zero bias $V = 0$ is determined by

$$S_0(\zeta) = 4\int_0^{+\infty}\frac{d\omega}{2\pi} \times \int_0^{+\infty}\frac{qdq}{2\pi}\frac{\sin^2(\omega\zeta)}{\omega+Dq^2}\frac{U_{\text{eff}}(q)}{\omega+Dq^2+\sigma q^2U_{\text{eff}}(q)}, \tag{60}$$

where σ and D are the conductivity and the diffusive constant of the electron system, respectively. They are related by the Einstein formula $\sigma = e^2\rho D$.

Using asymptotic expression (42) for the static dielectric function $\varepsilon(q)$, we evaluate action (60) in the large time limit $\zeta \gg 1$ as

$$S_0(\zeta) = \frac{e^2}{8\pi^2\sigma\eta}\ln\frac{2\zeta}{\tau_0}\ln\left(\frac{2\zeta}{\tau_0}\beta^{4\eta}\right), \tag{61}$$

where we introduce two dimensionless parameters

$$\beta = \frac{a_B}{\sqrt{2}l_{el}}, \quad \eta = \left(1-\frac{\pi}{6\omega_c\tau}\right)\left(\frac{R_c}{\sqrt{2}l_{el}}\right)^2 \tag{62}$$

with the bare elastic mean free path $l_{el} = R_c\omega_c\tau_0$. In accordance with the inequality $a_B \ll R_c \ll l_{el}$, the parameters β and η are small, $\beta \ll 1$ and $\eta \ll 1$.

Taking the work done by the voltage source into account, we obtain the total action of the spreading charge

$$S(\zeta) = S_0(\zeta) - 2eV\zeta.$$

We must then find the optimum time ζ_* determined by the minimum of the action $S(\zeta)$. The optimum time ζ_* plays the role of the charge accommodation time in the problem. It can be written as

$$\zeta_* = \tau_0\frac{V_0}{2V}\ln\frac{V_0}{\beta^{2\eta}V}, \quad eV_0 = \left(1-\frac{\pi}{6\omega_c\tau}\right)^{-1}\frac{1}{\pi mR_c^2}. \tag{63}$$

The theory must be self-consistent in the hydrodynamical limit, i.e., for $\zeta_* \geq \tau_0$. Therefore, the theory is applicable for the bias $V \leq V_0$.

Assuming the contribution from the barrier to be a constant at a small bias, we can write the tunneling conductivity as

$$G(V) = G_0\exp[-S_0(\zeta_*) + 2eV\zeta_*]. \tag{64}$$

After the evaluation, we obtain the dependence of the tunneling conductivity for a small bias,

$$G(V) = G_0\left(\frac{V}{V_0}\right)^{\alpha(V)}, \quad \alpha(V) = \frac{e^2}{8\pi^2\sigma\eta}\ln\frac{V_0}{V\beta^{4\eta}}. \tag{65}$$

Equation (65) shows that the screening of the electron–electron interaction results in increasing the suppression of the tunneling conductivity. We note that the above result is valid for the bias in the range $V \leq V_0$.

Expression (65) for the tunneling conductivity contains the energy scale eV_0 that coincides with the "hydrodynamic" term E_h in Eq. (52) except for the logarithm. A hydrodynamical model for the low-energy excitations of a clean ($\tau^{-1} = 0$) electron liquid in a weak magnetic field was considered by Aleiner, Baranger, and Glazman [6]. They showed that the tunneling density of states exhibits the gap $2E_h$ related to the Fermi energy. Equation (52) describes the same gap for a weak disorder $\omega_c\tau \gg 1$. Apparently, the disorder is responsible for the fact that the gap is about $0.05\omega_c$ in a wide range of the applied magnetic field [5].

As the magnetic field increases, the factor α increases and becomes of the order of unity. The zero-bias anomaly in the tunneling conductivity crosses over from weak to strong. Expression (65) shows that the factor α depends on the bias V and the magnetic field. This results in the shift of the crossover point V_c along the bias V as the applied magnetic field changes,

$$V_c = V_0 \exp\left(-\frac{4\pi\mu}{\omega_c^2\tau_0}\right), \tag{66}$$

where μ is the chemical potential. The crossover was observed by Ashoori *et al.* [5] in the tunneling current from a normal metal into two-dimensional electrons in the presence of a magnetic field. In the experiment, the ohmic conductance was measured as a function of the temperature T. For low temperatures, the conductance corresponds to the zero temperature conductance taken at $V = T/e$. The two-dimensional electrons were relatively clean, with the elastic collision time $\tau_0 \approx 4 \times 10^{-12}$ s. The chemical potential calculated from the electron density was $\mu = 10$ mV. Using Eq. (66), the dependence of the crossover temperature on the magnetic field can be written as

$$T_c = 2.9 \exp\left(-\left[\frac{3.2}{H}\right]^2\right), \tag{67}$$

where the temperature is measured in kelvin and the magnetic field in tesla. Equation (67) demonstrates a good agreement with the results reported in [5].

6. CONCLUSIONS

We have considered the system of a two-dimensional electron gas in the presence of disorder and Coulomb interaction in a weak perpendicular magnetic field. The effective low-energy theory describing electrons at the partially filled Nth Landau level was derived in the case of a weak magnetic field ($Nr_s \gg 1$) and a weak interaction ($r_s \ll 1$). The modified electron–electron interaction for electrons on a partially filled Landau level involves the screening from the other electrons on the occupied Landau levels. We also presented the exchange corrections to the thermodynamic and chemical potentials in the presence of disorder.

The theory proposed here allows us to account for the effects of disorder in the problems connected with the behavior of a two-dimensional electron gas in a weak magnetic field. It can be investigated how disorder affects the formation of stripes, bubble phase, tunneling density of states, spin excitations, tunneling conductivity, etc.

We discussed the effect of disorder on the exchange enhancement of the g factor and the simplest spin excitations on the partially filled Landau level. We obtained an additional dependence of the effective g factor as a function of the magnetic field, the suppression of the effective mass and the decay rate of the spin-wave excitations.

We also investigated the electron tunneling into a two-dimensional electron liquid with a weak disorder in a weak magnetic field. We obtained the enhancement of the gap in the tunneling density of states and a nonlinear dependence of the tunneling conductivity on the applied bias.

ACKNOWLEDGMENTS

The author is grateful to M.A. Baranov, A.S. Ioselevich, and M.V. Feigel'man for useful discussions. The author thanks M.A. Baranov for some critical remarks during preparation of the manuscript.

APPENDIX A

In this appendix, we calculate the term S_μ in action (29). This term appears after performing integration over the longitudinal fluctuations and is given by

$$S_\mu = \frac{1}{2}\langle (S_2[\overline{\Psi}, \Psi, \delta P, \lambda])^2 \rangle_{\delta P}, \tag{A.1}$$

where $\langle \ldots \rangle_{\delta P}$ denotes the average with the propagator of the δP field in Eq. (28). We then obtain

$$S_\mu = \delta S_1 + \delta S_2 + \delta S_3, \tag{A.2}$$

where

$$\delta S_1 = T^2 \iint d\mathbf{r}_1 d\mathbf{r}_2 (\Psi^\dagger \hat{\lambda} \tilde{G}_0 \hat{\lambda} \Psi - 2\mathrm{tr}\hat{\lambda}\tilde{G}_0\hat{\lambda}G_0), \tag{A.3}$$

$$\begin{aligned}\delta S_2 = -T^2 \int d\mathbf{r}_1 \ldots \int d\mathbf{r}_4 \sum_{k,n,m}^{\alpha,\beta} \lambda_m^\beta(\mathbf{r}_1) D_{nm}^{\alpha\alpha}(k; \mathbf{r}_3, \mathbf{r}_4) \\ \times \pi_0^n(m; \mathbf{r}_1, \mathbf{r}_2) L_{n,n+m}^{\alpha\beta,\alpha\alpha}(\mathbf{r}_2, \mathbf{r}_4),\end{aligned} \tag{A.4}$$

$$\begin{aligned}\delta S_3 = \frac{T^2}{2} \int d\mathbf{r}_1 \ldots \int d\mathbf{r}_4 \sum_{k,l,n,m}^{\alpha,\beta,\gamma,\delta} D_{nm}^{\alpha\beta}(k; \mathbf{r}_1, \mathbf{r}_2) \\ \times D_{nm}^{\gamma\delta}(l; \mathbf{r}_3, \mathbf{r}_4) L_{n,n+m}^{\alpha\beta,\gamma\delta}(\mathbf{r}_2, \mathbf{r}_4).\end{aligned} \tag{A.5}$$

Here,

$$\begin{aligned}D_{nm}^{\alpha\beta}(k; \mathbf{r}_1, \mathbf{r}_2) = (\overline{\Psi}_k^\alpha(\mathbf{r}_1)\Psi_{n+m}^\beta(\mathbf{r}_2) \\ -2\delta_{k,n+m}\delta^{\alpha,\beta}G_0^{n+m}(\mathbf{r}_2, \mathbf{r}_1))\lambda_{n-k}^\alpha(\mathbf{r}_1)\tilde{G}_0^n(\mathbf{r}_1, \mathbf{r}_2) \\ + (\overline{\Psi}_n^\beta(\mathbf{r}_1)\Psi_k^\alpha(\mathbf{r}_2) - 2\delta_{k,n}\delta^{\alpha,\beta}G_0^n(\mathbf{r}_2, \mathbf{r}_1)) \\ \times \lambda_{k-n-m}^\beta(\mathbf{r}_1)\tilde{G}_0^{n+m}(\mathbf{r}_1, \mathbf{r}_2),\end{aligned} \tag{A.6}$$

and $L_{m_1,m_2}^{\alpha\beta,\gamma\delta}$ is the propagator of longitudinal fluctuations in Eq. (28).

Integrating over the plasmon field, we obtain from the S_μ term

$$S_\mu \longrightarrow \frac{\Delta\Omega_1 + \Delta\Omega_2 + \Delta\Omega_3}{T} + (\delta\mu_1 + \delta\mu_2 + \delta\mu_3)\int d\mathbf{r}\Psi^\dagger(\mathbf{r})\Psi(\mathbf{r}), \tag{A.7}$$

where second-order corrections to the thermodynamic potential are given by

$$\frac{\Delta\Omega_1}{L_xL_y} = -T^2\sum_{n,\,m}\int d\mathbf{r}G_0^n(0,\mathbf{r})\tilde{G}_0^m(\mathbf{r},0)U_{\text{eff}}(m-n,\mathbf{r}),$$

$$\begin{aligned}\frac{\Delta\Omega_2}{L_xL_y} &= 2T^2\sum_{n,\,m}\int d\mathbf{r}d\mathbf{r}_1d\mathbf{r}_2\\ &\times(G_0^n(0,\mathbf{r})\tilde{G}_0^{n+m}(\mathbf{r},0) + \tilde{G}_0^n(\mathbf{r},0)G_0^{n+m}(0,\mathbf{r}))\\ &\times U_{\text{eff}}(m,\mathbf{r}_1)\pi_0^n(m,\mathbf{r}_1-\mathbf{r}_2)L_{n,\,n+m}^{\alpha\alpha,\,\alpha\alpha}(\mathbf{r}_2-\mathbf{r}),\\ \frac{\Delta\Omega_3}{L_xL_y} &= 2T^2\sum_{n,\,m}\int d\mathbf{r}d\mathbf{r}_1d\mathbf{r}_2\\ &\times(G_0^n(0,\mathbf{r})\tilde{G}_0^{n+m}(\mathbf{r},0) + G_0^{n+m}(0,\mathbf{r})\tilde{G}_0^n(\mathbf{r},0))\\ &\times(G_0^n(0,\mathbf{r}_1)\tilde{G}_0^{n+m}(\mathbf{r}_1,0) + G_0^{n+m}(0,\mathbf{r}_1)\tilde{G}_0^n(\mathbf{r}_1,0))\\ &\times U_{\text{eff}}(m,\mathbf{r}_2)L_{n,\,n+m}^{\alpha\alpha,\,\alpha\alpha}(\mathbf{r}-\mathbf{r}_1+\mathbf{r}_2).\end{aligned} \tag{A.8}$$

The above corrections are negligible in the parameter N^{-1} compared to the correction determined by Eq. (35). The corrections to the chemical potential are given by

$$\frac{\delta\mu_1}{2\pi l^2} = T\sum_m d\mathbf{r}P_N(0,\mathbf{r})\tilde{G}_0^m(\mathbf{r},0)U_{\text{eff}}(m,\mathbf{r}),$$

$$\begin{aligned}\frac{\delta\mu_2}{2\pi l^2} &= -4T\sum_m\int d\mathbf{r}d\mathbf{r}_1d\mathbf{r}_2P_N(0,\mathbf{r})\tilde{G}_0^m(\mathbf{r},0)\\ &\times U_{\text{eff}}(m,\mathbf{r}_1)\pi_0^0(m,\mathbf{r}_1-\mathbf{r}_2)L_{0,\,m}^{\alpha\alpha,\,\alpha\alpha}(\mathbf{r}_2-\mathbf{r}),\\ \frac{\delta\mu_3}{2\pi l^2} &= -8T^2\sum_m\int d\mathbf{r}d\mathbf{r}_1d\mathbf{r}_2P_N(0,\mathbf{r})\tilde{G}_0^m(\mathbf{r},0)\\ &\times U_{\text{eff}}(m,\mathbf{r}_1)L_{0,\,m}^{\alpha\alpha,\,\alpha\alpha}(\mathbf{r}-\mathbf{r}_1+\mathbf{r}_2)\\ &\times(G_0^0(0,\mathbf{r}_1)\tilde{G}_0^m(\mathbf{r}_1,0) + G_0^m(0,\mathbf{r}_1)\tilde{G}_0^0(\mathbf{r}_1,0)).\end{aligned} \tag{A.9}$$

The second and third corrections are negligible in the parameter N^{-1} compared to the first term. The shift of the chemical potential $\delta\mu$ is therefore mainly determined by the first correction $\delta\mu_1$.

APPENDIX B

In this appendix, we evaluate the polarization operator $\Pi(\omega_n, q)$. The condition $\omega_c\tau \gg 1$ is assumed to hold. Then,

$$\Pi(\omega_n, q) = T\sum_m \pi^m(\omega_n, q). \tag{B.1}$$

The calculation of the polarization operator $\Pi(\omega_n, q)$ is analogous to that given in [7]. The wave vectors $q \ll R_c/l^2$ are considered.

Using Eq. (31), we immediately obtain

$$\begin{aligned}\Pi(\zeta_n, Q) &= \frac{m}{\pi}\\ &\times\left[1 - \sum_{j=-\infty}^{+\infty}\frac{\mathcal{J}_j^2(Q)}{j^2+\zeta_n^2}\left(\zeta_n^2 + \frac{\pi}{3\tau\omega_c}\mathcal{L}_j(\zeta_n)\right)\right],\end{aligned} \tag{B.2}$$

where

$$\begin{aligned}\mathcal{L}_j(\zeta_n) &= \Theta(|j|-2)\\ &\times\sum_{n=1}^{|j|-1}\left[\frac{|j|}{n} - \frac{1}{2}\frac{j^2-\zeta_n^2}{j^2+\zeta_n^2}\ln\frac{\zeta_n^2+n^2}{n^2} + \frac{2|j|\zeta_n}{(j^2+\zeta_n^2)^2}\arctan\frac{\zeta_n}{n}\right]\\ &+ \Theta(|j|-1)\left[\frac{1}{2} + \frac{\zeta_n|j|\arctan\dfrac{\zeta_n}{|j|}}{j^2+\zeta_n^2} + \frac{1}{4}\frac{j^2-\zeta_n^2}{j^2+\zeta_n^2}\right.\\ &\left.\times\ln\left(\frac{\zeta_n^2+j^2}{j^2}(1+2\tau\omega_c\zeta_n)^2\right)\right].\end{aligned} \tag{B.3}$$

Two parameters $\zeta_n = \omega/\omega_c$ (with $\omega = 2\pi Tn$) and $Q = qR_c$ are introduced here. The transformation of series (B.2) into the integral form yields the asymptotic form of the polarization operator in the different regimes. In the static limit $\zeta_n \ll 1$,

$$\begin{aligned}\Pi(\zeta_n, Q) &= \frac{m}{\pi}\left(1 - \mathcal{J}_0^2(Q) - \frac{\pi}{3\omega_c\tau}\eta(Q)\right.\\ &\left.+ \frac{\ln(1+2\omega_c\tau\zeta_n)}{2\pi\omega_c\tau}\xi(Q) + O(\zeta^2)\right),\end{aligned} \tag{B.4}$$

where the function $\eta(x)$ is defined as

$$\eta(x) = \frac{1}{2\pi}\int_0^\pi\frac{dy}{\pi}\mathcal{J}_0\left(2x\sin\frac{y}{2}\right)\left[\ln^2\left(2\sin\frac{y}{2}\right) - \frac{\pi^2}{12}\right], \tag{B.5}$$

and its asymptotic form is given in Eq. (41). The function $\xi(x)$ is defined as

$$\xi(x) = \int_0^\pi \frac{dy}{\pi} \mathcal{J}_0\left(2x\sin\frac{y}{2}\right)\left[(y-\pi)^2 - \frac{\pi^2}{3}\right], \tag{B.6}$$

and its asymptotic form is given by

$$\xi(x) = \begin{cases} x^2, & x \ll 1 \\ \dfrac{\pi}{3x}(2 - \sin 2x), & x \gg 1. \end{cases} \tag{B.7}$$

In the hydrodynamic limit $qR_c \ll 1$, we obtain

$$\begin{aligned} \Pi(\zeta_n, Q) = {} & \frac{m}{2\pi}\frac{Q^2}{1+\zeta_n^2}\left[1 - \frac{\pi}{6\tau\omega_c} + \frac{1}{2\pi\tau\omega_c}\frac{1-\zeta_n^2}{1+\zeta_n^2}\right. \\ & \left.\times \ln((1+\zeta_n^2)(1+2\tau\omega_c\zeta_n)^2) - \frac{\pi}{3\tau\omega_c}\frac{\zeta_n \arctan\zeta_n}{1+\zeta_n^2}\right]. \end{aligned} \tag{B.8}$$

APPENDIX C

In this appendix, we evaluate the corrections to the thermodynamic and chemical potentials.

C.1. Correction to the Thermodynamic Potential

Using Eq. (35), we can split the correction to the thermodynamic potential into the exchange and correlation ones as

$$\Delta\Omega = \Delta\Omega_{\text{ex}} + \Delta\Omega_c, \tag{C.1}$$

$$\frac{\Delta\Omega_{\text{ex}}}{L_x L_y} = \frac{T}{2}\sum_n \int \frac{d\mathbf{q}}{(2\pi)^2} U_0(q)\Pi(n, q), \tag{C.2}$$

$$\frac{\Delta\Omega_c}{L_x L_y} = -\frac{T}{2}\sum_n \int \frac{d\mathbf{q}}{(2\pi)^2}\int_0^1 d\alpha \frac{\alpha U_0^2(q)\Pi^2(n, q)}{1 + \alpha U_0(q)\Pi(n, q)}. \tag{C.3}$$

The exchange correction gives the leading contribution [7] and can be written as

$$\begin{aligned} \frac{\Delta\Omega_{\text{ex}}}{L_x L_y} = {} & -\frac{e^2}{2\pi l^3}\sum_{m \neq N}\int_0^\infty dx e^{-x^2/2} L_N^1\left(\frac{x^2}{2}\right) \\ & \times L_m\left(\frac{x^2}{2}\right)\left(\Theta(N-m) + \frac{1}{\pi\omega_c\tau}\frac{1}{m-N}\right), \end{aligned} \tag{C.4}$$

where L_n^m stands for the Laguerre polynomials. The above equation goes to Eq. (48) in the case where $N \gg 1$.

C.2. Correction to the Chemical Potential

Using Eq. (37), we can split the correction to the chemical potential into the exchange and correlation ones,

$$\delta\mu = \delta\mu_{\text{ex}} + \delta\mu_c, \tag{C.5}$$

$$\delta\mu_{ex} = 2\pi l^2 T\sum_n \int d\mathbf{r}\, U_0(r) P_N(0, \mathbf{r})\tilde{G}_0^n(\mathbf{r}, 0), \tag{C.6}$$

$$\begin{aligned} \delta\mu_c = {} & -2\pi l^2 T\sum_n \int \frac{d^2\mathbf{q}}{(2\pi)^2} P_N(q)\tilde{G}_0^n(q) \\ & \times \frac{U_0^2(q)\Pi(n, q)}{1 + U_0(q)\Pi(n, q)}. \end{aligned} \tag{C.7}$$

The exchange correction gives the leading contribution [7] and can be written as

$$\begin{aligned} \delta\mu_{\text{ex}} = {} & -\frac{e^2}{l}\sum_{m \neq N}\int_0^\infty dx e^{-x^2/2} L_N\left(\frac{x^2}{2}\right) L_m\left(\frac{x^2}{2}\right) \\ & \times\left(\Theta(N-m) + \frac{1}{2\pi\omega_c\tau}\frac{1}{m-N}\right). \end{aligned} \tag{C.8}$$

For $N \gg 1$, the above equation leads to

$$\begin{aligned} \delta\mu_{\text{ex}} = {} & \frac{2e^2}{\pi l^2}(2N)^{1/2}\left[1 - \frac{\ln N}{8N} + \frac{1}{4\pi\omega_c\tau}\frac{1}{2N}\right. \\ & \left.\times\left(\int_1^\infty \frac{dt}{t}\ln(1 - e^{-t}) + \int_0^1 \frac{dt}{t}\ln\frac{1 - e^{-t}}{t} - \frac{\pi^2}{3}\right)\right]. \end{aligned} \tag{C.9}$$

REFERENCES

1. *The Quantum Hall Effect*, Ed. by R. E. Prange and S. M. Girvin (Springer-Verlag, Berlin, 1987).
2. A. H. MacDonald and S. M. Girvin, Phys. Rev. B **33**, 4009 (1986); R. Morf and N. d'Ambrumenil, E-print archives cond-mat/9409008; L. Belkhir and J. Jain, E-print archives cond-mat/9409020.
3. A. P. Smith, A. H. MacDonald, and G. Gumbs, Phys. Rev. B **45**, 8829 (1992).
4. J. P. Eisenstein, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **69**, 3804 (1992).
5. R. C. Ashoori, J. A. Lebens, N. P. Bigelow, *et al.*, Phys. Rev. Lett. **64**, 681 (1990); Phys. Rev. B **48**, 4616 (1993).
6. I. L. Aleiner, H. U. Baranger, and L. I. Glazman, Phys. Rev. Lett. **74**, 3435 (1995).
7. I. L. Aleiner and L. I. Glazman, E-print archives cond-mat/9505026.
8. A. A. Koulakov, M. M. Fogler, and B. I. Shklovskii, Phys. Rev. Lett. **76**, 499 (1996); Phys. Rev. B **54**, 1853 (1996); R. Moessner and J. T. Chalker, Phys. Rev. B **54**, 5006 (1996).

9. M. P. Lilly, K. B. Cooper, J. P. Eisenstein, *et al.*, Phys. Rev. Lett. **82**, 394 (1999).
10. M. M. Fogler, E-print archives cond-mat/0111001.
11. K. B. Efetov, A. I. Larkin, and D. E. Khmel'nitzkii, Zh. Éksp. Teor. Fiz. **79**, 1120 (1980) [Sov. Phys. JETP **52**, 568 (1980)].
12. A. M. Finkel'stein, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 436 (1983); Zh. Éksp. Teor. Fiz. **84**, 168 (1983); Zh. Éksp. Teor. Fiz. **86**, 367 (1984).
13. A. M. M. Pruisken, Nucl. Phys. B **235**, 277 (1984).
14. A. M. M. Pruisken, M. A. Baranov, and B. Škorić, Phys. Rev. B **60**, 16807 (1999).
15. T. Ando and Y. Uemura, J. Phys. Soc. Jpn. **36**, 959 (1974); **36**, 1521 (1974); T. Ando, J. Phys. Soc. Jpn. **37**, 1233 (1974).
16. M. A. Baranov and A. M. M. Pruisken, private communication.
17. T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
18. I. V. Kukushkin, S. V. Meshkov, and V. B. Timofeev, Usp. Fiz. Nauk **155**, 219 (1988) [Sov. Phys. Usp. **31**, 511 (1988)].
19. X. G. Wu and S. L. Sondhi, E-print archives cond-mat/9502111.
20. T. Ando and Y. Uemura, J. Phys. Soc. Jpn. **37**, 1044 (1974).
21. C. Kallin and B. I. Halperin, Phys. Rev. B **30**, 5655 (1984).
22. K. Suzuki and Y. Kawamoto, J. Phys. Soc. Jpn. **35**, 1456 (1973).
23. Yu. A. Bychkov, S. V. Iordanskiĭ, and G. M. Eliashberg, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 152 (1981) [JETP Lett. **33**, 143 (1981)].
24. L. S. Levitov and A. V. Shytov, E-print archives cond-mat/9607136 and references therein.

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 145–153.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 164–174.
Original Russian Text Copyright © 2002 by Rostunov, Shchur.

NONLINEAR PHYSICS

Early Stages of Generation of Two-Dimensional Structures by the Hastings–Levitov Method of Conformal Mapping Dynamics

T. A. Rostunov[a], * and L. N. Shchur[a, b, c]

[a]*Landau Institute for Theoretical Physics, Russian Academy of Sciences,*
Chernogolovka, Moscow oblast, 142432 Russia
[b]*Laboratoire de Physique des Matériaux, Université Henri Poincaré,*
Nancy B.P. 239, F-54506 Vandoevre les Nancy Cedex, France
[c]*Istituto Nazionale di Fisica Nucleare, Universitá Milano-Bicocca I-20126, Milano, Italy*
**e-mail: rostunov@itp.ac.ru*

Received September 28, 2001

Abstract—Two-dimensional structures obtained by the Hastings–Levitov conformal mapping were studied for a relatively small number of mappings n. The fractal dimension D of these structures is computed by the recent Davidovitch–Procaccia technique [6] as a function of n. For small $n < n_0$ (where n_0 is the number of particles at the first layer), D exponentially decreases, which should have supported the conclusion made in [6] about the possibility of determining the fractal dimension with an arbitrary accuracy using a relatively small number of mappings $n \approx n_0$. On the other hand, it turned out that D irregularly deviates from a certain quantity D_0 depending on the initial size of the bump $\sqrt{\lambda_0}$, which contradicts the main assertion of [6]. However, our analysis does not rule out the possibility of determining the fractal dimension of two-dimensional structures by the original Hastings–Levitov method. *© 2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

There are many objects in nature that grow due to diffusion of particles forming a cluster (for example, crystal growth, formation of frost work, formation of mineral inclusions in rocks, formation of river networks, and similar phenomena, such as dielectric breakdown).

These phenomena can sometimes be considered as dynamic critical processes. The most interesting characteristics of such processes are the fractal dimension D of the generated objects and multifractal properties of their surface of growth. These properties are actually the properties of an ensemble of object surfaces, and their study requires proper techniques of averaging over the ensemble [1].

An intensive study of growth processes started about 20 years ago when Witten and Sanders proposed the DLA (Diffusion Limited Aggregation) model in their 1981 work [2]. This model produces two-dimensional structures that are qualitatively similar to those mentioned above (see, e.g., the recent review [3]).

In the DLA model, the process starts with placing a seed particle at the origin of coordinates, and then the cluster grows at the expense of particles that diffuse from infinity (in the model, infinity is replaced by a circle with radius much greater than the possible size of the cluster). When a particle touches the cluster, it sticks to it, and this is how the cluster grows. As a particle sticks to the cluster or goes to infinity (at a distance much greater than the radius of the initial circle), the next particle is released. The resulting cluster appears to have a fractal structure.

In recent years, considerable progress has been made in studying DLA-like objects due to the novel techniques proposed by Hastings and Levitov [4] for modeling the growth of structures. In this model, clusters are generated using a sequence of mappings of a unit circle exterior to the exterior of the growing object. At every step, a conformal mapping is used that maps the exterior of the unit circle to the exterior of the unit circle with a bump in the form of a semiellipse. Thus, each mapping adds a new particle to the cluster. The parameter of the mapping, λ, which is responsible for the bump area, is chosen depending on the problem. For example, in the DLA model, λ is chosen so that the sizes of all bumps are identical after the sequential application of all transformations to the unit circle (see Section 2 for details).

In 1996, Hastings [5] applied the renormalization group approach to the approximate analytical calculation of the fractal dimension in the DLA model using the conformal mapping technique. He obtained the rational value $D = 17/10$, which is rather close to the value $D = 1.71$, which is usually obtained as a result of direct numerical simulation.

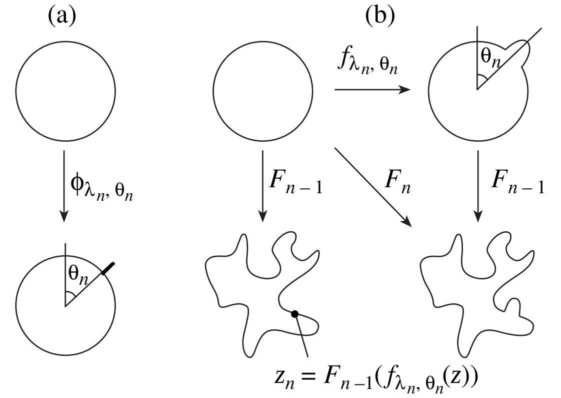


Fig. 1. Action of the mappings $\phi_{\lambda_n,\theta_n}$, f_{λ_n,θ_n}, and F_{n-1}, F_n.

In the Hastings–Levitov method, the fractal dimension is determined by the dependence of the first term in the Laurent series of the cluster generation function on the number of particles [4]. Later, Davidovitch and Procaccia proposed a modified method for the numerical computation of the cluster fractal dimension [6] (see below for details). They claimed that the fractal dimension can be determined with a high accuracy at early stages of the growth, whereas conventional methods require that huge clusters be analyzed.[1] The main purpose of this paper is to verify this very strong claim.

The paper is organized as follows. In Section 2, the Hastings–Levitov model [4] is described. In Section 3, the method for determining the fractal dimension proposed in [6] is outlined, and its implementation used in this paper is presented. Details of computations and the results are presented in Section 4. We also discuss self-averaging of the fractal dimension computed by the Davidovitch–Procaccia method at early stages of the growth and the variation of the dimension depending on the number of particles for a certain set of problem parameters. In Section 5, conclusions are made on the basis of the simulation results.

2. THE HASTINGS–LEVITOV MODEL

Let $u(z)$ be the probability density of finding a particle at the point z. It is determined by solving the Laplace equation

$$\Delta u = 0 \tag{1}$$

with the boundary conditions

$$u = 0 \tag{2}$$

on the surface of the cluster and

$$u = \frac{1}{2\pi}\ln(|z|) \tag{3}$$

on the circle of a large radius ($|z| \longrightarrow \infty$).

Equation (1) describes diffusion without sources. Condition (2) states that the particle that reaches the surface sticks to it and is not free anymore (in the DLA model, this ensures the cluster growth), and condition (3) tells that particles diffuse equiprobably from every direction at infinity.

The probability of the DLA cluster growth at a certain point of its boundary is determined as the probability for a particle to touch the element dl of the boundary containing this point:

$$dP(z) \propto |\nabla u| dl$$

(this probability is proportional to the flux at the boundary).

In the general case, the probability of the cluster growth can be determined as

$$dP \propto |\nabla u|^{\alpha} dl \tag{4}$$

with an arbitrary value of α. The value $\alpha = 1$ corresponds to the DLA model, which is a particular case of the dielectric breakdown model (DBM) [8]. By definition, the DBM is a model in which the probability of growth (i.e., the probability of the dielectric breakdown) is proportional to a certain power α of the field (of the potential gradient) in the vicinity of the surface.

Problem (1)–(3) can be modeled using the iteration of conformal mappings of the exterior of the unit circle to the exterior of the growing cluster [4]. At every step, the composition of two mappings is used. The function

$$\phi_{\lambda,\theta}(z) = e^{i\theta}\phi_{\lambda}(e^{-i\theta}z) \tag{5}$$

maps the exterior of the unit circle to the exterior of the unit circle with a δ-shaped bump of the size

$$2\sqrt{\lambda} + O(\lambda^{3/2}), \quad \lambda \ll 1,$$

at the point $z = e^{i\theta}$, which is schematically shown in Fig. 1a. In the original paper [4], the function $\phi_{\lambda}(z)$ was taken in the form

$$\begin{aligned}\phi_{\lambda}(z) &= \frac{1+\lambda}{2z}(z+1) \\ &\times\left(z + 1 + \sqrt{z^2 + 1 - 2z\frac{1-\lambda}{1+\lambda}}\right) - 1.\end{aligned} \tag{6}$$

Then, the transformation

$$f_{\lambda,\theta}(z) = z^{1-a}\phi^{a}_{\lambda,\theta}(z)$$

maps the exterior of the circle to the exterior of the circle with a bump of a certain shape. The shape of the bump around the circle point $z = e^{i\theta}$ depends on the parameter a. At $a = 2/3$, the bump has the same size $(4/3)\sqrt{\lambda}$ in any direction, and it can be considered as an attached circular particle. At the nth iteration step, the resulting function $F_{n-1}(z)$, which is the superposition of $n - 1$ mappings f performed at the preceding steps,

---

[1] For example, two-dimensional clusters consisting of 10^8 particles were analyzed in [7].

transforms the circle to a cluster consisting of $n - 1$ particles, and the circle with an bump is transformed to a cluster consisting of n particles (see Fig. 1b).

Thus, a cluster consisting of n particles is obtained from the unit circle by a sequence of n mappings

$$F_n(z) = F_{n-1}(f_{\lambda_n, \theta_n}(z)),$$

where the initial function $F_0(z) = z$. Here, λ_n is defined in such a way that the area of the nth bump is proportional to $\lambda_0|\nabla u|^{\alpha - 1}$ upon the transformation F_{n-1}. The probability of growth at the element of the cluster surface is proportional to $|\nabla u|$, where λ_0 is the parameter determining the initial size of the particles. Since the linear dimensions at the point z change proportionally to $(F'(z))^{-1}$ under conformal mappings (the prime denotes the differentiation), we have

$$\lambda_n = \frac{\lambda_0}{\left(F'_{n-1}(z)\Big|_{z = F_{n-1}(e^{i\theta_n})}\right)^{1+\alpha}}. \qquad (7)$$

The proportional dependence of the probability of growth (4) is taken into account through the corresponding change of the particle size (7): whereas the area of an object consisting of several bumps is initially proportional to $\lambda_n dl_0$, where dl_0 is the length element of the initial nucleus (i.e., the circle), it becomes proportional to

$$\lambda_n (F'_{n-1})^2 dl_0$$

upon the transformation F_{n-1}. On the other hand, the area is proportional to the length element of the resulting cluster, dl, and to the probability of growth, which, in turn, is proportional to a power of the field gradient at the surface. Therefore, the object area must be

$$\lambda_0 |\nabla u|^{\alpha} dl \propto \lambda_0 |F'_{n-1}|^{-\alpha + 1} dl_0.$$

Equating both expressions for the area, we obtain (7).

It was shown in [4] that θ_n is uniformly distributed in the interval $[0; 2\pi]$. Indeed, using the fact that the factor $|\nabla u|^{\alpha - 1}$ is accounted for by the size of the new particle, we have the following expression for the probability of growth:

$$dP \propto |\nabla u| dl \propto |\nabla u| \frac{dl_0}{|\nabla u|} \propto dl_0 \propto d\theta.$$

The characteristic linear size of the cluster is determined as the coefficient $F_n^{(1)}$ of z in the expansion of $F_n(z)$. Since u satisfies the electrostatic equations with the zero potential on the cluster surface, we can introduce an effective circle that has the same electrostatic properties at infinity as the cluster. Initially, the asymptotic behavior of the potential is

$$u \propto \ln|z|.$$

The function $f_{\lambda, \theta}$ is expanded into the Laurent series

$$f_{\lambda, \theta}(z) = (1 + \lambda)^a z + \sum_{k \le 0} a_k z^k$$

and has the asymptotic behavior

$$f_{\infty}(z) = (1 + \lambda)^a z.$$

This expansion contains no powers greater than unity since $f_{\lambda, \theta}$ is chosen so as not to affect remote areas by anything except for scale transformations. The asymptotic behavior of the resulting transformation is

$$F_{\infty}(z) = F_n^{(1)} z,$$

where

$$F_n^{(1)} = \prod_{k=1}^{n} (1 + \lambda_k)^a. \qquad (8)$$

Under this transformation, the potential at infinity has the form

$$u \propto \ln\left(\frac{|z|}{F_n^{(1)}}\right),$$

which corresponds to the potential of a circle of the radius $F_n^{(1)}$. For this reason, the linear size of the cluster consisting of n particles is taken to be $F_n^{(1)}$.

The above reasoning suggests that the DLA and Hastings–Levitov models are partially equivalent at $\alpha = 1$. In addition, there are examples of clusters in [4] constructed for various α, which are similar (at $\alpha = 1$) to DLA clusters. However, no rigorous proof of the equivalence of those two models has yet been found.

3. THE DAVIDOVITCH–PROCACCIA METHOD FOR DETERMINING THE FRACTAL DIMENSION

The fractal dimension is one of the fundamental characteristics of fractal objects. The Hastings–Levitov model provides a unique possibility of investigating various structures and determining their fractal dimension, in particular, for the DLA and DBM models, in a unified way.

The fractal dimension can be found by analyzing the asymptotic behavior of the dependence of the cluster linear size $F_n^{(1)}(\lambda_0)$ on the number of particles n at the fixed parameters λ_0, α, and a (see [5]). For large n, the linear size is fitted by the power function $n^{1/D}$, whence the fractal dimension D is found.

The value of the fractal dimension for DLA structures defies numerical computation with a relative error less than ~0.01. This is due to the very slow convergence of D with the growth of the number of particles when standard methods for the computation of the fractal dimension are used (see, for example, papers in [9]).

In [6], the authors claim that they succeeded in finding a procedure for the computation of D that converges rather rapidly and provides much more accurate results. More precisely, they claim that the dependence $F_n^{(1)}(\lambda_0)$ can be represented in the form of a universal function $F_*^{(1)}(x)$ even for small $x \ll 1$, where

$$x = \sqrt{\lambda_0} n^{1/D}, \tag{9}$$

and $F_n^{(1)}(\lambda_0)$ converges to $F_*^{(1)}(x)$ already at $n \geq n_0$. Here,

$$n_0 = \frac{3}{2}\frac{\pi}{\sqrt{\lambda_0}} \tag{10}$$

gives (for a given λ_0) the number of particles that can cover the nucleus by a single layer. Then, determining the values of n corresponding to the same value of the linear size of the cluster $F^{(1)}$ (i.e., to the same x) for various λ_0,

$$F_n^{(1)}(\lambda_0) = F_{n'}^{(1)}(s\lambda_0),$$

we obtain the fractal dimension

$$D(n; \lambda_0) = \frac{2(\ln n - \ln n')}{\ln s}, \tag{11}$$

where n is the number of the iteration step (the number of particles in the cluster) corresponding to the parameter of the particle size λ_0, and n' is the number of the iteration step corresponding to another parameter of the particle size $s\lambda_0$ at which the same value of the linear size of the cluster is reached. Since n and n' can be rather small, if this method is valid, statistically significant number of realizations can easily be accumulated and then highly accurate results can be obtained by averaging $D(n; \lambda_0)$ over a large number of realizations with a fixed λ_0. The authors of [6] claim that as λ_0 decreases, one can expect a rapid convergence of the value of the fractal dimension even at a small number of mappings $n \approx n_0$.

In this paper, we analyze this scheme for the computation of the fractal dimension.

4. SIMULATION RESULTS

4.1. Simulation Process

The growth of clusters was simulated as follows.

1. The values of the area parameter λ_0, the shape parameter $a = 2/3$ (which corresponds to a circular particle), and the power of the potential gradient ($\alpha = 1$ for the DLA model and $\alpha = 2.5$ for the DBM model) were chosen.

2. The unit circle centered at the origin of the complex plane was used as the nucleus.

3. The function $F_0(z) = z$, which maps the nucleus circle to itself, was used as the initial mapping function. The initial value of the coefficient of the linear term in the Laurent expansion of $F_0(z)$ was taken as $F_0^{(1)} = 1$. The initial step number was $n = 1$.

4. At the nth step, a pseudorandom number θ_n in the interval $[0, 2\pi]$ was chosen, which determined the location of the new particle corresponding to the point $z_n = F_{n-1}(e^{i\theta_n})$ on the cluster surface (see Fig. 1b).

5. At this point, the derivative $F'_{n-1}(z)\big|_{z = z_n}$ was computed, which determined the linear dilatation of the point neighborhood relative to its neighborhood on the nucleus.

6. The value of the parameter λ_n, which characterizes the size of the nth particle, was determined by Eq. (7).

7. Using λ_n, the next mapping function $F_n(z) = F_{n-1}(f_{\lambda_n, \theta_n}(z))$ of the unit circle on the cluster consisting of n particles (see Fig. 1b) and the corresponding new value of the cluster size $F_n^{(1)} = F_{n-1}^{(1)}(1 + \lambda_n)^a$ were determined.

The simulation is reduced to the iteration of steps 4–7 with n increased by one at each step. Thus, we find the dependence of the linear term coefficient of $F_n^{(1)}(\lambda_0)$ on n. The process described above is referred to as a realization. Under identical initial parameters selected in point 1, realizations differ from each other in different sets of random numbers θ_n. Average values and their variances are computed over an ensemble of realizations.

4.2. The Behavior of D at Small $n < n_0$

The Davidovitch–Procaccia approach [6] is based on the claim that the fractal dimension $D(n; \lambda_0)$ rapidly decreases with the number of the iteration step n and converges to a certain D_0, which remains almost invariable for $n \gtrsim n_0$. In this paper, we first investigated the behavior of the fractal dimension $D(n; \lambda_0)$ at early stages ($n < n_0$). The simulation was carried out for $\alpha = 1$ (the DLA model) with $a = 2/3$ and $\lambda_0 = 10^{-5}$ for 9000 various realizations in the interval $n \in [0; n_0/3]$.

For every basic realization of the cluster growth, an additional realization with a doubled initial value $\lambda_0 = 2 \times 10^{-5}$ ($s = 2$ in (11)) was simulated. From this realization, the number of iteration step n' was determined at which the values of $F_n^{(1)}(\lambda_0)$ and $F_{n'}^{(1)}(2\lambda_0)$ were identical. The values of D as a function of n in the above interval were obtained by averaging over those realizations: for each pair of realizations, $D(n; \lambda_0)$ was calculated by Eq. (11); then, it was averaged over the realizations.

Figure 2 shows $\ln(D(n; \lambda_0) - D_0)$ as a function of the dimensionless ratio n/n_0 for three possible asymptotic values of $D(n; \lambda_0)$ at large n: $D_0 = 1.75$, 1.76, and 1.77. The closer D_0 to the asymptotic value, the closer is $\ln(D(n; \lambda_0) - D_0)$ to the linear function. It is seen that, for $D_0 = 1.75$ and 1.77, the curves deviate from the linear function up and down, respectively; therefore, the intermediate value $D_0 = 1.76$ can be taken for the value of the Davidovitch–Procaccia fractal dimension for $\lambda_0 = 10^{-5}$.

4.3. The Distribution of D

In order to average the values of D correctly, one must know their statistical distribution.

The distribution function for D was computed for $\alpha = 1$ and 2.5. In all the realizations, we used $\lambda_0 = 10^{-5}$. For every α, ten series of realizations were carried out; then, the results were averaged over these series, and the standard deviation was computed.

Using the data of each series, a histogram was constructed from 11 700 points obtained in 900 realizations (13 values of D were taken for each realization at different stages of the cluster growth for certain n in the interval $[n_0; 4n_0]$). The histogram step was $\delta D = 0.05$ for $\alpha = 1.0$ and $\delta D = 0.1$ for $\alpha = 2.5$. The results are shown in Figs. 3a and 3b. It is seen that the distributions are close to Gaussian ones.

We denote by $\bar{D}$ the mean value of all $D(n; \lambda_0)$ for $n > n_0$ at fixed values of α and λ_0.

For $\alpha = 1$, $\bar{D} = 1.76$; and for $\alpha = 2.5$, $\bar{D} = 1.55$. In both cases, the mean values coincide with the location of the distribution maximum within the accuracy of computations, which confirms that the distribution is symmetric. For $\alpha = 1$, the maximizer of the distribution is $D_{\max} = 1.761 \pm 0.001$, and the mean value is $\bar{D} = 1.762 \pm 0.004$. For $\alpha = 2.5$, we have $D_{\max} = 1.538 \pm 0.002$ and $\bar{D} = 1.538 \pm 0.008$. Note that, for $\alpha = 1$ (i.e., for the DLA cluster), the value obtained by the Davidovitch–Procaccia method, $\bar{D} = 1.76$, differs from the "classical" value $\bar{D} = 1.71$ (see Subsection 4.5 for a more detailed discussion).

4.4. Self-Averaging of D

In systems with a second order phase transition, the correlation length in the neighborhood of the critical point can become very large and can exceed the finite size of the system when approaching the critical point. In disordered systems, some quantities can lack the property of self-averaging. This kind of behavior can be expected, for example, for the correlation function in spin lattice models with quenched impurities [10].

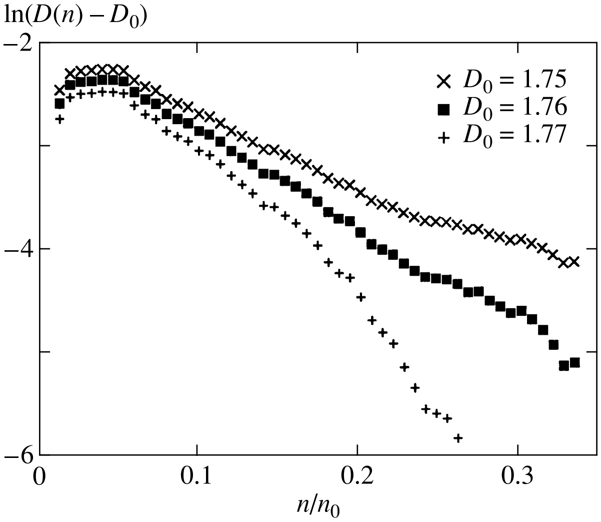


Fig. 2. Exponential relaxation of $D(n; \lambda_0)$ at small $n \leq 0.35n_0$ and $\lambda_0 = 10^{-5}$.

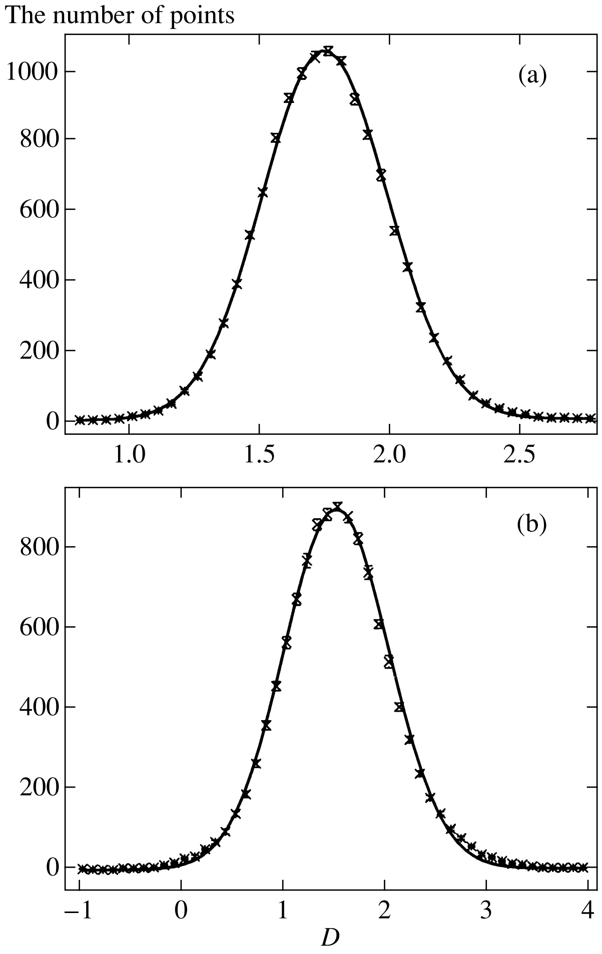


Fig. 3. The distribution of the fractal dimension D for $\alpha = 1$ (a) and 2.5 (b). Solid curves are Gaussian fits to the distributions.

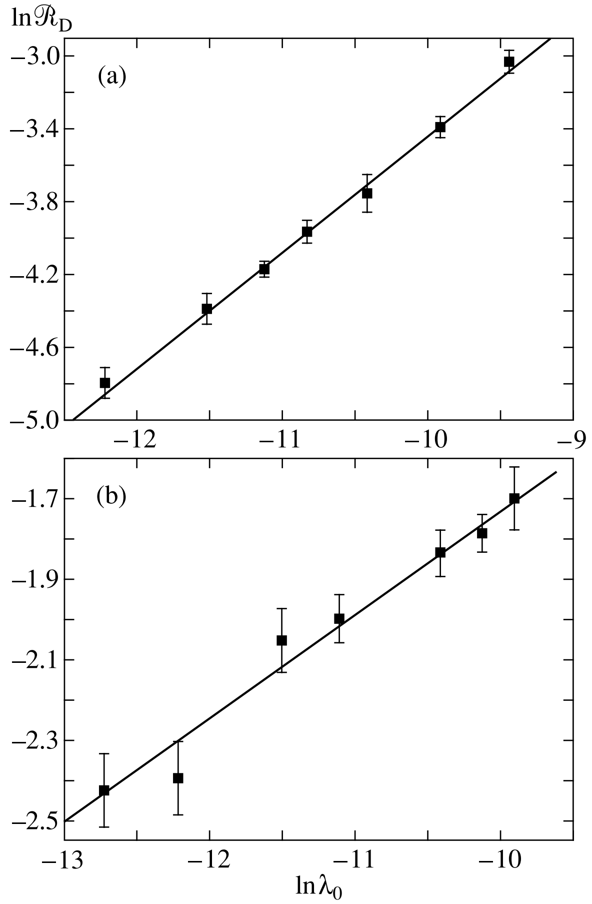


Fig. 4. The plots of $\ln\mathcal{R}_D(\ln\lambda_0)$ for $\alpha = 1$ (a) and 2.5 (b).

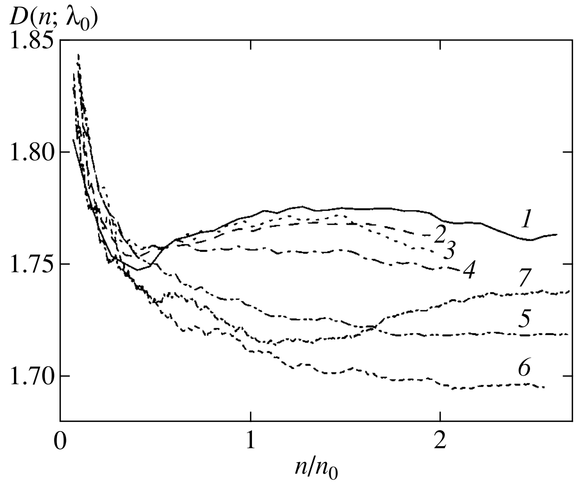


Fig. 5. The plots of $D(n;\lambda_0)$ for $n \lesssim 3n_0$ for models with various values of λ_0: 4×10^{-5} (*1*), 10^{-5} (*2*), 2.5×10^{-6} (*3*), 10^{-6} (*4*), 10^{-7} (*5*), 10^{-8} (*6*), and 10^{-9} (*7*).

Self-averaging of χ is determined by the behavior of its relative fluctuation

$$\mathcal{R}_\chi = \frac{\langle\chi^2\rangle - \langle\chi\rangle^2}{\langle\chi\rangle^2}$$

depending on the characteristic linear size of the system, L.

In the thermodynamic limit $L \longrightarrow \infty$, this quantity usually decreases inversely proportionally to the system volume:

$$\mathcal{R}_\chi \propto 1/L^d,$$

where d is the dimension of the system. If $\mathcal{R}_\chi$ decreases slower, i.e.,

$$\mathcal{R}_\chi \propto 1/L^\gamma, \quad \gamma < d,$$

then it is said that the averaging of χ for the system under consideration is weak. If $\mathcal{R}_\chi$ tends to a nonzero constant with the increase in the system volume L^d, it is said that no averaging of χ in the system occurs.

The method used for constructing clusters in the DLA model is stochastic, although no complete analogy with quenched impurities in a thermodynamic system can be established. Nevertheless, one may expect that weak self-averaging would occur in the DLA model.

In our system, the linear size is determined by the cluster radius $F^{(1)}$. One can vary the size of particles λ_0 instead, while leaving $F^{(1)}$ invariable; in other words, one can determine the power γ in the expression for the relative fluctuation of D:

$$\mathcal{R}_D \propto \lambda_0^{\gamma/2} \propto n^{-\gamma/D}. \qquad (12)$$

Thus, in the Hastings–Levitov model, the dependence of the relative fluctuation of the fractal dimension on the initial size of the bump contains information on the degree of its self-averaging.

To analyze self-averaging of D, we carried out several series of simulations for

$$\lambda_0 = (0.5;\ 1.0;\ 1.5;\ 2.0;\ 3.0;\ 5.0;\ 8.0) \times 10^{-5}$$

($\alpha = 1$) and for

$$\lambda_0 = (0.3;\ 0.5;\ 1.0;\ 1.5;\ 3.0;\ 4.0;\ 5.0) \times 10^{-5}$$

($\alpha = 2.5$). For every λ_0, nine series 100 realizations in each were obtained. For each realization, 14 points for certain values of $F^{(1)}$ up to $F^{(1)} = 1.1$ were taken. The values of $\mathcal{R}_D$ were then averaged over nine series.

The plots for $\mathcal{R}_D(\lambda)$ and their approximations are presented in Figs. 4a and 4b on a log-log scale. The linear approximation yields the following values of γ in (12): $\gamma = 1.268 \pm 0.034$ for $\alpha = 1$ and $\gamma = 0.510 \pm 0.036$ for $\alpha = 2.5$.

For both models, D is weakly self-averaging. This follows from the fact that the exponent γ/D in (12) is less than unity (with regard for the fact that $D = 1.71$ for $\alpha = 1$ and D is considerably greater than unity for $\alpha = 2.5$). For example, for $\gamma = 1$, the exponent $\gamma/D = 0.74$. Therefore, $\mathcal{R}_D$ decreases slower than the inverse "cluster volume" $1/n$ with the increase in n (whereas in systems with full self-averaging, these values decrease with the same rate).

If we assume that the behavior of fluctuations of the fractal dimension computed by the conventional method [4, 7, 13] (i.e., on the basis of the dependence $F^{(1)} \propto n^{1/D}$) is similar to the behavior of fluctuations of the Davidovitch–Procaccia fractal dimension (11), then our result also explains the slow convergence of the fractal dimension with increasing system size.

4.5. *The Behavior of D at a Large Number of Mappings*

According to the Davidovitch–Procaccia method [6], at a large number of mappings n, the values of the fractal dimension $D(n;\lambda_0)$ are independent of both n and λ_0, which makes it possible to assume the existence of a unique function $F_*^{(1)}(x)$. This assumption provides a basis for the method of determining the fractal dimension.

To check the validity of this assumption, we numerically determined the fractal dimension D as a function of n and λ_0 at $n > n_0$.

The function $D(n;\lambda_0)$ for various values of λ_0 was investigated for the DLA model ($\alpha = 1$) at $n \lesssim 3n_0$. We used the technique described in Subsection 4.2. The plots of $D(n;\lambda_0)$ for various λ_0 are shown in Fig. 5. Everywhere (except of the points where the curves meet), the errors are much less than the distance between the curves.

It is seen from Fig. 5 that $D(n;\lambda_0)$ do not converge to a certain constant as the number of mappings n increases; moreover, there is a strong dependence of the fractal dimension $D(n;\lambda_0)$ on λ_0 at large n. The table presents the values of λ_0, the corresponding values of n_0, the number of realizations N, and the number $n_{\max}$ of mappings used in the computations. The last column of the table contains the fractal dimension $\bar{D}$ obtained by averaging $D(n;\lambda_0)$ over the interval $n \in [1.0 : 2.5]$. It is seen from the table that $\bar{D}$ varies from 1.77 at $\lambda = 10^{-5}$ to 1.70 at $\lambda = 10^{-8}$; i.e., they already differ in the second digit beyond the decimal point.

The behavior of $D(n;\lambda_0)$ at $\lambda_0 = 10^{-9}$ differs from its behavior at other λ_0. This fact can be explained as follows. The derivative of the conformal mapping function in (6) is strongly nonuniform at sizes of order $\sqrt{\lambda_0}$. For this reason, if a particle grows in the vicinity of an existing particle, the latter substantially distorts the size and shape of the former. This problem and possible ways to overcome it are thoroughly discussed in [13]. With a large number of particles, this effect can be substantial since, with decreasing λ_0 (and increasing n), the probability of forming a large number of particles at the same place increases. This feature casts doubt on the possibility of obtaining accurate results without monitoring the growth of particles and on the exact correspondence between the DLA and Hastings–Levitov models (see a detailed discussion in [13]).

The values of n_0, $n_{\max}$, N, and $\bar{D}$ for various λ_0

λ_0	n_0	$n_{\max}/n_0$	N	$\bar{D}$
4×10^{-5}	745	2.62	3200	1.769
10^{-5}	1490	1.95	3600	1.765
2.5×10^{-6}	2980	1.98	560	1.766
10^{-6}	4712	2.68	210	1.754
10^{-7}	14901	2.10	64	1.724
10^{-8}	47123	2.55	20	1.704
10^{-9}	149018	2.69	7	1.723

For a small depth of the layer of attached particles and a small size of particles $\sqrt{\lambda_0}$ compared with the initial radius, the problem under consideration is close to the problem of a flat layer growth [12]. Then, as the size of particles decreases, one can expect the Davidovitch–Procaccia fractal dimension to tend to the dimension of the flat layer structures, which is approximately equal to 1.66–1.68, (see [12] and references therein). If we plot the value of $\bar{D}$ given in the table versus $(-1/\ln\lambda_0)$, then the four of them (for $\lambda_0 = 2.5 \times 10^{-6}$, 10^{-6}, 10^{-7}, and 10^{-8}) fall onto a line, and the linear extrapolation yields $D = 1.6(1)$. Certainly, this extrapolation is very approximate; nevertheless, it supports our assumption that the fractal dimension in the problem of the growing flat layer can be roughly determined by the Davidovitch–Procaccia technique.

As further verification of the claims made by Davidovitch and Procaccia, we constructed analogs of the dependences of the function $F^{(1)}(x;\lambda) - 1$ on the “invariant” variable $x = \sqrt{\lambda_0}\,n^{1/D}$ (such dependences were obtained in [6]). The coincidence of those curves for various λ_0 would confirm the existence of a universal function $F_*^{(1)}(x)$.

These dependences are shown in Figs. 6a and 6b for various λ_0 and two values of the fractal dimension $D = 1.70$ and 1.75.[2] We chose these values of the fractal dimension, since the value $D = 1.70$, which is close to the conventional value of the dimension in the DLA model, was used in [6] in similar figures, and $D = 1.75$ yields the densest arrangement of curves corresponding to various values of λ_0.

---

[2] These dependences are similar to those shown in Figs. 3–5 in [6]; the difference is only in the scale used in the figures (we use a larger scale).

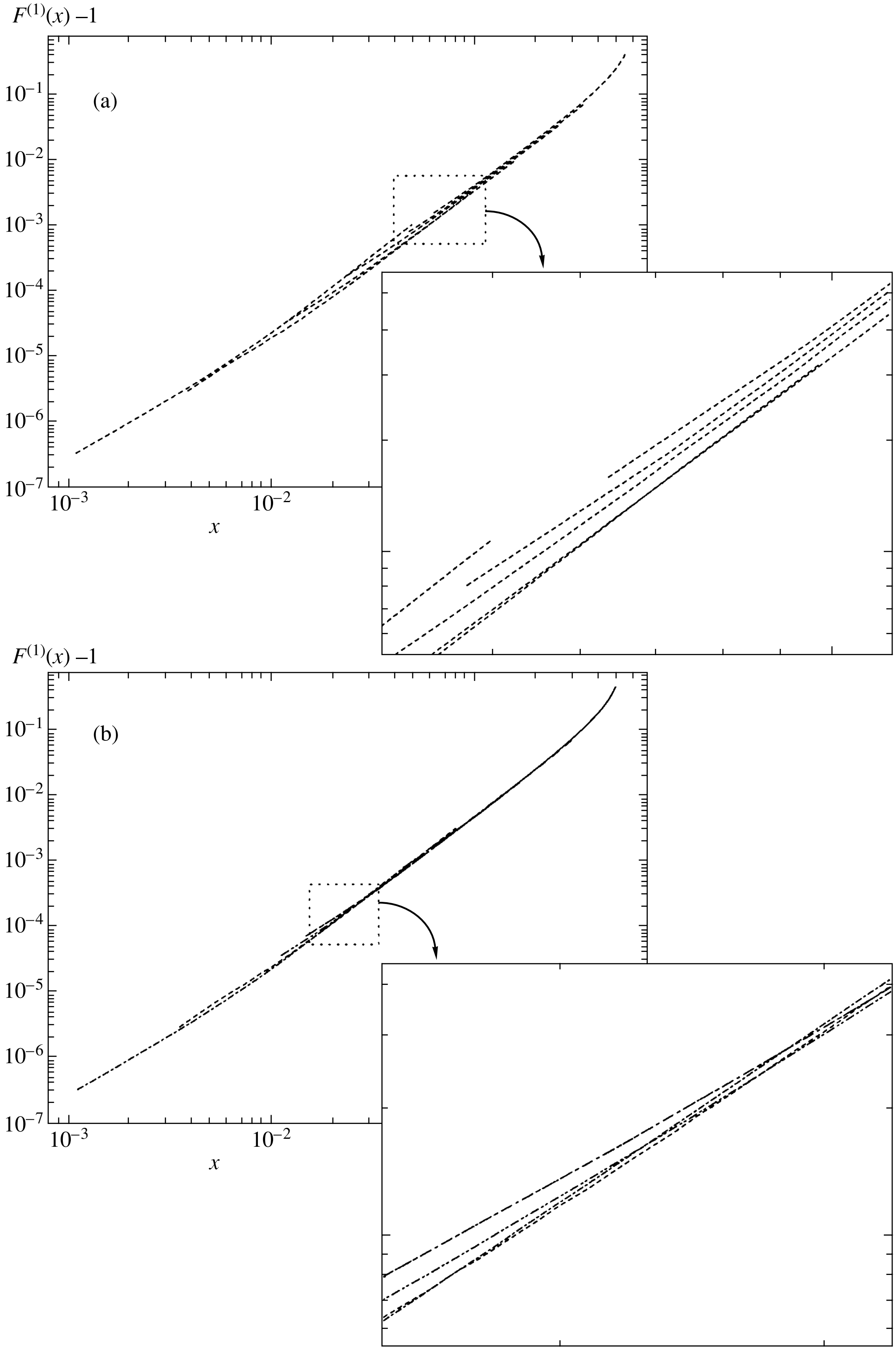


Fig. 6. $F^{(1)}(x; \lambda_0)$ as a function of $x = \sqrt{\lambda_0}\, n^{1/D}$ at $D = 1.70$ (a) and $D = 1.75$ (b) for various values of λ_0: 4×10^{-5}, 10^{-5}, 2.5×10^{-6}, 10^{-6}, 10^{-7}, 10^{-8}, and 10^{-9}. It is seen in the insets made on a larger scale that the curves are close to each other, but do not collapse.

An analysis of the figures suggests the following conclusions.

1. Since the curves corresponding to various λ_0 are closer to each other at $D = 1.75$ than at $D = 1.70$ (the latter case was considered in [6]), the existence of a unique function $F_*^{(1)}(x)$ at $D = 1.75$ seems more probable.

2. On a not very large scale (as the one used in Figs. 3–5 in [6]), it may seem that all $F^{(1)}(x; \lambda)$ collapse into a single function $F_*^{(1)}(x)$ both at $D = 1.75$ and at $D = 1.70$.

3. For different values of λ_0, the functions $F^{(1)}(x; \lambda)$ are close to each other, but they do not collapse. Moreover, their behavior is such that they would not collapse at any D. (At $D = 1.75$, there exist both parts of the plot where the curves collapse and parts where they do not. For another D, the curves will depart from each other on the intervals where they collapsed for $D = 1.75$.) Therefore, it seems probable that no unique function $F_*^{(1)}(x)$ exists.

Thus we can draw the following conclusion: the approach proposed in [6] does not have enough support and hardly can be used to obtain highly accurate results.

5. CONCLUSIONS

The simulation described in this paper suggests the following conclusions.

1. The distribution of D in various realizations of clusters is near Gaussian. D is weakly averaged, which can explain a slow convergence in the conventional computation of the fractal dimension.

2. $D(n; \lambda_0)$ relaxes exponentially to a certain value (the relaxation occurs at sizes of order $n_0/2$); therefore, the cluster considered at the size of order n_0 can be assumed to be steady.

3. The analysis of the behavior of $D(n; \lambda_0)$ shows that the method proposed in [6] cannot yield an arbitrarily accurate value of the fractal dimension at $\lambda_0 \longrightarrow 0$. It appears that, in this case, the dimension tends to the dimension of the growing flat layer problem. This is a quite different problem, which requires a special investigation.

ACKNOWLEDGMENTS

We are grateful to L.S. Levitov and M.G. Stepanov for critical remarks. The work was supported in part by the Russian Foundation for Basic Research, project no. 99-02-18412. L.N. Shchur also thanks the Cariplo Foundation and the Centro Volta–Landau Network (Italy), as well as the cooperation program between École Normale Superior (Paris) and Landau Institute for Theoretical Physics.

REFERENCES

1. T. C. Halsey, B. Duplantier, and K. Honda, Phys. Rev. Lett. **78**, 1719 (1997).
2. T. A. Witten and L. M. Sanders, Phys. Rev. Lett. **47**, 1400 (1981).
3. T. C. Halsey, Phys. Today **53**, 36 (2000).
4. M. B. Hastings and L. S. Levitov, Physica D (Amsterdam) **116**, 244 (1998).
5. M. B. Hastings, Phys. Rev. E **55**, 135 (1997).
6. B. Davidovitch and I. Procaccia, Phys. Rev. Lett. **85**, 3608 (2000).
7. P. Ossadnik, Physica A (Amsterdam) **195**, 319 (1993).
8. L. Nimeyer, L. Pietronero, and H. J. Wiessmann, Phys. Rev. Lett. **52**, 1033 (1984).
9. *Fractals and Disordered Systems*, Ed. by A. Bunde and S. Havlin (Springer-Verlag, Berlin, 1996).
10. S. Wiseman and E. Domany, Phys. Rev. Lett. **81**, 22 (1998); Phys. Rev. E **58**, 2938 (1998).
11. E. Somfai, L. M. Sander, and R. C. Ball, Phys. Rev. Lett. **83**, 5523 (1999).
12. B. Kol and A. Aharony, Phys. Rev. E **63**, 046117 (2001).
13. M. G. Stepanov and L. S. Levitov, Phys. Rev. E **63**, 061102 (2001).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 154–165.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 175–187.
Original Russian Text Copyright © 2002 by Vecheslavov, Chirikov.

Diffusion in Smooth Hamiltonian Systems

V. V. Vecheslavov* and B. V. Chirikov**
Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 11,
Novosibirsk, 630090 Russia
**e-mail: vecheslavov@inp.nsk.ru*
***e-mail: chirikov@inp.nsk.ru*

Received September 6, 2001

Abstract—A family of models determined by a smooth canonical 2D-map that depends on two parameters is studied. Preliminary results of numerical experiments are reported; they are evidence of substantial suppression of global diffusion in a wide range of perturbation values. This effect is caused by the little-known phenomenon of the conservation of resonance separatrices and other invariant curves under the conditions of strong local dynamic chaos. Such a total suppression of diffusion occurs although invariant curves are only conserved for a countable zero-measure set of parameter values. Simple refined estimates of diffusion rates in smooth systems without invariant curves were obtained and numerically substantiated. The principal boundary of diffusion suppression in a family of models with invariant curves was described by a semiempirical equation in dimensionless variables. The results were subjected to a statistical analysis, and an integral distribution for diffusion suppression probability was obtained. *© 2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

One of the principal concepts of the modern theory of nonlinear Hamiltonian systems is the assertion that splitting of a nonlinear resonance separatrix and the formation of a chaotic layer in its place in a typical (that is, nonintegrable) system occurs under an almost arbitrary perturbation. It is also believed that precisely separatrices are destroyed first, because the period of motion along them is infinite, and the interaction of nonlinear resonances in their vicinity is always substantial (e.g., see [1–4]). Invariant curves (surfaces) experience breakup and disappear as perturbation increases. This as a rule causes overlapping of chaotic layers of all resonances and the appearance of so-called “global” chaos.

The conditions of global chaos formation and the possibility of diffusion over the whole unified chaotic component in the phase space depend not only on the value but also on the smoothness of perturbation. The smoothness can conveniently be characterized by the rate of decreasing Fourier amplitudes. An analytic perturbation decreases exponentially, and a threshold perturbation value ε_{tr} always exists. Global diffusion only arises at $\varepsilon \gtrsim \varepsilon_{tr}$. If $\varepsilon < \varepsilon_{tr}$, chaos is localized in comparatively narrow chaotic layers (which are formed at arbitrary $\varepsilon > 0$ values), and there can be no global diffusion in a conservative system with the number of degrees of freedom $\mathcal{N} \le 2$.

Note that, if $\mathcal{N} > 2$, global diffusion can only take place under special initial conditions (Arnold diffusion, e.g., see [1]). The rate of this diffusion and the measure of its region decrease exponentially with respect to the $1/\varepsilon$ parameter as $\varepsilon \longrightarrow 0$.

The character of motion changes considerably for a smooth perturbation of the Hamiltonian whose Fourier amplitudes decrease as some power $\beta + 1$ of their number n (e.g., see [5] and the references therein). In the simplest case of a 2D map, to which our analysis will be restricted, the $\varepsilon_{tr} > 0$ threshold of the appearance of global chaos always exists if $\beta > \beta_{cr} = 3$. This critical smoothness value was obtained from the simple estimate made in [5] (also see Section 3), but it nevertheless requires verification by numerical experiments. So far as we know, a rigorous proof can only be obtained for $\beta_{cr} = 5$ (see [6], where the suggestion is made that, in reality, $\beta_{cr} = 4$). This uncertainty is of no significance for our purposes, because, for the model under consideration (Section 2), the inequality $\beta = 2 < \beta_{cr}$ is always satisfied. Interestingly, the situation has long remained unclear precisely for the $\beta = 2$ index.

Even in early numerical experiments on systems whose smoothness was lower than critical, trajectories that did not go beyond some limited phase space region in long-time computations were observed along with global diffusion [7, 8]. This was, however, nothing more than a suspicion of diffusion suppression or weakening. A rigorous result was obtained by Bullett [9], who proved the existence of global invariant curves with both irrational and rational rotation numbers for a symmetrical piecewise linear 2D map [$\beta = 2$—see (2.1) and (2.2) below; also see [10] and Section 2]. Precisely global invariant curves have a complete phase extent, which prevents unlimited diffusion over action. In [9], it was found for the first time that, among invariant curves with rational rotation numbers, there are also (at special perturbation parameter values) intact nonlinear

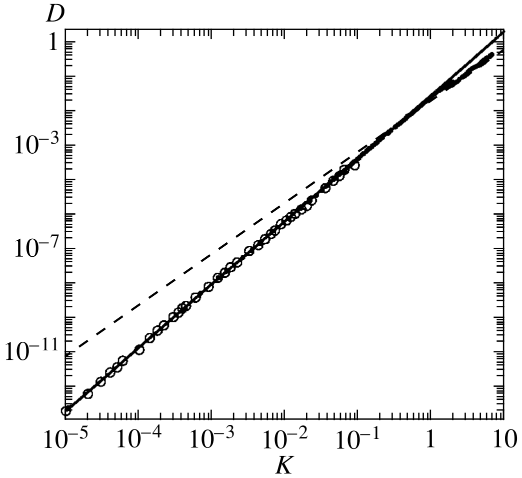


Fig. 1. Diffusion rate $D(K)$ in model (2.2) with the $d = 0$ parameter (without invariant curves): dots are the data from [17], open circles are our data averaged over 250 trajectories $t_0 = 4 \times 10^7$ iterations long with random initial conditions, solid line is power dependence (4.1), and dashed line is limiting mode (3.10).

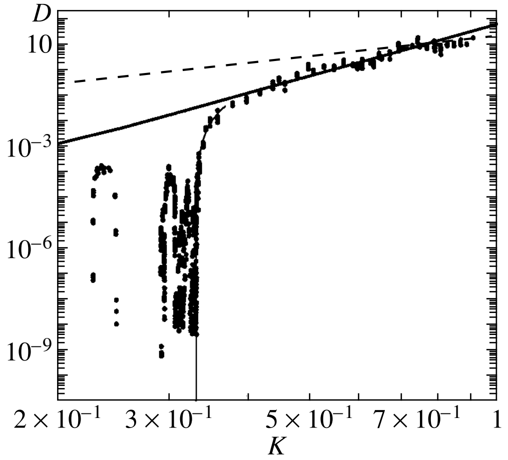


Fig. 2. Example of diffusion in a smooth map with invariant curves of measure zero: model (2.2) with the $d = 1/2$ parameter, 50 trajectories, computation time $t_0 = 2.5 \times 10^6$. The solid straight line is $D_2 = 0.8K^5$ according to (5.1), and the dashed line is limiting mode (3.10).

resonance separatrices. Especially important and unexpected was that the system nevertheless remained nonintegrable, but the separatrix not only was conserved under strong chaos conditions but also prevented global diffusion.

A similar theorem for the same model was later proved by Ovsyannikov [11], who not only specified a countable set of parameter values at which the separatrix of an integer resonance is conserved but also found an explicit and very simple expression for the separatrix. Although Ovsyannikov proved his theorem independently, such a coincidence of the models was not fortuitous, because solving a linear (even though piecewise) map considerably simplifies the problem. Note that completely solving even a linear map is only possible if the separatrix is conserved, because otherwise two branches of a split separatrix form random trajectories. For the same reason, a symmetrical piecewise linear 2D map cannot be simplified to a purely linear map of the type of the Arnold map, in which nonlinear resonance separatrices are always split. The mathematical works by Bullett and Ovsyannikov are therefore restricted to studies of only new-type invariant curves themselves. The first examples of such curves were predicted in [10].

Precisely the Ovsyannikov theorem prompted us to thoroughly study the symmetrical piecewise linear 2D map and its modifications [12–15]. Unfortunately, this theorem was not published by its author (the complete formulation of the theorem can be found in Appendix in [14]). Instead, the theorem was generalized in [16] to arbitrary map parameter values. The result obtained in [16] contradicts that of [9] and our numerical experiment data.

A certain perturbation parameter value corresponds to each invariant curve in the symmetrical piecewise linear 2D map (including invariant curves of the new type with rational rotation numbers, that is, also integer and fractional resonance separatrices). The set of all such values is a Cantor set (see Figs. 2 and 3 in [9]), and there are intervals of parameter values in which global diffusion certainly takes place (one of such intervals is identified in [9]; also see Section 5). As the density of this set is fairly high, we can expect strong (although incomplete) suppression of global diffusion at an arbitrary perturbation parameter value. Studies in this direction were performed in the present work.

2. MODEL

The selected model is a two-dimensional map in the canonical variables of action p and phase x,

$$\bar{p} = p + Kf(x), \quad \bar{x} = x + \bar{p} \mod 1. \tag{2.1}$$

Here, $K = \varepsilon > 0$ is the perturbation parameter (not necessarily small), and "force" $f(x)$ has the form of an antisymmetric $[f(-y) = -f(y),\ y = x - 1/2]$ piecewise linear "saw" with period 1.

We will study the whole family of sawtooth perturbations[1] (see Fig. 1 in [15]),

$$f(x) = \begin{cases} \dfrac{2x}{1-d}, & \text{for } |x| \le \dfrac{1-d}{2}, \\ -\dfrac{2y}{d}, & \text{for } |y| \le \dfrac{d}{2}, \end{cases} \tag{2.2}$$

---

[1] A family similar but not identical to that of the model used in [9].

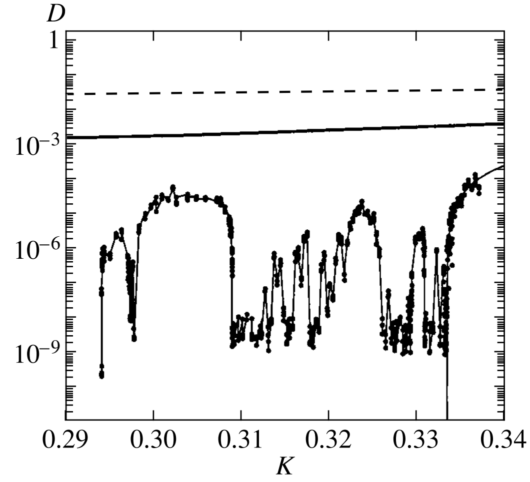


Fig. 3. Part of fractal diffusion region extended along K with invariant curves shown in Fig. 2. The computation parameters are the same as in Fig. 2 except that $t_0 = 4 \times 10^7$ for most of the points. For some points, including the leftmost with $K = 0.294$, computation time $t_0 = 10^9$. The smooth curve at the right was constructed by (5.5); it approximates the boundary of the principal diffusion region.

where $y = x - 1/2$, and $0 \le d \le 1$ is the distance between saw "teeth" $|f(x)| = 1$ situated at the points $y = y_\pm = \pm d/2$. The most thoroughly studied particular case of the symmetrical piecewise linear 2D map corresponds to the bias parameter $d = 1/2$.

We are interested in global diffusion over momentum, which is suppressed by invariant curves with a complete extent over phase. We call these invariant curves global and, in what follows, only consider such curves unless otherwise stated. Note that these invariant curves (including unbroken separatrices of integer and fractional resonances) exist for an arbitrary bias parameter value in the interval $0 < d < 1$ at special K values [9, 12–15].

If $0 < d < 1$, force (2.2) can be written as the Fourier series

$$f(x) = \sum_{n \ge 1} \frac{f_n}{n^{\beta}} \sin(2\pi n x), \qquad (2.3)$$

where

$$f_n = -\frac{2}{\pi^2} \frac{\cos(n\pi)\sin(n\pi d)}{d(1-d)}, \quad \beta = 2. \qquad (2.4)$$

The passage to the $d \longrightarrow 0$ limit in (2.3) and (2.4) for the discontinuous saw yields

$$f_n = -\frac{2}{\pi}\cos(n\pi), \quad \beta = 1. \qquad (2.5)$$

One can see that, in the $d = 0$ limit, the smoothness index of the system, β, is smaller by one than within the interval, and both indices are smaller than the $\beta_{\text{cr}} = 3$ critical value. The second $d \longrightarrow 1$ limit is not considered in this work, because motion then becomes regular (see [15]).

Map (2.1) can be written as a continuous system with a Hamiltonian that explicitly depends on time and perturbation in the form of impulses [1–3, 14]:

$$\begin{aligned} H(x, p, t) &= \frac{p^2}{2} \\ &+ K \sum_{n \ge 1} \frac{f_n}{2\pi n^{\beta+1}} \cos(2\pi n x)\delta_1(t), \end{aligned} \qquad (2.6)$$

where

$$\delta_1(t) = 1 + 2\sum_{m \ge 1} \cos(2\pi m t)$$

denotes the δ function with period 1. Note that selecting the δ function in this form also defines the time unit of the continuous system; this unit equals one iteration of the initial map.

Each term of the double sum in (2.6) is proportional to $\cos[2\pi(nx - mt)]$ with integer m and n and determines "its own" primary nonlinear resonance (details are given, e.g., in [1]). Supposing that these resonances do not interact with each other, we can describe any of them by the Hamiltonian of a "pendulum"

$$H_{nm}(x, p, t) = \frac{p^2}{2} + \frac{K f_n}{2\pi n^{\beta+1}} \cos[2\pi(nx - mt)]. \qquad (2.7)$$

Introducing the resonance momentum value $p_{nm} = \dot{x}_{nm} = m/n$, we find that, in the new variables $\tilde{x} = nx - mt$ and $\tilde{p} = (p - p_{nm})/n$, each such solitary resonance is a conservative system with motion strictly bounded with respect to momentum. Returning to the old variables, we determine the frequency of phase oscillations

$$\Omega_n^2 = \frac{2\pi K f_n}{n^{\beta-1}} \qquad (2.8)$$

and the total momentum width of the resonance

$$(\delta p)_n = 4\sqrt{\frac{K f_n}{2\pi n^{\beta+1}}}. \qquad (2.9)$$

In the next section, these equations are used to obtain very simple and unexpectedly accurate estimates of the rate of diffusion for a smooth map without invariant curves.

3. DIFFUSION RATE ESTIMATES

Our estimates are based on the criterion of overlapping of nonlinear resonances (see [1–3, 5]), which can be written in the simplest form as

$$\mathcal{P} \sim 1, \qquad (3.1)$$

where

$$\mathcal{P} \sim \sum_{n \ge 1} n(\delta p)_n = 4\sqrt{\frac{Kf_0}{2\pi}}\mathcal{S}, \qquad \mathcal{S} = \sum_{n \ge 1} n^{(1-\beta)/2}. \tag{3.2}$$

Here, $\mathcal{P}$ is the approximate sum of widths (2.9) of all primary resonances in a unit interval with respect to momentum p. For simplicity, we assume that all Fourier coefficients are equal, $f_n = f_0$.

Note that the sum diverges at $\beta \le 3$, and this determines the critical smoothness value specified above, $\beta_{cr} = 3$, in the approximation that we use. Global diffusion then occurs at an arbitrary K value, including $K \longrightarrow 0$, and, generally, its rate depends on all resonances (2.7) and is described by complex and cumbersome equations (cf. [17]). As in [5], our simple estimates are based on the following hypothesis, which we consider physically plausible and which is substantiated by the numerical experiments described below. We assume that the mean rate of global diffusion at $\beta \le 3$ is largely determined by a finite number of resonances up to some critical harmonic $n = n_c$. These resonances, in combination with various m values in (2.7), provide overlapping (3.1). Indeed, stronger resonances ($n < n_c$) cause faster diffusion, but this diffusion is local because of incomplete overlapping of resonances. On the other hand, weaker resonances ($n > n_c$) more than provide overlapping, but the rate of diffusion for them

$$D = \frac{\overline{(\delta p)^2}}{t} \sim \frac{(\delta p)_n^2 \Omega_n}{2\pi} \tag{3.3}$$

rapidly decreases as n increases. Here, the total resonance width (2.9) and the period of the corresponding phase oscillations $2\pi/\Omega_n$ [see (2.8)] are used as dynamic diffusion scales.

Replacing the sum in (3.2) by the integral in $n \gg 1$ yields

$$\mathcal{S} \approx \frac{2}{3-\beta} n^{(3-\beta)/2}. \tag{3.4}$$

It follows from (3.2) that the number of the critical harmonic is

$$n_c \sim \left[\frac{\pi}{32}\frac{(3-\beta)^2}{Kf_0}\right]^{1/(3-\beta)}. \tag{3.5}$$

Lastly, (3.3) is used to obtain the rate of diffusion,

$$D_{\beta<3}(K) \sim \frac{4\sqrt{2\pi}}{\pi^2} \times \left(\frac{32}{\pi(3-\beta)^2}\right)^{(3\beta+1)/2(3-\beta)} (Kf_0)^{5/(3-\beta)}. \tag{3.6}$$

Generally, this equation only gives an estimate of the order of magnitude. We, however, deliberately leave various numerical coefficients in it in the hope that its accuracy can be substantially increased through the introduction of empirical correction factors. This will be done in the next section in considering a simple and thoroughly studied example of such diffusion.

Note that all these estimates are only valid at $K \ll 1$. Indeed, as mentioned, diffusion in the system under consideration has two dynamic scales,

$$(\delta p)_c \sim 4\sqrt{\frac{Kf_0}{2\pi n_c^{\beta+1}}} \sim K^{2/(3-\beta)} \lesssim 1 \tag{3.7}$$

with respect to momentum (2.9) and

$$t_c \sim \sqrt{\frac{2\pi n_c^{\beta-1}}{Kf_0}} \sim K^{-1/(3-\beta)} \gtrsim 1 \tag{3.8}$$

with respect to time [from (2.8)]. Both scales are bounded because p is periodic and t discrete. This imposes the limitation on the system parameter

$$K \lesssim 1. \tag{3.9}$$

The same limitation arises also from the condition imposed on critical harmonic (3.5), namely, $n_c \gtrsim 1$. At $K \ll 1$, the time of the decay of perturbation correlations satisfies the inequality $t_c \gg 1$. At $K \sim 1$, this time is shortened to $t_c \sim 1$, and at $K \gg 1$, correlations between closely spaced impulses become negligibly small. The rate of diffusion is then determined by the mean square perturbation,

$$D_{K\to\infty} = K^2\int_0^1 f^2(x)dx = \frac{K^2}{3}, \tag{3.10}$$

and ceases to depend on bias parameter d for the whole family of maps (2.1), (2.2).

4. NUMERICAL EXPERIMENTS FOR A MODEL WITHOUT INVARIANT CURVES

First, consider the simplest example of a smooth 2D map, which corresponds to the $d = 0$ parameter in family (2.2). According to [15], invariant curves are absent in this limiting case, and global diffusion occurs at arbitrary positive $K > 0$. Diffusion in such a model was thoroughly studied numerically and analytically fairly long ago [17]. Note that both the diffusion rate and perturbation parameter K are normalized differently in [17] and this work, and the data obtained in [17] and cited below were therefore recalculated to our model.

Substituting $\beta = 1$ and $f_0 = 2/\pi$ [see (2.5)] into (3.6) yields $D_{\beta=1}(K) = 0.84K^{2.5}$. Recalculating the value numerically obtained in [17] yields

$$\begin{aligned} D_1(K) &= A_1 K^{B_1} = C_D D_{\beta=1}(K), \\ A_1 &= 0.5680 \pm 0.0034, \\ B_1 &= 2.4940 \pm 0.008, \end{aligned} \tag{4.1}$$

where the correction factor for the diffusion rate $C_D = 0.68$. In what follows, we use the assumption made by the authors of [17], namely, $B_1 = 2.5$, for comparison with their theory.

The C_D correction to our theoretical estimate is small, but it heavily depends on the smoothness index β. By way of example, we set $\beta = 2$. Equation (3.6) gives $D_{\beta=2}(K) = 1199K^5$ against the numerical calculation result $D_2(K) \approx 0.8K^5$ [see (5.1) and Fig. 2]. A comparison of these values gives a correction factor $C_D \approx 6.7 \times 10^{-4}$ (!), which makes this factor physically meaningless.

Physically unreasonable results are obtained because the principal approximation for constructing estimates is related to an intermediate result in the formulation of the condition of resonance overlapping, Eqs. (3.1) and (3.2), rather than directly to diffusion rate D. In such a form, this condition always overestimates the overlapping effect because it contains maximum widths of resonance separatrices, whereas, in reality, separatrices may have mutual phase shifts. This circumstance can be taken into account by introducing correction C_s in place of C_D,

$$\mathcal{P} = C_s > 1. \tag{4.2}$$

C_s is essentially different from C_D in (4.1) because it is raised to some power, which depends on β and can be fairly large,

$$\begin{aligned} D_{\beta<3}(K) &\approx \frac{4\sqrt{2\pi}}{\pi^2} \\ &\times \left(\frac{32}{\pi(3-\beta)^2 C_s^2}\right)^{(3\beta+1)/2(3-\beta)} (K f_0)^{5/(3-\beta)}. \end{aligned} \tag{4.3}$$

To obtain $C_D \approx 1$ for $\beta = 2$ (see above), it suffices to set $C_s \approx 2.84$ [see (5.1)]. This is evidence that such a method for introducing an empirical correction into order-of-magnitude estimates is very effective.

The C_s correction is much smaller for $\beta = 1$, $C_s \approx 1.2$, because the perturbation spectrum at $\beta = 1$ contains all harmonics, whereas only odd harmonics remain at $\beta = 2$. As a result, sum (3.2) decreases twofold. This additional effect is easy to take into account in (4.3) by the replacement $C_s \longrightarrow 2C_s$. The necessary correction for $\beta = 2$ then decreases from 2.84 to 1.42, which is close to the $C_s \approx 1.2$ value found above for $\beta = 1$.

The most important results obtained in studying the simple model with $d = 0$ and the approximating straight line found in [17] [Eq. (4.1)] are shown in Fig. 1. This straight line is also fairly well described by our simple theory (4.3) with correction $C_s = 1.2$. Both equations closely agree with the empirical data up to the $K \sim 1$ value, at which the transition to mode (3.10) occurs.

Numerical data were compared in [17] with a very complex theory developed by the authors, which was also based on the concept of overlapping of resonances (more exactly, of their destroyed separatrices). This theory did not include adjustment parameters of any kind, but the accuracy that it provided was in reality not high (approximately 10%, see Fig. 2 in [17]). The theory actually referred to a qualitatively different model with an analytic Hamiltonian with a finite number of harmonics n retained in its Fourier transform. Such a "cutting off" of the spectrum was accompanied by the appearance of a threshold for the arising of global diffusion, as is characteristic of analytic systems, which limited the applicability of this theory to the most interesting region of small parameter K values, $K \lesssim K_c(n)$. In the example given in [17] (see Fig. 2 in [17]), $n_c = 21$ and $K_c \sim 0.02$ (in our normalization). The mechanism of this limitation resembles that of restriction (3.5) imposed on the critical harmonic in our theory, but our restriction is much weaker. For instance, at $n_c = 21$ and $C_s = 1.2$, the minimum value is

$$K \approx K_c(n_c) \approx \left(\frac{\pi C_s}{4 n_c}\right)^2 \approx 0.002, \tag{4.4}$$

which is one order of magnitude smaller than in [17].

We turn to the most interesting part of our study, when the bias parameter of model (2.2) is $d \neq 0$. It has been proved in [9] that there exists a critical perturbation parameter value K_B such that, at $K > K_B$, there is no global invariant curves in the system. The exact equation for K_B in our normalization takes the form

$$K_B(d) = \frac{2d^2}{1+d}, \quad 0 < d < 1. \tag{4.5}$$

According to [9–15], generally, there is a countable set of special $K \leq K_B$ values at which invariant curves are formed in the system under strong local chaos conditions.

Our main interest is how strongly the existence of these invariant curves suppresses global diffusion at arbitrary K values, although the measure of the set of special K values and the probability of fortuitously falling into it are zero.

5. DIFFUSION SUPPRESSION BY "VIRTUAL" INVARIANT CURVES

First consider the most thoroughly studied example of family (2.2) with the bias parameter $d = 1/2$, for which the results of our numerical experiments are shown in Figs. 2 and 3.

In the computations, the whole time interval was divided into four equal portions, and the diffusion rate was output at the end of each interval. For this reason, four points generally correspond to each K value in

Figs. 2–5. Their arrangement along vertical lines allows the diffusive case to be distinguished from the nondiffusive one. Characteristic of the latter is a decrease in the rate of diffusion in time, which results in mutual "dispersal" of these points. The spread of points for diffusion characterizes the accuracy of diffusion rate values. In the region without invariant curves, this accuracy is quite satisfactory (approximately 10%), as for $d = 0$. However, in the region with virtual invariant curves and in its vicinity (at $K \sim K_B$), the error increases and sometimes reaches 2. Of the same order is a systematic decrease in the mean diffusion rate caused by nonergodicity of motion in this region. Nonergodicity of motion results in the formation of a stable motion component of a very complex structure (so-called critical structure, e.g., see [5]). Currently, we have not been able to substantially increase the accuracy of determining the rate of diffusion. However, we believe the attained accuracy to be sufficient for our purposes.

The $d = 1/2$ value is the only one (except the $d = 0$ limiting value) when the $|f_n|$ = const simplification in (3.2) is possible. This simplification substantially facilitates a theoretical analysis of numerical data. By virtue of this simplification, Eq. (4.3) with $|f_n| = 8/\pi^2$ is applicable to odd harmonics, and the same equation with $|f_n| = 0$, to even harmonics at $\beta = 2$ with correction $C_s = 2.84$ [see (2.4) and the preceding section], but only in the region without invariant curves. In addition, (4.3) is only valid in a very limited range of perturbation parameter values (see Fig. 2), namely,

$$D_2(K) = 0.8K^5, \quad 0.4 \lesssim K \lesssim 0.8. \tag{5.1}$$

The upper bound is determined by the well-known transition to the limiting diffusion conditions without correlations (3.10). An essentially new feature of the diffusion picture is the lower bound, clearly related to the appearance of invariant curves, which suppress diffusion. According to (4.5), invariant curves at $d = 1/2$ are completely absent if

$$K > K_B\left(\frac{1}{2}\right) = \frac{1}{3}. \tag{5.2}$$

It is, however, obvious from Figs. 2 and 3 that strong diffusion suppression begins much earlier, that is, in the region of K values where invariant curves are actually absent! Hence our new term "virtual invariant curve." In other words, every real invariant curve, which is formed at some strictly definite special $K = K_0$ value, in reality substantially distorts the structure of the phase plane of the system in some finite neighborhood of K_0.

Diffusion in the vicinity of a single invariant curve in system (2.2) with $d = 1/2$, in the vicinity of an undestroyed integer resonance separatrix formed at $K = K_0 = 1/8$, was for the first time studied in [13]. The first thing observed was sharp asymmetry at $K > K_0$ and $K < K_0$.

At $K > K_0$, the separatrix begins to transmit other trajectories, but the mean time (the number of iterations) T_c of passing the resonance depends on the detuning $K - K_0 > 0$. The following measurements were preformed to determine this dependence in the interval $1.25 \times 10^{-7} \le K - K_0 \le 1.25 \times 10^{-5}$, where there was no other invariant curves. In the region between two neighboring integer resonances (see Fig. 4 in [13]), 100 random chaotic trajectories were generated, and time T_c of the first appearance of each of them in the region either below the lower or above the higher resonance was fixed. To facilitate comparison of these data with the results of the present work, we give the equation for the diffusion rate

$$\begin{gathered} D_c(K) \approx \frac{1}{\langle T_c \rangle} = 0.089(K - K_0)^{1.193}, \\ K > K_0 = \frac{1}{8}. \end{gathered} \tag{5.3}$$

The $F_{\text{ex}} = D_2(K)/D_c(K)$ ratio, where, according to (5.1), $D_2(K)$ determines the diffusion rate on the assumption of the complete absence of invariant curves in the system, is the quantitative measure of diffusion suppression. For instance, in the whole $K - K_0$ range that we studied, this coefficient changed from 40 000 to 200, which was evidence of substantial diffusion suppression (also see Figs. 2 and 3).

Note beforehand that (5.3) is in a certain way similar to Eq. (5.5) for the boundary of the principal diffusion region; this similarity is discussed below.

At $K < K_0$ and in the region arbitrarily close to the separatrix, many closely spaced invariant curves were observed; because of their presence, the problem of determining the resonance passage time was virtually unsolvable. The question of diffusion in this region remains open.

All these fairly simple observations are, we believe, most important at the same time, because they show that the zero measure of the set of invariant curves and even their finite density do not prevent strong diffusion suppression in the model under consideration.

The next important problem is that of quantitatively estimating diffusion suppression. Complete diffusion suppression is likely to be possible only at special K_0 values, that is, only for real invariant curves. Generally, everything depends on their structure in the space of system parameters (K, d). This structure appears to be fairly complex and is likely to be fractal. In particular, it also includes whole regions of finite width without invariant curves. One of such regions, $0.2295 < K < 0.2500$, has been predicted in [9] and is well seen at the left of Fig. 2. The rate of diffusion rapidly decreases at the boundaries of this region, as at the principal boundary $K = 1/3$ (5.2). This causes diffusion suppression

Parameters of diffusion regions shown in Fig. 3

$K_{\max}$	$D_{\max}$	F_{ex}	C_F	$\Delta K \times 100$	$\delta K \times 100$
0.3322	9.9×10^{-7}	3.26×10^{3}	1.99×10^{0}	0.080	0.051
0.3309	4.0×10^{-6}	7.93×10^{2}	1.30×10^{0}	0.155	0.130
0.3282	1.0×10^{-8}	3.05×10^{5}	1.20×10^{2}	0.060	0.003
0.3270	5.0×10^{-8}	5.98×10^{4}	1.85×10^{1}	0.051	0.007
0.3240	1.1×10^{-5}	2.60×10^{2}	1.65×10^{0}	0.382	0.274
0.3216	2.0×10^{-6}	1.38×10^{3}	1.15×10^{0}	0.099	0.090
0.3204	1.0×10^{-7}	2.70×10^{4}	5.58×10^{0}	0.039	0.012
0.3196	8.0×10^{-7}	3.34×10^{3}	1.73×10^{0}	0.072	0.050
0.3178	2.0×10^{-6}	1.30×10^{3}	4.01×10^{-1}	0.051	0.094
0.3163	1.0×10^{-6}	2.53×10^{3}	1.84×10^{0}	0.090	0.060
0.3146	4.0×10^{-7}	6.16×10^{3}	1.58×10^{0}	0.045	0.033
0.3139	8.0×10^{-7}	3.05×10^{3}	7.81×10^{-1}	0.045	0.053
0.3130	5.0×10^{-8}	4.81×10^{4}	2.04×10^{1}	0.063	0.008
0.3043	4.0×10^{-5}	5.22×10^{1}	1.59×10^{0}	1.090	0.798
0.2978	3.0×10^{-7}	6.25×10^{3}	4.05×10^{-1}	0.018	0.033
0.2960	2.5×10^{-6}	7.27×10^{2}	3.22×10^{0}	0.301	0.138
0.2941	4.3×10^{-7}	4.06×10^{3}	2.00×10^{-1}	0.0150	0.044

Note: $K_{\max}$ and $D_{\max}$ are the perturbation parameter and the rate of diffusion in the center of the region, respectively; F_{ex} is the experimental diffusion suppression factor in the center of the region; $C_F = F_{\text{ex}}/F_{\text{th}}$ is the ratio between the experimental and theoretical suppression factors; F_{th} is calculated by (5.6); ΔK is the experimental region width; and δK is the region width recalculated by (5.5) under the additional requirement of providing the $C_F \equiv 1$ equality.

even at the maximum. The width of this region is $\Delta K \approx 0.02$, and the diffusion suppression factor is

$$F_{\text{ex}} = \frac{D_2(K_{\max})}{D_{\max}} \approx 14. \tag{5.4}$$

Here, $D_{\max} \approx 4.6 \times 10^{-5}$ is the maximum measured diffusion rate near the center of the region $K_{\max} \approx 0.24$, and $D_2(K_{\max}) \approx 6.4 \times 10^{-4}$ is the diffusion rate expected in the complete absence of invariant curves in the system, Eq. (5.1).

A crude estimate of F_{ex} can be made by comparing the boundary of this region and the $D_B(K)$ principal boundary, which, according to Fig. 2, is satisfactorily approximated by the equation (to the left of the arrow)

$$\begin{aligned} D_B(K) &\approx \frac{1}{2}\left(K - \frac{1}{3}\right)^{3/2} \\ \longrightarrow D_b(K) &\approx \frac{1}{2}(K - K_b)^{3/2}. \end{aligned} \tag{5.5}$$

This particular case substantiates the natural suggestion that the K_B critical value from (4.5), which is the boundary value for the principal region without invariant curves, coincides with K_D (or is close to it) at the $D_B(K_D) = 0$ boundary of the principal diffusion region.

In addition, we make the suggestion (which should also be verified) that the boundaries of all diffusion regions exhibit similar behaviors, and Eq. (5.5) to the right of the arrow is therefore applicable to an arbitrary region without invariant curves with the K_b left boundary.

Equation (5.5) yields

$$F_{\text{th}} \approx \frac{2D_2(1/3)}{D_b(K_{\max})} \approx 13. \tag{5.6}$$

Here, coefficient 2 characterizes interference of two region boundaries, which amplifies diffusion suppression; $K_b \approx 0.23$ and $K_{\max} \approx 0.24$ are the left boundary and the center of the region under consideration, and $D_b(K_{\max}) \approx 5 \times 10^{-4}$ is the rate of diffusion in the center obtained from Eq. (5.5) for the boundary. Lastly, $D_2(1/3) \approx 3.28 \times 10^{-3}$ is the diffusion rate at the boundary of the principal region calculated by (5.1) on the assumption of the absence of invariant curves (we ignore the small $K_{\max} - K_b$ correction and use $K = 1/3$).

For the model under consideration, the measured (5.4) and theoretical (5.6) values very closely agree with each other, which substantiates the suggestion made above that diffusion region boundaries are similar. By analogy with (4.1), we may introduce an empirical correction factor $G_F = F_{\text{ex}}/F_{\text{th}} \approx 1.1$, which is close to one. Note that approximation (5.5) is only valid for $d = 1/2$ (see below), and applying it to narrow diffusion regions appears to be justified only in the vicinity of the principal boundary.

We stress once more that dependence (5.3) for diffusion rate $D_c(K)$ close to one invariant curve and dependence (5.5) for the $D_B(K)$ boundary of the principal region are similar; not only both are power dependences but also their exponents differ insignificantly. This difference is likely to arise because (5.3) is calculated at $K = 1/8$ and (5.5), at $K = 1/3$.

Several other narrower diffusion regions are also shown in Fig. 2. They are reproduced in Fig. 3 on an enlarged scale. The empirical values are connected by a polygonal line, which helps us to distinguish between closely spaced narrow regions. In the centers of these regions without invariant curves, diffusion is also suppressed the stronger, the narrower the region. For 17 regions that we were able to discern, calculations by (5.4)–(5.6), similar to those made above, were performed. These data are summarized in the table, which also contains the $C_F = F_{\text{ex}}/F_{\text{th}}$ empirical correction factors. Given in the last column are region widths calculated by (5.5) under the additional requirement to provide fulfillment of the equality $C_F \equiv 1$. For the widths, agreement with empirical estimates is much worse, which appears to be caused by difficulties of determining the width of a narrow ΔK region based on a limited number of perturbation parameter K values used in the calculations (see below).

Diffusion suppression at several bias parameter d values is shown in Fig. 4. Similarity of $D(K)$ dependences at different d values attracts attention. This similarity gives promise that a unified description of diffusion in some dimensionless variables can be constructed.

On the assumption that $K_D \approx K_B$ (see above), it is natural to write dimensionless perturbation parameter K^* as

$$K^* = \frac{K}{K_B} = \frac{1+d}{2d^2}K. \qquad (5.7)$$

The determination of the second dimensionless variable D^* involves serious difficulties because, generally, $|f_n| \neq \text{const}$ (see Section 3) and an explicit expression for the $D(K)$ function cannot therefore be obtained. This function can, however, be described approximately if the argument of the sine function in (2.4) is small. We then have $\sin(\pi nd) \approx \pi nd$ and [see (2.3)]

$$|f_n| \approx nf_0, \quad f_0 \approx \frac{2}{\pi}, \quad \beta = 1. \qquad (5.8)$$

Clearly, the $D(K)$ dependence at $d \ll 1$ is the same as in the $d = 0$ limit.

This approximation [and, therefore, asymptotic equation (4.1)] is valid at [see (3.5) with $\beta = 1$]

$$K \gtrsim K_1 = \frac{\pi^4 d^2}{16} \approx \frac{\pi^4}{32}K_B \approx 3K_B. \qquad (5.9)$$

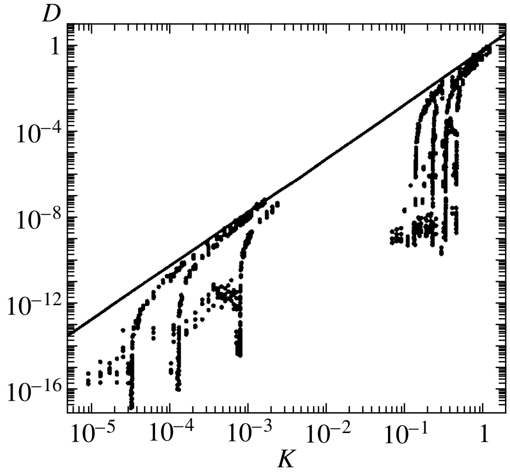


Fig. 4. General picture of diffusion in model (2.2) at seven bias parameter d values, $d = 0.004, 0.008, 0.02, 0.3, 0.4, 0.5$, and 0.6 (from left to right). The number of trajectories 50–250, computation time $t_0 = 2.5 \times 10^6$ iterations. The straight line is the upper bound of diffusion rate (4.1); also see text and Fig. 5.

On the other hand, if the inverse inequality holds ($\pi n_c d \gtrsim 1$, $d \ll 1$, $\beta = 2$), $\sin(\pi nd)$ in (2.4) can be averaged over n. As the sum in (3.2) satisfies the proportionality relation

$$\sqrt{f_n} \propto \sqrt{|\sin(\pi nd)|},$$

we can introduce the new parameter

$$S_0 = \langle\sqrt{|\sin(\pi nd)|}\rangle \approx 0.76 \qquad (5.10)$$

and set ($d \ll 1$)

$$f_0 \approx \frac{2}{\pi^2}\frac{S_0^2}{d}, \quad \beta = 2. \qquad (5.11)$$

The averaging condition can also be written as

$$K \lesssim K_2 \approx \left(\frac{\pi^2 d}{8S_0}\right)^2 \approx 2.6d^2 \approx 1.3K_B < K_1. \qquad (5.12)$$

We then have

$$D(K) \approx A_2K^5, \quad A_2 \approx \frac{A_1}{K_{\text{cr}}^{2.5}}, \qquad (5.13)$$

where the A_1 value is taken from (4.1), and K_{cr} is the intersection point between two asymptotic dependences, (5.13) and (4.1). The position of this point is not known, and it is not clear how it can be determined, because, at $d \ll 1$ (as distinguished from the case of $d = 1/2$ in Fig. 2), there is no diffusion region with $\beta = 2$. At the same time, both K_1 and K_2 boundaries are fairly close to K_B. For this reason, the abscissa of the intersec-

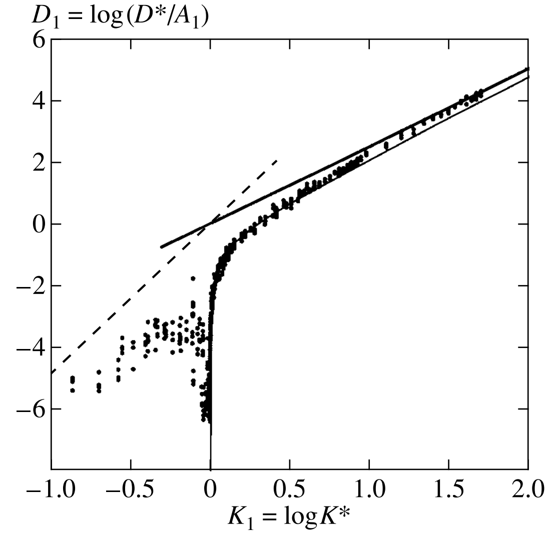


Fig. 5. General picture of diffusion in model (2.2) for the three smallest d parameter values, d = 0.004, 0.008, and 0.02, in dimensionless variables (5.7) and (5.15). The curve is the empirical approximation of the principal boundary of diffusion (5.16). The straight lines describe asymptotic behaviors of diffusion without invariant curves: the solid line corresponds to (4.1) with $\beta = 1$, and the dashed line, to (5.13) with $\beta = 2$ in approximation (5.14).

tion point between the two asymptotic dependences can be assumed to be

$$K_{\rm cr} \approx K_B. \tag{5.14}$$

In this approximation and taking into account (5.7) and (5.13), the second dimensionless variable can be selected in the form

$$D^* = \left(\frac{1+d}{2d^2}\right)^{5/2} D. \tag{5.15}$$

Strictly, variables (5.7) and (5.15) are suitable only if $d \ll 1$, when simple asymptotic dependences (4.1) and (5.13) can be used. The general similarity picture, however, persists to $d \sim 1$, but not for $d \longrightarrow 1$.

The results of our numerical experiments for $d \ll 1$ are shown in dimensionless variables in Fig. 5 together with two asymptotic dependences (4.1) and (5.13). The smooth curve, which can with difficulty be traced in the dense system of points, corresponds to the purely empirical universal boundary of the principal diffusion region found by us,

$$\begin{gathered} D_l = \frac{5}{2}K_l - \frac{0.4}{\sqrt{K_l}}, \\ D_l = \log\left(\frac{D^*}{A_1}\right), \quad K_l = \log K^*, \end{gathered} \tag{5.16}$$

which is written in the decimal logarithms of dimensionless variables.

This equation contains dimensionless variables (4.5) and (5.15), in which the suggestion of the equality of the K_B critical number and the K_D boundary of the principal diffusion region is implicit. Preliminary experiments (see Figs. 3–5) confirm this suggestion to high accuracy,

$$\frac{K_D - K_B}{K_B} \lesssim 10^{-3}.$$

The interesting question of exact equality of these two parameters, however, remains open.

In [9], only one region without invariant curves was found (see above), and the suggestion was made that the number of such regions grew infinitely as $K \longrightarrow 0$. Our results (Fig. 3) show that a set of such regions also exists in the immediate vicinity of the principal diffusion boundary. An important problem is the statistics of such regions related to the distribution of diffusion rates with respect to perturbation parameter K and, accordingly, to its suppression by virtual invariant curves.

We performed a statistical treatment of the experimental diffusion suppression coefficients $F_{\rm ex}$ for N = 134 parameter K values in the interval $0.293 < K < 0.333$ of width as small as $\Delta \approx 0.04$ (see Fig. 3). An effective method for obtaining a statistically significant distribution $P(F_{\rm ex})$ for such poor statistics is the special method for constructing an integral distribution with a “floating” cell width (see [18] and the references therein). This method is also called “rank-ordering statistics of extreme events.” This effective procedure was for the first time suggested in 1949 and used in mathematical linguistics [19]. It turns out that it suffices to arrange all $F_{\rm ex}(n)$, $n = 1, 2, \ldots, N$ values in decreasing order, $F_{\rm ex}(n+1) < F_{\rm ex}(n)$. The sought distribution is then given by the approximate equality

$$P(F_{\rm ex}) \approx \frac{n}{N}, \quad n = 1, 2, \ldots, N. \tag{5.17}$$

The distribution obtained in this simple way is shown in Fig. 6. Its most interesting feature is an exceedingly slow decrease in the probability of strong diffusion suppression,

$$\begin{gathered} P(F_{\rm ex}) \approx \frac{2}{F_{\rm ex}^{0.15}}, \\ 100 < F_{\rm ex} < 3 \times 10^5 = F_{\max}(t_0). \end{gathered} \tag{5.18}$$

Here, the left boundary is related to the very narrow interval of K values used in the calculations, $\Delta \approx 0.04$. The rapid decrease in probability $P(F_{\rm ex})$ at $F_{\rm ex} > 3 \times 10^5$ is explained by the limited time of calculations ($t_0 = 4 \times 10^7$). Indeed, as regular oscillations $\Delta p \sim K \approx 0.3$, the minimum observed diffusion rate is given by

$$D_{\min} \sim \frac{(\Delta p)^2}{t_0} \sim 10^{-9}, \tag{5.19}$$

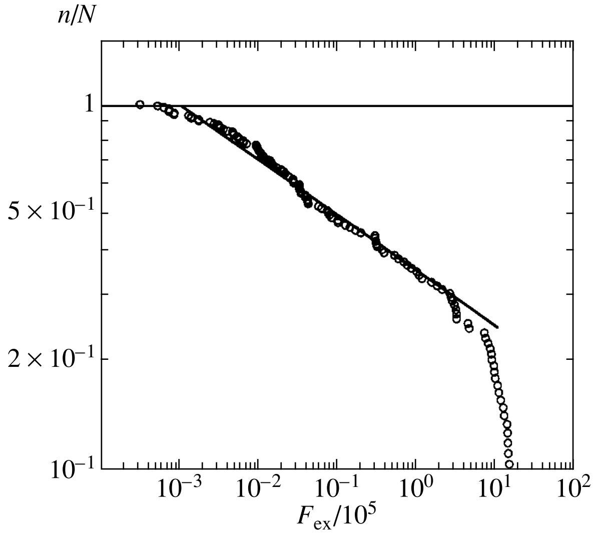


Fig. 6. The first empirical results on the statistic of the F_{ex} factor (5.4) of diffusion suppression by virtual invariant curves according to our numerical experiments shown in Fig. 3 ($d = 1/2$). The slanted straight line corresponds to integral power distribution (5.18). The total number of F_{ex} values is 134; 100 of them lie within the principal interval $F_{ex} < 3 \times 10^5$ (see text).

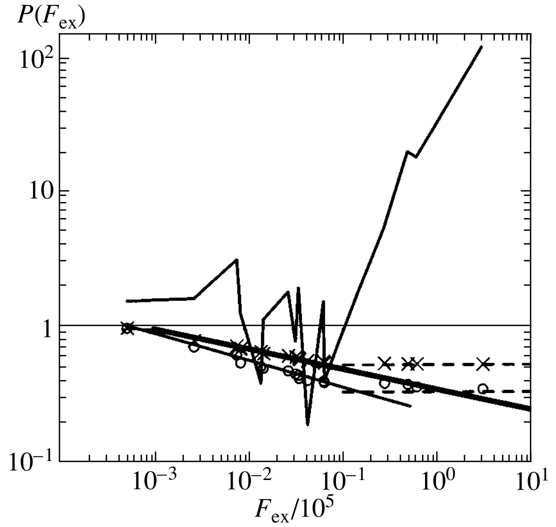


Fig. 7. The same as in Fig. 6 but over the width of 17 discernible diffusion regions (see table and text). Thick line corresponds to the data from Fig. 6 (5.18), circles are integral probabilities (5.21) determined from the experimental ΔK width of diffusion regions, crosses are integral probabilities for the δK width recalculated by (5.5), polygonal line is the experimental C_F correction to (5.6), and dashed horizontal lines are the fractions of "lost" (indiscernible) regions at large $F \gtrsim 10^4$.

which closely agrees with the experimental data shown in Fig. 3. At smaller K values, this minimum decreases even to $D \approx 10^{-17}$ at $K \approx 3 \times 10^{-5}$, $t_0 = 4 \times 10^8$ (see Fig. 4). Interestingly, at such a slow decrease in probability (5.18), both the mean F_{ex} value and its variance are determined by the F_{max} value and diverge approximately proportionally to $t_0 \longrightarrow \infty$ [see (5.19)].

Small deviations of the empirical distribution shown in Fig. 6 from law (5.18) are likely to be related to nonuniformities of the K_n values used in the calculations. This may easily be corrected, but will require considerable computation time or a substantial statistical reduction.

Empirical law (5.18) can, in particular, be used to analyze enigmatic trajectory "jamming" observed in old work [7], which still remains unexplained. This phenomenon is also related to map (2.2) with $d = 1/2$, but in a somewhat different normalization, as in [9]. In the notation that we use, it corresponds to $K = 0.29 < 1/3$ and, therefore, it fortuitously falls into the region with invariant curves. At a $t_0 = 3 \times 10^6$ time of computations, the minimum diffusion coefficient is $D_{min} \sim 3 \times 10^{-8}$, which corresponds to the minimum diffusion suppression factor $F_{min} \sim 10^5$ with a reasonable probability $P \approx 36\%$.

Instead of separate $F_{ex}(n)$ values, we can take the F_m values [here, the notation is simplified: F_m stands for $F_{ex}(m)$] for the centers of all $m = 1, 2, \ldots, M$ ($M = 17$) discernible in the $K = 0.293$–0.333 selected interval of diffusion regions (see table). The probability is then proportional to the normalized sum of the widths of these regions dK_m, which are found either empirically from Fig. 3 (ΔK in the table) or by recalculation according to (5.5) from the empirical suppression factor F_m value (δK in the table). The normalization is performed with respect to the total width of the K interval as $dK_m \longrightarrow dK_m/\Delta$, where $\Delta = 0.04$. In addition, it should be taken into account that the normalized sum of the widths of all $M = 17$ diffusion regions is $S \approx 0.664$ for ΔK and $S \approx 0.468$ for δK. In both cases, significant loss of narrow regions with large $F \gtrsim 10^4$ occurs (see Fig. 7). We eventually obtain

$$P(F_m) \approx 1 - S + \sum_{i=m}^{M} dK_i. \qquad (5.20)$$

The result of such a processing of the empirical data summarized in the table is shown in Fig. 7 together with the data from Fig. 6, which are plotted by a thick line corresponding to (5.18). Agreement for such limited statistics can be considered satisfactory. This especially refers to the δK data (within the $F \lesssim 10^4$ limitation introduced above). Note that outside this region, that is, at $F \gtrsim 10^4$, where the $P(F)$ probability is almost constant, we simultaneously observe a sharp increase in the C_F empirical correction (see Fig. 7).

Agreement is worse for ΔK, and we observe not only spread of data but also a systematic although small deviation. As previously, the empirical distribution remains a power function (see the lowest slanted straight line in Fig. 7) but with a somewhat different exponent,

$$P(F) \approx \frac{2.3}{F^{0.20}}, \quad 64 < F \lesssim 10^4, \qquad (5.21)$$

where the left boundary is determined by the minimum F value in the studied K range. The reason for these discrepancies is not known and requires further inquiries. We suggest that the observed discrepancies are largely caused by underestimation of the empirical width of the ΔK region, which increases as F grows. This in all probability occurs as a result of overlapping of neighboring regions, which also increases with F. The difference between empirical equations (5.18) and (5.21) amounts to about 30% in the exponent and 20% in the probability [within the range of the applicability of (5.21)], which is not bad for preliminary results.

Note that diffusion rate fluctuations mentioned above (not exceeding twofold rate changes) change probability (5.18) by a factor of $2^{0.15} \approx 1.11$; that is, by as little as 11%.

6. CONCLUSION

Studies of a family of piecewise linear maps of types (2.1) and (2.2) have a long history (e.g., see [7, 9, 11–15, 17] and the references therein). In this work, we use very simple models to study a comparatively new and little-known but very complex phenomenon of fractal diffusion under the action of virtual global invariant curves and under the conditions of strong local chaos.

However, first, it would be useful to understand why studies of such simplified constructions as piecewise linear maps deserve attention. Let us return to work [17], where a complex analytic function with 21 Fourier harmonics was used to study the properties of such a map, and a certain similarity between the dynamic behaviors of this function and the map was observed. We can therefore use the opposite approach and, for a complex continuous analytic function, for instance, for a function with sharp turns, seek a piecewise linear function close to the analytic function and study the corresponding map, which is much simpler. This approach, we believe, offers much promise, but requires special consideration.

Our studies show that, in the family of models (2.1), (2.2), there always exists a comparatively wide (principal) region of "normal" diffusion, as in other smooth systems without invariant curves. We were able to obtain fairly simple and fairly accurate diffusion rate estimates in this region (Sections 3 and 4), which were of considerable help in analyzing the most important empirical data on fractal diffusion in the region with virtual invariant curves (Section 5).

Our studies were performed in the range of bias parameter values $0 < d < 0.6$ [with invariant curves at $K < K_B$, see (4.5)]. We found that the presence of a set of invariant curves, although of measure zero, caused strong diffusion suppression at $K < K_B$, which prompted us to suggest a new term, namely, "virtual invariant curve."

The case of $d = 1/2$ was studied most thoroughly. For this case, we were able to obtain quantitative estimates of diffusion suppression. Preliminary treatment of empirical data allowed us to construct integral distribution (5.18) for the probability of the $F_{\rm ex}$ diffusion suppression coefficient. The most interesting feature of this distribution was an exceedingly slow decrease in the probability of $F_{\rm ex}$, slow to the extent that the mean $F_{\rm ex}$ value diverged as the time of computations increased (see Section 5). We also thoroughly studied a series of narrow diffusion regions to the left of the principal boundary (see table). This also allowed us to substantiate important result (5.18) by another method [see (5.21)]. The reason for such unusual statistics of $F_{\rm ex}$ [and, therefore, of the rate of fractal diffusion $D_{\rm max}$ in (5.4)] is not known and requires further inquiries.

Studies of the $D(K)$ dependences at different d values revealed obvious similarity of their behavior, which was an indication that their universal description might be possible. We were able to give such a description on the additional assumption that $d \ll 1$ and construct empirical dependence (5.16) for the boundary of the principal diffusion region in dimensionless variables. This simple dependence fairly well described the characteristic sharp transition from the chaotic region without invariant curves with well-known regular diffusion to an also chaotic region but with a dense system of invariant curves of a new form with absolutely unknown very irregular (fractal) diffusion.

On the whole, we consider this new phenomenon fairly interesting and important; in our view, it deserves further investigation.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 01-02-16836) and by the Russian Academy of Sciences ("Nonlinear dynamics and solitons" Program).

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer-Verlag, New York, 1992, 2nd ed.).
3. G. M. Zaslavsky and R. Z. Sagdeev, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988; Harwood, Chur, 1988).
4. J. Green, J. Math. Phys. **20**, 1183 (1979).
5. B. V. Chirikov, Chaos, Solitons and Fractals **1**, 79 (1991).
6. J. Moser, *Stable and Random Motion in Dynamical Systems* (Princeton Univ. Press, Princeton, 1973).

7. B. V. Chirikov, E. Keil, and A. Sessler, J. Stat. Phys. **3**, 307 (1971).
8. M. Hénon and J. Wisdom, Physica D (Amsterdam) **8**, 157 (1983).
9. S. Bullett, Commun. Math. Phys. **107**, 241 (1986).
10. M. Wojtkowski, Commun. Math. Phys. **80**, 453 (1981); Ergodic Theory Dyn. Syst. **2**, 525 (1982).
11. L. V. Ovsyannikov, private communication (May, 1999).
12. V. V. Vecheslavov, Preprint IYaF 99-69 (Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 1999).
13. V. V. Vecheslavov, Preprint IYaF 2000-27 (Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000); nlin.CD/0005048.
14. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **119**, 853 (2001) [JETP **92**, 744 (2001)].
15. V. V. Vecheslavov and B. V. Chirikov, Preprint IYaF 2000-68 (Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000); Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
16. L. V. Ovsyannikov, Dokl. Akad. Nauk **369**, 743 (1999).
17. I. Dana, N. W. Murray, and I. C. Percival, Phys. Rev. Lett. **62**, 233 (1989).
18. D. Sornette, L. Knopoff, Y. Kagan, and C. Vanneste, J. Geophys. Res. **101**, 13883 (1996).
19. G. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, 1949).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 166–174.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 188–197.
Original Russian Text Copyright © 2002 by Guz.

MISCELLANEOUS

Browninan Motion in a Fluctuating Periodic Potential

S. A. Guz
Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
e-mail: guz@lit.ru

Received November 23, 2001

Abstract—A classical Brownian particle is considered in a periodic potential field with a rapidly oscillating phase. The concept of effective potential is used for describing a slow averaged motion of a particle. It is shown that there exists a certain region in which a particle performs a stationary random motion without appreciable drift. By analogy with the ideal case, this region can be called an effective locking region. The situation described is valid for stationary fluctuations of the phase of a potential function, provided that they have a sufficiently small but finite correlation time. The study of the problem is reduced to the analysis of a stochastic system with external noise whose spectral density is zero at zero frequency ("green" noise [1]). The analysis of the first- and second-approximation equations of the averaging method exhibits the high stability of the locking phenomenon. This result has been verified by the numerical solution of appropriate stochastic equations. In this case, a predictor–corrector algorithm was used that allowed one to carry out a numerical simulation to a sufficiently high degree of accuracy. The result of the simulation is in good agreement with the theoretical results. The effective locking bandwidth calculated analytically by the averaging method actually coincides with the value obtained by the simulation. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The investigation of many nonlinear systems that can be described by ordinary stochastic differential equations usually involves the Markov approximation. If external noises are Gaussian white noises, then the problem is reduced to the Fokker–Planck diffusion equation (see, for example, [2–6]). From the viewpoint of statistical physics, this approach is quite justified since the diffusion coefficients can be determined by the well-known fluctuation–dissipation theorem [7]. For example, if we consider a Brownian particle, this theorem provides a relation between the diffusion coefficient and the coefficient of viscous friction, because diffusion and viscosity are attributed to the same phenomenon, the collision of a Brownian particle with a large number of particles of the surrounding medium—a statistical ensemble. Since the number of such particles is large and the time interval between collisions is small, the noise affecting the Brownian motion can be considered as Gaussian white noise with zero correlation time. The study of the phenomena that occur within physically small but finite time intervals is primarily aimed at determining the intensity of the white noise. In this approach, the particle trajectory is described by a Wiener process [2, 3, 8, 9], the concept that underlies stochastic calculus [10–12].

Suppose, however, that, in addition to weak white noise associated with the interaction of the particle with a statistical reservoir, it undergoes external noise that does not contain a constant component in its spectrum. For example, this may be a noisy acoustic or electromagnetic radiation transmitted through a finite aperture of an antenna or a lens. Radiation with the wavelength greater than the size of the aperture will be cut out from the spectrum of the external noise. Hence, the noise $\zeta(t)$ that actually affects the particle has the spectral density $S(\omega)$ equal to zero at zero frequency, i.e., $S(0) = 0$. On the other hand, the diffusion coefficient D in the Fokker–Planck equation is determined as the ratio of the noise increment variance in a certain sufficiently small time interval to the length of this interval. However, to allow for the diffusion approximation, this interval must be greater than the correlation time of noise because otherwise one cannot apply the white noise model. In this case, the diffusion coefficient is determined in terms of the noise correlation function $\psi(t_1 - t_2) = \langle\zeta(t_1)\zeta(t_2)\rangle$ (everywhere below, we assume that $\langle\zeta\rangle = 0$) by the relation [2, 3]

$$D = \int_{-\infty}^{\infty} \psi(\tau)d\tau = 2\pi S(0). \tag{1}$$

This formula takes into account that, according to the Wiener–Khinchin theorem, the functions $\psi(\tau)$ and $S(\omega)$ are related through the Fourier transformation [3, 5]. It follows from (1) that $D = 0$ if $S(0) = 0$. A good example of noise with zero diffusion coefficient is given by blackbody radiation. The correlation function for this case is given in [5]. The condition $D = 0$ implies that, under the effect of external noise with such spectral density, the motion of a free (i.e., when potential forces are neglected) Brownian particle is a stationary process, rather than a process with uncorrelated increments, whose variance grows by a diffusion law [5]. This fact provides sufficient grounds for selecting random stationary processes with $S(0) = 0$ into a separate class. By analogy with optics, noise is often characterized by its "color." A classical example is white noise, where $S(\omega) = \text{const}$. When $S(\omega) \neq \text{const}$, the term "colored noise" has recently found wide application [13–15]. A simple and popular example of colored noise is the Ornstein–Uhlenbeck process [12, 13, 15]. One often considers harmonic (or quasi-monochromatic) noise and broadband colored noise (see, for example, [16, 17]). Stationary noise with $S(0) = 0$ should also be regarded as colored noise. However, now its spectral density decreases at low frequencies, i.e., in the region of red color. Therefore, we call such noise green noise. In a number of our works [1, 18–23], as well as in [17, 24], it was shown that green noise causes a qualitative change in the dynamics of a stochastic system. The main subject of the present work is the study of trapping a Brownian particle in a periodic potential with a fluctuating phase. Below, we will show that this problem leads to green noise.

The main property of green noise is the existence of considerable negative autocorrelation. Only in this case, the integral of the correlation function may vanish. One should expect that such an aftereffect will result in an appreciable stabilization of a stochastic system. The system actually becomes quasi-stationary, and a weak diffusion process is only possible due to the nonlinearity of the system, when the zeroth harmonic of noise is detected during nonlinear transformations of the external green noise in the system itself. Accordingly, in [1, 19, 20], this phenomenon was called a nonlinear diffusion.

From the fact that the diffusion coefficient (1) vanishes, one cannot draw a general conclusion that the Markov approximation is inapplicable. In many cases, we can apply a multidimensional nonstationary Fokker–Planck equation; however, in this case, the number of variables increases, and the solution of this equation grows into a separate, quite serious, problem [4]. It was shown in the above-listed works [1, 18–23] that it is much easier to investigate the role of green noise by the Krylov–Bogolyubov averaging method [25]. It is important that this method allows one to consider the role of non-Gaussian external noise virtually without changing the scheme of calculation, which can hardly be done by generalizing the Fokker–Planck equation.

Note that the averaging method is widely used in the theory of nonlinear oscillations. As for stochastic oscillatory systems, this method is frequently applied to the investigation of high-Q self-oscillatory devices. In this case, a narrowband filter (an oscillatory circuit or a resonator) actually performs the time averaging of external broadband noise. Apparently, this was done for the first time in [26], where the averaging method was applied to investigating the fluctuations of a tube oscillator. Since the correlation time of the external noise and the period of oscillations are small as compared with the time constant of the filter, one can carry out the averaging of the main equation over the period and obtain equations for slowly varying amplitude and phase. The fluctuation sources in these equations are represented by white noise with the spectral density equal to that of the external noise at the oscillator frequency (see also [2, 3]). Nevertheless, it is noteworthy that, even in this case, the nonlinearity and the external noise were assumed to be small (on the order of a small parameter), so that a trivial (zeroth-order) approximation has the form of a simple harmonic oscillation. Such an approach agrees with the Krylov–Bogolyubov theory. In this relation, we note that a fundamental theorem was proved in [27] on normal deviations from the averaged system, which actually provides mathematical grounds for the procedure of analytic calculations described below. The only difference from the scheme given in [27] lies in the fact that, as the zeroth approximation, we use a solution to the equation of motion of a free particle under the action of additive external noise. In this case, one does not need to assume that the noise is small. However, if this noise is white, which is frequently assumed, then the zeroth approximation, obtained for the zero value of the small parameter, is a diffusion process, i.e., will initially contain a certain "slow" motion. At the same time, as we have already mentioned, a simple harmonic oscillation in oscillation theory, which can be interpreted, in a sense, as a stationary function, usually serves as an approximation describing a "fast" motion. The characteristic time of the fast motion is the period of these oscillations. This contradiction is removed if we consider the effect of green, rather than white, noise on a free system. Then, there exist sufficient conditions under which the motion of a free particle represents a stationary process [1, 23], and this process can be used as the zeroth approximation. Obviously, the averaging in this case should be carried out over time intervals greater than the correlation time of the zeroth approximation thus obtained.

Note also that the linearization of the potential field at a certain point is a convenient and frequently used method to obtain a stationary diffusion approximation. However, such an approach does not allow one to describe the free motion of a system. An example is given by an ordinary parabolic potential, when a particle cannot leave a certain closed region unless we introduce the concept of absorbing walls, which always looks somewhat artificial. On the other hand, if we restrict ourselves to the first approximation of the averaging method, which takes into account potential forces, it suffices to know one- and two-dimensional probability densities to obtain information on the nonstationary motion of the system and thus to solve the problem. The study of every subsequent approximation requires the calculation of all the preceding approximations; however, as a rule, the asymptotic series of the averaging method converge well under an appropriate choice of the expansion parameter, and the first approximation usually proves to be sufficient. Nevertheless, we have to point out the basic drawback of the averaging method: it does not give the distribution functions

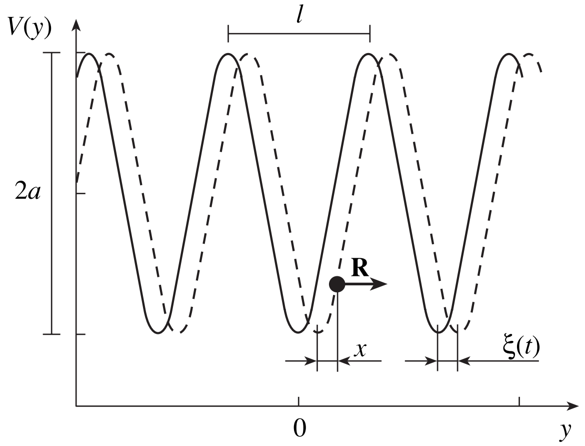


Fig. 1. Brownian particle in a periodic potential with a fluctuating phase.

of the variables of the system, but only their moments. However, this fact is not essential for many applications.

2. MODEL

Below, we consider a one-dimensional motion of a Brownian particle in a certain periodic potential $V(y)$ (Fig. 1) whose phase is perturbed by a stationary field $\xi(t, y)$ with zero mean $\langle\xi\rangle = 0$, variance σ^2, and the characteristic wavelength much greater than the spatial period l of the unperturbed potential. An example is given by the effect of an acoustic field on a periodic structure of a crystal or a semiconductor of sufficiently small dimension. Choose l as the length scale and assume that the coordinate y is dimensionless. Represent the potential function as $V(y) = aU(y)$, where $U(y)$ is a dimensionless function defining the shape of the potential and a is a certain amplitude. Then, the equation of motion of a particle in the laboratory system of coordinates can be expressed as

$$m\ddot{y} + k\dot{y} = \frac{\alpha}{l}F + \frac{a}{l}f(y + \xi), \tag{2}$$

where the dot denotes time differentiation, m is the mass of the particle, k is the coefficient of viscous friction, and $R = aF/l$ is a certain constant external force acting on the particle (Fig. 1). As we have already mentioned above, we neglect weak white noise and apply a long-wavelength approximation:

$$f(y+\xi) = -\frac{\partial U(y+\xi)}{\partial y}$$

$$= -\left(1 + \frac{\partial\xi}{\partial y}\right)\left[\frac{\partial U(s)}{\partial s}\right]_{s = y+\xi} \approx -\left[\frac{\partial U(s)}{\partial s}\right]_{s = y+\xi}.$$

Assume that the function $f(y)$ is bounded and periodic with zero constant component. To analyze the motion of the particle with respect to a fixed potential, we pass to the system of coordinates connected with this potential by the change of variables

$$x = y + \xi. \tag{3}$$

Then, Eq. (2) is reduced to the equation of motion that we used in [1, 20, 22]:

$$\vartheta\ddot{x} + \dot{x} = \varepsilon[F + f(x)] + \zeta(t), \tag{4}$$

where $\vartheta = m/k$, $\varepsilon = a/lk$, and

$$\zeta = \vartheta\ddot{\xi} + \dot{\xi} \tag{5}$$

is an external perturbation. The introduction of the parameter ϑ seems convenient because the case when $\vartheta = 0$ corresponds to the simpler and widely used model of an overdamped particle. Note that the form of noise (5) is similar to the noise arising in the phase-locking system of the second kind when there exists a random stationary modulation of the phase of the external signal [2, 20].

When $\varepsilon = 0$, Eq. (4) has a particular solution that coincides with $\xi(t)$ and, as we have already noted, represents a stationary process rather than a process with uncorrelated increments, as would be the case should the external noise $\zeta(t)$ be white. According to relation (5), the spectral density $S(\omega)$ of the random force $\zeta(t)$ in Eq. (4) is given by

$$S(\omega) = \omega^2[(\vartheta\omega)^2 + 1]G(\omega) \tag{6}$$

and vanishes at zero frequency if the spectral density $G(\omega)$ of the process $\xi(t)$ grows no faster than ω^2 as $\omega \longrightarrow 0$. In other words, the external noise $\zeta(t)$in this case is green noise. It should be noted that we do not consider the effect of the motion of the Brownian particle on the phase modulation itself of the periodic potential. Therefore, the green noise (5) is an external action that is specified *a priori* and does not depend on the state of the Brownian particle. For example, if we consider a crystalline lattice in solids, this means that we neglect the scattering of the Brownian particle by phonons. This situation is quite analogous to the Mössbauer effect. In mechanics, we can consider an example of a particle moving along a randomly vibrating sinuous-shaped surface of a bed. In this case, we have to assume that the mass of the bed is much greater than the mass of the particle and that we can neglect the reaction force exerted by the particle on the bed.

Henceforth, we use the parameter ε as a parameter of asymptotic expansion. Therefore, as we noted in the Introduction, the stationary process $\xi(t)$ represents the zeroth approximation of this theory. Note that, when $\vartheta \neq 0$ and $\omega \longrightarrow \infty$, the spectral density (6) diverges, which is due to the noninertial character of the frame of reference in which Eq. (4) is written [22].

In the present paper, we carry out specific analytic calculations for the case when the process $\xi(t)$ is the

Ornstein–Uhlenbeck process. We represent its spectral density G and the correlation function Ψ as

$$G(\omega) = \frac{D}{2\pi}\frac{1}{\omega^2+\gamma^2}, \qquad \Psi(\tau) = \frac{D}{2\gamma}\exp(-\gamma|\tau|). \tag{7}$$

Then, a complete system of initial equations can be written as

$$\vartheta\ddot{x} + \dot{x} = \varepsilon[F + f(x)] + \vartheta\ddot{\xi} + \dot{\xi}, \qquad \dot{\xi} = -\gamma\xi + N(t), \tag{8}$$

where $N(t)$ is white Gaussian noise with the correlation function $\langle N(t)N(t+\tau)\rangle = D\delta(\tau)$. When $\vartheta \neq 0$, the presence of the second-order derivative of the external noise on the right-hand side of the first equation in (8) substantially complicates the application of the Fokker–Planck equation (see [6]). However, the averaging method allows one to analyze this problem by sufficiently simple means.

3. AVERAGING METHOD

Let us return to Eq. (2) and rewrite it as

$$\hat{L}y = \left(\vartheta\frac{d^2}{dt^2} + \frac{d}{dt}\right)y(t) = \varepsilon[F + f(y+\xi)]. \tag{9}$$

The main idea of the averaging method is the time averaging of the right-hand side of this equation to extract a "slow" motion. Depending on the type of problem, there exist a variety of algorithms of the averaging method (see, for example [25–30]). In our works [1, 18–23], we primarily proceeded from the algorithms described in [27, 28], where the characteristic averaging time indefinitely increases in the limit. However, the substance of the method does not experience crucial changes if one fixes T and decreases the correlation time of the external noise. In both cases, it is essential that the ergodicity [3, 5, 27] of the random function $f(y + \xi)$ is used for fixed $y = c =$ const, $|c| < \infty$. This property is the main approximation of the theory, although it was shown in [27] that, in general, it is sufficient that the function $f(c + \xi)$ satisfies a strong mixing condition.

Consider an auxiliary differential equation

$$\hat{L}\bar{y} = \varepsilon[F + q(\bar{y})], \tag{10}$$

which is called, as in [30], a comparison equation. Here, $\bar{y}(t)$ is a deterministic (not random) function, and $\varepsilon q(\bar{y}) = \varepsilon\langle f(\bar{y} + \xi)\rangle$ is a certain deterministic force that is assumed to be a smooth function in what follows. Then, it is convenient to define this force with the use of the effective, or stochastic, potential

$$U_{\text{eff}}(\bar{y}) = \langle U(\bar{y}+\xi)\rangle \tag{11}$$

by the conventional rule

$$q(\bar{y}) = -\frac{\partial U_{\text{eff}}(\bar{y})}{\partial \bar{y}}.$$

The possibility of introducing an effective potential is the essential factor in this theory.

The function $\bar{y}(t)$ is called the averaged motion of the system. It is important that, to determine this function, we applied the statistical averaging of the right-hand side of Eq. (9) rather than the time averaging. The accuracy of such a substitution depends on the actual finite interval of the averaging time. In general, this interval cannot be greater than the minimal nonzero time $T_{\min}$ of all characteristic times of transient processes defined by the general solution to Eq. (10). Since the random process $f(c + \xi)$ is ergodic, for $\tau_0 \ll T_{\min}$, where τ_0 is the correlation time of the process $\xi(t)$, the accuracy of the substitution of the averaging over an ensemble for the time averaging is of order $\tau_0/T_{\min}$. Hence, it is this ratio that serves as a measure of the actual dimensionless small parameter $\mu \sim \tau_0/T_{\min}$ in this theory. Often, this condition is denoted as follows. One introduces a "slow" time $t' = \mu t$; then, the parameter μ becomes a parameter of asymptotic expansions [5, 25]. However, to give a clearer physical interpretation of the results, we do not introduce two times: "fast" and "slow" [29]. In the problem considered here, the linearized equation (4) with $\zeta = 0$ and $F = 0$ near the stable points entails a characteristic equation of the form $\vartheta\lambda^2 + \lambda \approx -\varepsilon$, whence we find that one has to take, as $T_{\min}$, the lesser of the two times

$$T_{1,2} \propto 2\vartheta/(1 \pm \sqrt{1-4\varepsilon\vartheta}). \tag{12}$$

To avoid a difficult question on the averaging time in the case of oscillatory relaxation of the system, below we assume that the inequality $4\varepsilon\vartheta \lesssim 1$ holds (the case of aperiodic relaxation).

Consider the difference $v(t) = y(t) - \bar{y}(t)$ between the exact solution to Eq. (9) and the averaged solution $\bar{y}(t)$ to Eq. (10). This difference satisfies the equation

$$\hat{L}v = \varepsilon[f(\bar{y} + v + \xi) - \langle f(\bar{y}+\xi)\rangle]. \tag{13}$$

The right-hand side of this equation can be represented as the Taylor series in powers of v; thus we obtain an equation for successive approximations. Therefore, it is convenient to specify the initial condition for $v(t)$ as $v(t_0) = 0$. Then, we obtain $x(t_0) = x_0 + \xi(t_0)$, where x_0 is a certain prescribed nonrandom variable. Moreover, if we set $\bar{y}(t_0) = x_0$, then the set of such initial conditions actually implies that, when $t < t_0$, the potential field is as if switched off, and the motion of a free Brownian particle in the reference frame connected with the potential is described by the random process $x(t) = x_0 + \xi(t)$.

Taking into account (3), we represent the required function $x(t)$ as

$$x(t) = y + \xi = \bar{y} + v - \langle v \rangle + \langle v \rangle + \xi$$
$$= \bar{x}(t, \bar{y}(t)) + \xi(t) + u(t, \bar{y}(t)),$$

where $\bar{x} = \bar{y} + \langle v \rangle$ describes the slow component of motion, while $u = v - \langle v \rangle$ together with ξ describes the fast component. For $t \geq t_0$, we represent the function $v(t)$ as the power series

$$v(t) = \sum_{m=1}^{\infty} \varepsilon^m v_m(t). \tag{14}$$

Then, with regard to the Taylor expansion, Eq. (13) is rewritten as

$$\sum_{m=1}^{\infty} \varepsilon^{m-1} \hat{L} v_m(t) = f(\bar{y} + \xi)$$
$$- \langle f(\bar{y} + \xi) \rangle + \sum_{m=1}^{\infty} \frac{1}{m!} \frac{d^m f(\bar{y} + \xi)}{d\bar{y}^m} v^m.$$

Substituting here (14) and collecting the terms of the same order in ε, we can obtain the equations of further approximations.

In particular, in the first approximation ($m = 1$), we have

$$\hat{L} v_1(t) = f(\bar{y} + \xi) - \langle f(\bar{y} + \xi) \rangle. \tag{15}$$

We can see that $\langle v_1 \rangle = 0$. Hence, $v_1 \equiv u_1$, and a correction for the averaged motion is equal to zero. The fast motion $u_1(t)$ is described by the equation

$$\hat{L} u_1(t, \bar{y}(t)) = f(\bar{y}(t) + \xi(t)) - \langle f(\bar{y}(t) + \xi(t)) \rangle. \tag{16}$$

In an important case when $y_0 = x_0 = \mathrm{const}$ is a stable point of Eq. (10), one may not consider transient processes defined by this equation, and the process $f(y_0 + \xi)$ will be stationary. In this case, Eq. (16) describes the above-described nonlinear diffusion if $t - t_0 \gg \tau_0$ and the spectral density $S_1(\omega)$ of the process on the right-hand side satisfies the condition $S_1(0) \neq 0$. By analogy with relation (1), the coefficient of this diffusion can be represented as

$$\Lambda = \varepsilon^2 \int_{-\infty}^{\infty} \Psi_1(\tau) d\tau, \tag{17}$$

where $\Psi_1(\tau)$ is the correlation function of the random process $f(y_0 + \xi)$. It is important that the diffusion time $1/\Lambda$ actually determines a time interval within which the use of the first-order approximation is most effective. If the time $t - t_0$ exceeds $1/\Lambda$, then the correction due to the nonlinear diffusion is of order one, and one should consider higher order approximations.

The equations of the second- and third-order approximations ($m = 2, 3$) are expressed as

$$\hat{L} v_2 = f'(\bar{y} + \xi) v_1,$$
$$\hat{L} v_3 = f'(\bar{y} + \xi) v_2 + \frac{1}{2} f''(\bar{y} + \xi) v_1^2, \tag{18}$$

where the prime denotes a partial derivative with respect to $\bar{y}$. Hence, the corrections $\langle v_2 \rangle$ and $\langle v_3 \rangle$ for the slow motion are generally different from zero. However, a specific calculation of these and higher order corrections depends on the form of the nonlinear function $f(y)$ and is usually complicated enough (see below).

4. HARMONIC POTENTIAL

Below, as an example of application of the theory presented, we consider the harmonic potential $U(x) = -\cos x$. In this case, Eq. (4) is rewritten as

$$\vartheta \ddot{x} + \dot{x} = \varepsilon(F - \sin x) + \zeta(t).$$

Equations of this type are frequently used to describe synchronization systems in the theory of oscillations [2], the theory of superconductivity (the Josephson effect [31]), etc. Indeed, when $|F| \leq 1$ and there is no noise ($\zeta = 0$), a Brownian particle will be fixed at one of the stable points $x_n = 2\pi n + \arcsin F$. This phenomenon is called locking, or synchronization (when dealing, for example, with the phases of a signal and of an oscillator in phase-locking systems or with the phases of the wave functions of Cooper pairs on different sides of the Josephson junction). The region $|F| \leq 1$ is called the locking (or synchronization) region. If $\vartheta = 0$, one can easily find out that the time-average velocity of a particle outside the locking region is expressed as $\langle \langle \dot{x} \rangle \rangle = \varepsilon \sqrt{F^2 - 1}$ [2, 31].

Since $\langle \xi \rangle = 0$, in the first-order approximation of the theory, the slow motion $\bar{x}(t)$ of a particle coincides with the averaged motion $\bar{y} = \bar{x} + o(\varepsilon)$ in the laboratory frame of reference and is described by the comparison equation (10):

$$\vartheta \ddot{\bar{x}} + \dot{\bar{x}} = \varepsilon \left[F - \frac{\partial U_{\mathrm{eff}}(\bar{x})}{\partial \bar{x}} \right]. \tag{19}$$

Hereupon, we assume that the noise $\xi(t)$ is an Ornstein–Uhlenbeck process. Therefore, the effective potential (11) has the form [3]

$$U_{\mathrm{eff}}(\bar{x}) = -\langle \cos(\bar{x} + \xi) \rangle = -\Omega_{\mathrm{eff}} \cos \bar{x},$$

where, according to (7),

$$\Omega_{\mathrm{eff}} = \exp\left(-\frac{\sigma^2}{2}\right) = \exp\left(-\frac{D}{4\gamma}\right). \tag{20}$$

Earlier [1], we called this quantity the effective locking half-bandwidth of the averaged motion of a Brownian particle. As we stressed in [1], the dependence of the effective locking region on the variance of the noise $\xi(t)$ is determined by the probability distribution density for this noise. In particular, in [1], we investigated the case of a uniform density of distribution.

Henceforth, we do not consider the properties of nonlinear diffusion, which we studied in detail in our earlier works [1, 20, 21, 23]. We only mention the following fact that is essential from the viewpoint of the physical picture: irrespective of the intensity (variance) of noise $\xi(t)$, the diffusion coefficient (17) is bounded and even begins to decrease as the latter increases. This is associated with the fact that, as the intensity of noise $\xi(t)$ increases, the effect of the nonlinearity of the system becomes weaker, and the motion of a Brownian particle tends to a stationary process $\xi(t)$ because one should get such a stationary motion of a free particle in the limit. Such a behavior of the diffusion coefficient is quite contrary to the case of white noise, when a decrease in the role of the nonlinearity leads to the situation where the motion of a particle tends to a nonstationary process with uncorrelated increments.

Let us dwell on the question of what the order is of the contribution of the second-order approximation to the slow motion. To simplify the calculations, consider the case of an overdamped particle ($\vartheta = 0$). Let $\bar{y} = y_0 =$ const be a certain stable point of the comparison equation (10) in which

$$q(y_0) = -\langle \sin(y_0 + \xi) \rangle = -\Omega_{\text{eff}} \sin y_0.$$

Then, according to (15) and (18), the equations of the first- and second-order approximation are given by

$$\begin{aligned} \dot{v}_1(t) &= f(\bar{y} + \xi) - \langle f(\bar{y} + \xi) \rangle \\ &= -\sin(y_0 + \xi) + \langle \sin(y_0 + \xi) \rangle, \end{aligned} \tag{21}$$

$$\dot{v}_2 = \frac{df(\bar{y} + \xi)}{d\bar{y}} v_1 = -\cos(y_0 + \xi) v_1. \tag{22}$$

Using a formula similar to (20) but for the sum or the difference of two Gaussian quantities $c_1 + \xi_1$ and $c_2 + \xi_2$, where c_1 and c_2 are arbitrary numbers, $\xi_1 = \xi(t)$, and $\xi_2 = \xi(t + \tau)$, we can obtain the following expression [3]:

$$\begin{aligned} &\langle \cos(c_1 \pm c_2 + \xi_1 \pm \xi_2) \rangle \\ &= \operatorname{Re}\left\{ \exp i(c_1 \pm c_2) \exp\left[-\frac{1}{2} \langle (\xi_1 \pm \xi_2)^2 \rangle \right] \right\} \\ &= \cos(c_1 \pm c_2) \exp[-(\sigma^2 \pm \langle \xi_1 \xi_2 \rangle)], \end{aligned}$$

whence we obtain

$$\begin{aligned} &\langle \sin[c_1 + \xi(t)] \sin[c_2 + \xi(t - \tau)] \rangle \\ &= \frac{1}{2} \cos(c_1 - c_2) \exp\{-[\sigma^2 - \Psi(\tau)]\} \\ &- \frac{1}{2} \cos(c_1 + c_2) \exp\{-[\sigma^2 + \Psi(\tau)]\}. \end{aligned} \tag{23}$$

Here, $\Psi(\tau) = \langle \xi_1 \xi_2 \rangle = \langle \xi(t) \xi(t + \tau) \rangle$ is the correlation function of the process $\xi(t)$. Let us integrate Eq. (21) from zero to t and substitute the result obtained into Eq. (22). Carrying out the statistical averaging of the latter equation and applying formula (23), we obtain the following equation for the second-order correction $\langle v_2 \rangle$ for the averaged motion:

$$\begin{aligned} \langle \dot{v}_2 \rangle &= -\frac{1}{2} \exp(-\sigma^2) \sin(2y_0) \\ &\times \left\{ \int_0^t [1 - \exp(-\Psi(\tau))] d\tau \right\}. \end{aligned}$$

In the case of the Ornstein–Uhlenbeck process, the integral on the right-hand side does not exceed the number

$$\begin{aligned} &\int_0^\infty \{1 - \exp[-\sigma^2 \exp(-\gamma\tau)]\} d\tau \\ &= \frac{\gamma_0 + 2\ln\sigma + E_1(\sigma^2)}{\gamma}, \end{aligned} \tag{24}$$

where $\gamma_0 \approx 0.577$ is the Euler constant and $E_1(z)$ is an integral exponential function [32]. The analysis of the growth rate of the correction $\langle v_2 \rangle$ for $t \gg \tau_0$, where $\tau_0 = 1/\gamma$ is the correlation time of the process $\xi(t)$, shows that this rate is proportional to $\tau_0 \sigma^2$ for a small variance ($\sigma^2 \ll 1$), whereas, for a large variance ($\sigma^2 \gg 1$), it decreases at least as $t_0 \exp(-\sigma^2)$. This growth attains its maximum at $\sigma^2 \approx 0.82$ and has an absolute value of about $0.15\tau_0 \sin(2y_0)$. This result shows that, for $\tau_0 \ll 1$, the first-order approximation of the theory provides a sufficiently accurate description of the particle motion. This fact is also confirmed by the numerical experiment described in the next section.

5. NUMERICAL SIMULATION

Apparently, the results we obtained in [1] were first confirmed in [33], where the author carried out a numerical experiment on the basis of the Heun scheme. Later on, in [23], we presented the result of a numerical experiment based on the predictor–corrector algorithm [34], which actually confirmed the existence of the effective locking region of a particle in a periodic potential with a deterministic phase. In both cases, the motion of an overdamped Brownian particle ($\vartheta = 0$)

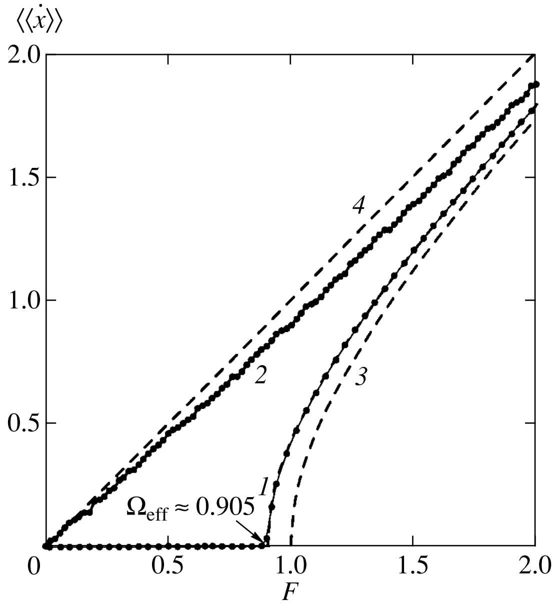


Fig. 2. The result of numerical calculation of the time-averaged velocity of an overdamped particle in the frame of reference connected with the potential curve as a function of an external constant force.

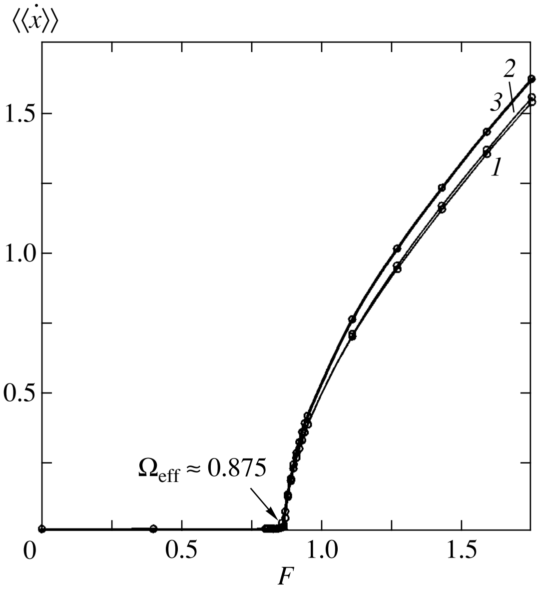


Fig. 3. The result of the numerical calculation of the time-averaged velocity of a particle with regard to the finiteness of its mass for the following values of the parameters: (*1*) $\tau_0 = 0.05$, $\vartheta \approx 0.17$, and $D \approx 10.7$; (*2*) $\tau_0 = 0.1$, $\vartheta = 0.25$, and $D \approx 5.34$; and (*3*) $\tau_0 = 0.1$, $\vartheta = 0.5$, and $D \approx 5.34$.

was considered under the action of external green noise defined as the time derivative of the Ornstein–Uhlenbeck process.

To verify the theoretical results presented above in the case of a periodic potential with a fluctuating phase, we numerically integrated the stochastic system of equations (8). As in [23], we applied the predictor–corrector algorithm of order $O(h^{3/2})$, where h is the time step of integration [34]. Time t in the calculations was measured in the units of $1/\varepsilon$. We calculated the time-averaged velocity of a particle $\langle\langle\dot{x}\rangle\rangle = (x_N - x_1)/Nh$ as a function of the external force F, where N is the number of steps. In the theory of synchronization, this function describes the mean difference between the frequencies of a signal and the voltage-controlled oscillator of a phase-locked system under the variation of the initial detuning between them [2]; in the theory of the Josephson effect, it describes the averaged current–voltage characteristic of the Josephson junction [31].

When $\vartheta = 0$, the algorithm has the form

$$x_{i+1} = x_i + gY_ih^{1/2} + (F - \sin x_i - \gamma\xi_i)h$$
$$- (\gamma + \cos x_i)gZ_ih^{3/2},$$

$$\xi_{i+1} = \xi_i + gY_ih^{1/2} - \gamma\xi_ih - \gamma gZ_ih^{3/2},$$

where Y_i and Z_i are two random normal independent numbers with the unit variance and $g^2 = D$. The correlation time of noise was equal to $\tau_0 = 1/\gamma = 0.1$; i.e., the actual small parameter of the system was $\mu \sim \varepsilon\tau_0 = 0.1$.

Below, we present the results of numerical calculations for $N = 2 \times 10^7$, $h = 0.0025$, and $D = 4$. The dots on curve *1* in Fig. 2 represent the results of a numerical experiment for the case of green noise. Curve *1* itself represents a theoretical result given by Eq. (19) with the effective locking bandwidth defined by (20), i.e., $\langle\dot{x}\rangle = 0$ for $|F| \le \Omega_{\rm eff}$ and $\langle\dot{x}\rangle = \sqrt{F^2 - \Omega_{\rm eff}^2}$. We can see that this theoretical result is in a good agreement with the result of the numerical experiment. Curve *2* corresponds to the case of white noise ($\gamma = 0$) and to the same sequence of numbers Y_i and Z_i. The dashed curve *3* represents the ideal case when there is no noise ($\Omega_{\rm eff} \equiv 1$). Finally, the dashed curve *4* corresponds to the absence of nonlinearity. We can see that the green noise significantly stabilizes the system in the region of $|F| \lesssim \Omega_{\rm eff}$.

In the case of $\vartheta \ne 0$, the algorithm becomes somewhat complicated:

$$x_{i+1} = x_i + gY_ih^{1/2} + (y_1 - \gamma\xi_i)h - \gamma gZ_ih^{3/2},$$

$$z_{i+1} = z_i + \Gamma(-z_i + F - \sin x_i)h - \Gamma gZ_ih^{3/2}\cos x_i,$$

$$\xi_{i+1} = \xi_i + gY_ih^{1/2} - \gamma\xi_ih - \gamma gZ_ih^{3/2},$$

where $z = \dot{x}$ and $\Gamma = 1/\vartheta$. Figure 3 represents the results of calculations by this algorithm. Curves *1* and *2* virtually coincide; the parameters for these curves are chosen so that $\Omega_{\rm eff}$ has the same value equal to 0.875, while the small parameter defined by (12) is $\mu \sim \tau_0/T_1 \approx 0.2$. In the case of curve *3*, the parameters of the numerical

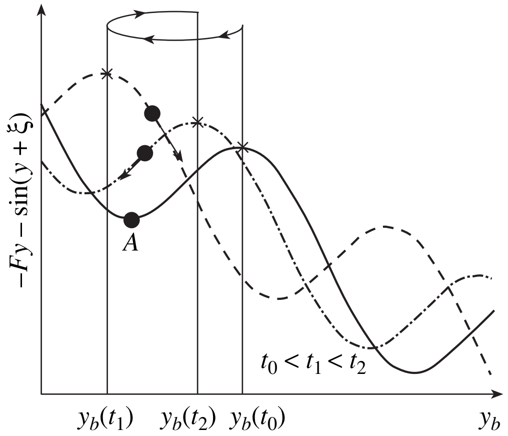


Fig. 4. The process of locking a Brownian particle near the stable point A under fluctuations of the phase of the potential. The cross denotes the position of the potential barrier at different moments of time.

experiment go beyond the region where the relaxation is aperiodic $(4\vartheta \lesssim 1)$; this fact has an immediate effect on the behavior of the curve: it is steeper than the first two curves.

All curves in Figs. 2 and 3 clearly display the effective locking region; the half-bandwidth of this region coincides with that obtained analytically. This fact suggests that the numerical simulation of the motion of a particle in a fluctuating potential fully confirms the theory based on the averaging method and the concept of the effective potential. Moreover, we can assert that, in the case of green noise, the concept of the locking, or synchronization, region in various stochastic systems virtually preserves its classical value.

6. CONCLUSION

The physical picture of the motion of a Brownian particle considered above can be explained as follows. Suppose that a particle is at a stable point x_0 of the periodic potential at the initial moment t_0. A random stationary displacement $\xi(t)$ of this potential curve with respect to the laboratory frame of reference will also result in a stationary random motion $x(t) + \xi(t)$ of the particle in the frame of reference of the potential. Due to the weakness of the nonlinear diffusion, such a motion of the particle will last for a long period of time. However, if the point x_0 is close to a certain value where the position of the particle is lower than the nearest maximum of the potential (barrier) by about Ω_{eff}, then the probability of overcoming this barrier due to the fluctuations of the random stationary motion of the particle becomes significant. Then, the particle will “fall over” the maximum and, in general, may leave the locking region. It is at this moment when the property of negative aftereffect of green noise plays an important role, because it can change the scenario and play the opposite role: make the particle return to the locking region even if it has passed the barrier for a moment. From the standpoint of the problem considered in this paper, this situation looks as follows. Suppose that, at a moment $t_1 > t_0$, the particle finds itself behind the barrier. However, by inertia, it may not leave the locking region if the fluctuation of the barrier quickly takes it back at a moment $t_2 > t_1$. This process is illustrated in Fig. 4, where y_b is the coordinate of the barrier in the laboratory frame of reference. If we assume that the noise $\xi(t)$ is white, then the increments of the particle coordinate with respect to the potential curve are independent; the displacements of the potential curve are accumulated according to a diffusion law; and the particle overcomes the barrier, going beyond the locking region virtually immediately after its coordinate reaches this maximum.

In conclusion, note that we investigated the effect of green noise on stochastic systems through an example of a Brownian motion in a randomly perturbed potential field. We obtained good agreement between the theory and a numerical experiment. The results obtained here and our works cited above are of a sufficiently general character and can be applied to the calculation of many radar and synchronization systems, laser systems, and systems of solid state physics, as well as the systems of other fields of science and technology where an important role is played by green-noise-type random processes.

REFERENCES

1. S. A. Guz and M. V. Sviridov, Phys. Lett. A **240**, 43 (1998).
2. V. I. Tikhonov and M. A. Mironov, *Markov Processes* (Sov. Radio, Moscow, 1977).
3. S. A. Akhmanov, Yu. E. D’yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981).
4. H. Risken, *The Fokker–Planck Equation* (Springer-Verlag, Berlin, 1989).
5. S. M. Rytov, *Introduction to Statistical Radiophysics* (Nauka, Moscow, 1976), Vol. 1.
6. S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Introduction to Statistical Radiophysics* (Nauka, Moscow, 1978), Vol. 2.
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1964; Pergamon, Oxford, 1980).
8. B. V. Gnedenko, *Theory of Probability* (Fizmatgiz, Moscow, 1961; Central Books, London, 1970).
9. V. I. Bogachev, *Gaussian Measures* (Nauka, Moscow, 1997).
10. I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes* (Nauka, Moscow, 1965; Saunders, Philadelphia, 1969).
11. R. Elliott, *Stochastic Calculus Applications* (Springer-Verlag, Berlin, 1983; Mir, Moscow, 1986).

12. A. N. Borodin and P. Salminen, *Handbook on Brownian Motion* (Birkhäuser Verlag, Basel, 1996; Lan', St. Petersburg, 2000).
13. W. Horsthemke and R. Lefever, *Noise Induced Transitions. Theory and Applications in Physics, Chemistry and Biology* (Springer-Verlag, Heidelberg, 1984; Mir, Moscow, 1987).
14. *Noise in Nonlinear Dynamical Systems,* Ed. by F. Moss and P. V. E. McClintock (Cambridge Univ. Press, Cambridge, 1989), Vols. 1–3.
15. J. Keiser, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Berlin, 1987; Mir, Moscow, 1990).
16. L. Schimansky-Geier and Ch. Zulicke, Z. Phys. B **79**, 451 (1990).
17. J. D. Bao and S. J. Liu, Phys. Rev. E **60**, 7572 (1999).
18. A. R. Arsen'ev, S. A. Guz, and M. V. Sviridov, Radiotekh. Élektron. (Moscow) **43**, 415 (1998).
19. S. A. Guz, Yu. G. Krasnikov, and M. V. Sviridov, Dokl. Akad. Nauk **365**, 34 (1999) [Dokl. Phys. **44**, 127 (1999)].
20. S. A. Guz, Radiotekh. Élektron. (Moscow) **45**, 444 (2000).
21. S. A. Guz, I. G. Ruzavin, and M. V. Sviridov, Int. J. Infrared Millim. Waves **21**, 285 (2000).
22. S. A. Guz, I. G. Ruzavin, and M. V. Sviridov, Phys. Lett. A **274**, 104 (2000).
23. S. A. Guz and M. V. Sviridov, Chaos **11**, 605 (2001).
24. J. D. Bao, Phys. Lett. A **256**, 356 (1999).
25. N. N. Bogolyubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Nauka, Moscow, 1974; Gordon and Breach, New York, 1962).
26. R. L. Stratonovich, *Selected Problems of Fluctuation Theory in Radio Engineering* (Sov. Radio, Moscow, 1961).
27. A. D. Venttsel' and M. I. Freĭdlin, *Fluctuations in Dynamic Systems under Small Random Perturbations* (Nauka, Moscow, 1979).
28. N. N. Moiseev, *Asymptotic Methods in Nonlinear Mechanics* (Nauka, Moscow, 1981).
29. A. N. Malakhov, *Fluctuations in Auto-Oscillation Systems* (Nauka, Moscow, 1968).
30. E. A. Grebennikov, *Averaging Method in Applied Problems* (Nauka, Moscow, 1986).
31. M. Buckingham, *Noise in Electronic Devices and Systems* (Ellis Horwood, Chichester, 1983; Mir, Moscow, 1986).
32. *Handbook of Mathematical Functions,* Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
33. R. Mannella, Fluctuation and Noise Letters **1** (2), L45 (2001).
34. R. Mannella, in *Noise in Nonlinear Dynamical Systems*, Ed. by F. Moss and P. V. E. McClintock (Cambridge Univ. Press, Cambridge, 1989), Vol. 3, p. 189.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 175–180.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 198–204.
Original Russian Text Copyright © 2002 by Bilchinskaya, Butkovskiĭ, Kravtsov, Rychka, Surovyatkina.

MISCELLANEOUS

The Probabilistic Symmetry Breaking of Periodic Regimes Rapidly Passing through a Zone of Chaos into the Transparency Window

S. G. Bilchinskaya[a], O. Ya. Butkovskiĭ[b, *], Yu. A. Kravtsov[c, d], I. A. Rychka[a], and E. D. Surovyatkina[d, **]

[a]*Kamchatka State Technical University, Petropavlosk-Kamchatskiĭ, 683003 Russia*
[b]*Vladimir State University, Vladimir, 600029 Russia*
[c]*Space Research Center, Polish Academy of Sciences, Bartycka 18A, Warsaw, 00716 Poland*
[d]*Institute for Space Research, Russian Academy of Sciences, ul. Profsoyuznaya 84/32, GSP-7, Moscow, 117810 Russia*
**e-mail: olegb@vpti.vladimir.ru*
***e-mail: selena@iki.rssi.ru*

Received December 29, 2001

Abstract—The problem of transition of a noisy dynamical system to a periodic oscillatory regime through a zone of chaos is considered. Using the noisy logistics map as an example, domains of attraction of energetically equivalent regimes of period three are found for various transition rates and various noise levels. The fine structure of the domains of attraction under the condition of fast transitions is revealed. It is discovered that the settling time of the stable cycle of period three heavily depends on the initial conditions, i.e., on the structure of the domains of attraction. The critical transition rate that separates the region of the probabilistic symmetry of final states from the region of the dynamic behavior of trajectories is estimated. *© 2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Dynamical systems that can reside in one of several energetically equivalent stable states or regimes hold a particular place among all nonlinear systems. For example, this class includes systems that undergo bifurcations of the spontaneous symmetry breaking [1], those that admit period-doubling bifurcations [2], and degenerate parametric systems [3]. Stable oscillations in such systems differ in the phase shift $\Delta\varphi = 2\pi/M$, where M is the degeneracy of multiplicity. Similar properties are characteristic of polarizationally degenerate laser systems [4]. In chemistry, isomer molecules are analogs of systems with several energetically equivalent states [5, 6].

A specific feature of bifurcation transitions in systems described above are the probabilistic symmetry of final states. This is directly related to the effect of noise acting on the system at the moment of the bifurcation transition. If there are M equiprobable regimes, the probability of the occurrence of each of them is $1/M$. At $M = 2$, each of two stable states occurs with the probability 1/2.

The experimentally observed equiprobable realization of energetically equivalent states contradicts the following fact known from the theory of dynamical systems: by the Cauchy theorem, trajectories of the system are uniquely determined by the initial conditions [7, 8]. This contradiction between equal probability of final states in noisy bifurcation systems and their nonequivalence in the absence of noise is known as the bifurcation paradox [9]. A solution to this paradox in the particular case of the period-doubling bifurcation for a one-dimensional map was obtained in [10, 11]. It was shown in those papers that, changing the relationship between the noise level σ_η^2 and the rate of variation, $s = dr/dt$, of the bifurcation parameter r, one can trace a continuous transition from the case of the probabilistic symmetry, which is characteristic of slow transitions in noisy systems, to the strong violation of the probabilistic symmetry, which is observed under fast transitions and in the case of low noise.

It was shown in [10, 11] that the critical noise level $(\sigma_\eta^2)_c$, which divides the plane noise level σ_η^2–transition rate s into the regions of the probabilistic symmetry and asymmetry of dynamically predictable final states, obeys the power law

$$(\sigma_\eta^2)_c \propto s^\alpha, \qquad (1)$$

where α takes values in the range from one through five depending on the initial conditions. The results obtained in [10, 11] for moderate values of σ_η^2 and s were extended to arbitrary values in [12]. Later, the

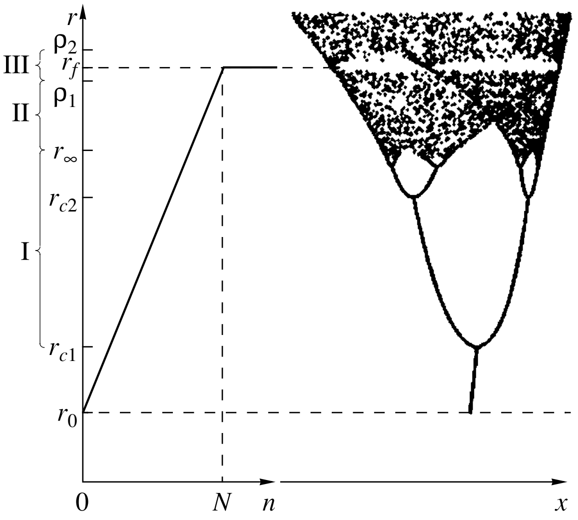


Fig. 1. The linear model of changing the control parameter r_n from the initial value r_0 to the final value r_f in N iteration steps. The bifurcation diagram is shown on the right, where the transparency window in the neighborhood of r_f is clearly seen (interval I corresponds to the cascade of period-doubling bifurcations, interval II corresponds to the zone of chaos, and interval III is the domain of stable period oscillations).

phenomenon of the probabilistic symmetry breaking was studied experimentally for a specific radiophysical system [13].

For other physical systems, including chemical reactions resulting in the emergence of right and left isomers, the bifurcation paradox has not received a satisfactory explanation yet [5, 6].

In this paper, we consider the conditions of the probabilistic symmetry breaking in a dynamical system that passes through a zone of chaos and settles itself in a regime of periodic oscillations. We consider the system governed by the logistics map, which admits the passage to an oscillatory regime of period three through a zone of chaos. The statement of the problem about the passage of a trajectory through the zone of chaos is presented in Section 2. In the absence of noise, the basins of attraction of the final periodic states are found in Section 3; in the presence of noise, this problem is solved in Section 4. The interpretation of results in terms of linear perturbation theory is discussed in Section 5, where the boundary between the regions of violation and validity of the probabilistic symmetry is approximately found. Due to an exponential rate of growth of fluctuations in the chaotic regime, this boundary is shifted to lower values of the noise level σ_η^2 compared to the power law (1).

2. STATEMENT OF THE PROBLEM

The simplest model of a dynamical system that can pass through a zone of chaos to periodic closed trajectories is the logistics map

$$x_{n+1} = F(x_n) + \eta_n = r_n x_n(1-x_n) + \eta_n, \qquad (2)$$

where η_n is the fluctuation and r_n is the controlling parameter, which depends on the discrete time n. In contrast to studies [9–13], in which only processes in the neighborhood of the first period-doubling bifurcation were analyzed ($r \approx r_{c1} = 3$), here we considerably extend the range of variation of r. More precisely, we assume that r linearly varies from the initial value $r_0 = 2.84$ ($r_0 < r_{c1}$) up to the final value r_f, which falls in the region $\rho_1 = 3.829 < r_f < \rho_2 = 3.845$ of periodic oscillations (transparency window) of period three (Fig. 1). Starting from the value r_0, the system first goes through a series of period-doubling bifurcations r_{c1}, r_{c2}, …, which extends up to the critical value $r_\infty = 3.659$; then, it crosses the zone of chaos; and finally falls into the periodicity window (ρ_1, ρ_2). For the numerical simulation, we used $r_0 = 2.84$ and $r_f = 3.84$. After the control parameter reaches the final value $r = r_f$ in N steps, the system performs a sufficiently large number of steps to complete the transient process of settling in one of the three possible stable oscillatory regimes of period three.

The time of transition from the initial state $r = r_0$ to the final state $r = r_f$ can be characterized both by the derivative

$$s = \frac{dr}{dn} \approx \frac{\Delta r}{\Delta n} = \frac{r_f - r_0}{N} \qquad (3)$$

and by the number of steps $\Delta n = N$ required to go through the interval $\Delta r = r_f - r_0$. According to (3), N is related to s by the equation $N = (r_f - r_0)/s$.

Thus, we assume that the system begins its motion from the initial state x_0 in the interval (0, 1); undergoes a cascade of period-doubling bifurcations; crosses the zone of chaos; and falls into the domain of attraction of one of the three periodic regimes of period three, which correspond to the closed trajectory shown in Fig. 2.

We shall refer to the sequences

$$\begin{aligned} x_n^{(1)} &= \{X_1X_2X_3X_1X_2X_3\ldots\}, \\ x_n^{(2)} &= \{X_2X_3X_1X_2X_3X_1\ldots\}, \qquad (4) \\ x_n^{(3)} &= \{X_3X_1X_2X_3X_1X_2\ldots\} \end{aligned}$$

as regimes (states) 1, 2, and 3, respectively. They differ from each other in the initial "phase," i.e., in the initial value X_1, X_2, or X_3 on the closed trajectory.

3. BASINS OF ATTRACTION OF PERIODIC REGIMES IN THE ABSENCE OF NOISE

The determination of attraction basins of final regimes $x_n^{(1)}$, $x_n^{(2)}$, and $x_n^{(3)}$ is reduced to finding the set of initial values $x_0^{(i)}$ ($i = 1, 2, 3$) that ultimately lead to regimes 1, 2, and 3, respectively. The latter problem defies analytical solution, and we had to use numerical simulation. The simulation results in the absence of noise are presented in Fig. 3, where the probability of settling in each of the three regimes is plotted against the initial conditions. The change from r_0 to r_f takes 20 steps, and then the system is allowed to complete the transient process, which takes up to 1000 additional iterations. The computations were carried out using the step $\Delta x_0 = 10^{-6}$.

The probabilities of settling in regimes 1, 2, and 3 are shown in Fig. 3. The calculations showed that there are wide stable domains on the axis of initial values from which the system surely goes into one of the regimes 1, 2, or 3. These wide attraction domains are intermittent with a series of short intervals of initial values of length 10^{-6} or less in which very small variations of x_0 result in the change of the final state. These series of short intervals will be referred to as riddle basins. To our knowledge, such a high sensibility to the initial values within riddle basins has not been discussed in the literature devoted to dynamic bifurcations. As the number of steps N increases, i.e., as the rate (s) of changing the parameter r decreases, the structure of attraction basins becomes finer.

Eventually, the basins of attraction acquire a much finer structure than in the case of the period-doubling bifurcation [10, 11].

The numerical simulation showed that the transient time, i.e., the settling time of the stable cycle of period three, depends heavily on the initial values. Figure 3d shows the dependence of the number of iteration steps, n_{trans}, required to complete the transient process on the initial values. Within wide stable attraction basins, 75–125 iterations are sufficient. Moreover, the wider the basin, the less the transient time n_{trans}. At the boundaries of basins, n_{trans} increases up to 225, and in the riddle basins, it is as high as 325.

4. BASINS OF ATTRACTION IN THE PRESENCE OF NOISE

In this section, we take into account the effect of fluctuations η_n. We assume that the mean value of η_n is zero, $\langle\eta_n\rangle = 0$; the fluctuations η_m and $\eta_n\rangle$ at different instances of time are independent, $\langle\eta_n\eta_m\rangle = \sigma_\eta^2\delta_{mn}$, where δ_{mn} is the Kronecker delta; and η is uniformly distributed on the interval $(-\gamma, +\gamma)$:

$$w(\eta_n) = \begin{cases} 0, & |\eta_n| > \gamma, \\ 1/2\gamma, & |\eta_n| \le \gamma. \end{cases}$$

In this case σ_η^2 and γ are related by the equation $\sigma_\eta^2 = \gamma^2/3$.

The inclusion of noise has two basic effects. The first one is the blurring of the boundaries of attraction basins, and the second effect is the equalizing of the probabilities of reaching the final regimes beginning with a certain critical value N_c. Both these phenomena were also observed in the period-doubling bifurcation [9–11]; however, in the case under consideration, this process is much faster due to the exponential growth of fluctuations in the domain of developed chaos.

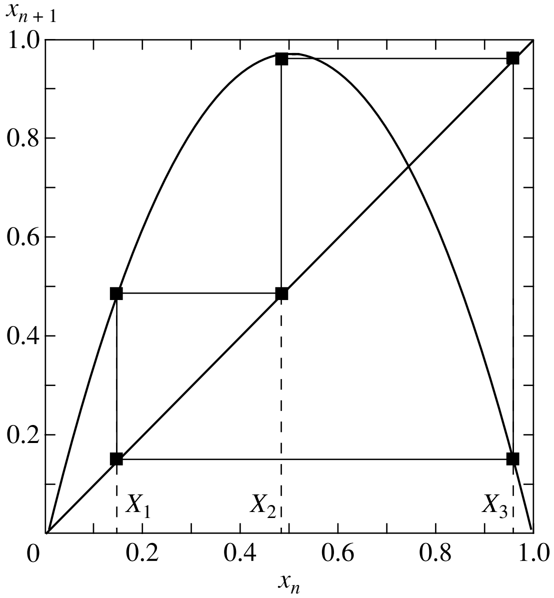


Fig. 2. The closed trajectory of period three corresponding to the control parameter $r_f = 3.84$, which belongs to the interval from $r = 3.829$ to $r = 3.845$. The closed trajectory corresponds to three periodic solutions, which differ in the initial values X_1, X_2, or X_3.

The intensity of noise required to equalize the probabilities heavily depends on the initial values. Within the basins of attraction, more intensive noise is required than on their boundaries and in the riddle basins.

These trends are clearly seen in Fig. 4, where the probabilities of settling in regimes 1, 2, and 3 are plotted against the initial values in the presence of noise. In Figs. 4a, 4b, and 4c, the noise with the variance $\sigma_\eta = 7.0 \times 10^{-4}$, 2.0×10^{-3}, and 1.1×10^{-2} was used, respectively. In the last case, the probabilities $P_1 = P_2 = P_3$ are effectively equal in the entire range of initial values; thus, $\sigma_\eta = 1.1 \times 10^{-2}$ is the critical value for $N = 20$ steps.

5. INTERPRETATION OF RESULTS ON THE BASIS OF THE LINEAR PERTURBATION THEORY

The specific features of the basins of attraction presented in Fig. 4 can be illustrated by a simple calculation based on perturbation theory. Assume that in the absence of noise the system is described by the unperturbed map

$$x_{n+1}^{(0)} = F(x_n^{(0)}). \tag{5}$$

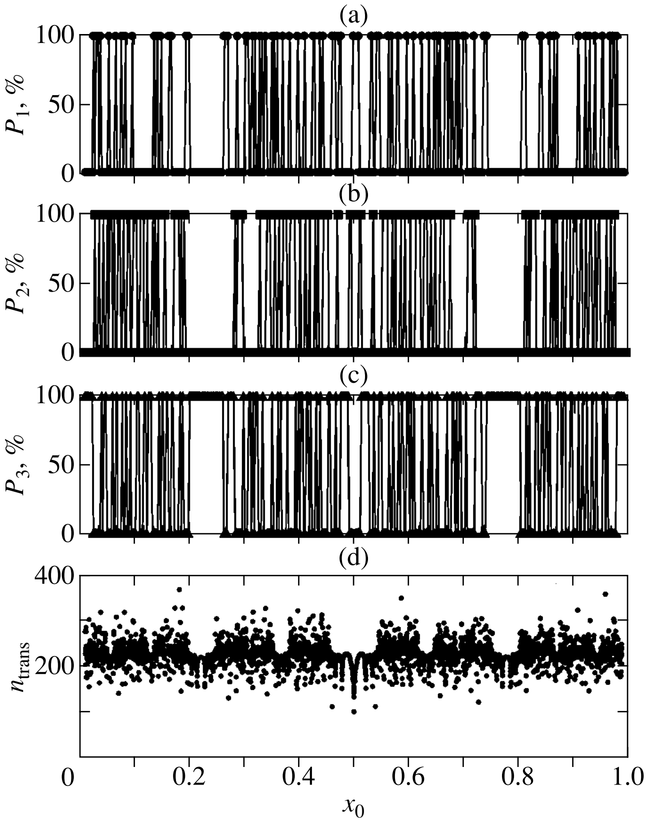


Fig. 3. Probabilities of settling in regimes 1 (a), 2 (b), and 3 (c), at $N = 20$, and the number of steps n_{trans} (d) required for the periodic regimes to settle as functions of the initial value x_0.

According to the linear perturbation theory, a small deviation x_n from the unperturbed value $x_n^{(0)}$, $\xi_n = x_n - x_n^{(0)}$, obeys the equation

$$\xi_{n+1} = \varphi_n \xi_n + \eta_n, \quad \xi_0 = 0, \tag{6}$$

where

$$\varphi_n = \left.\frac{\partial F(x_n)}{\partial x_n}\right|_{x_n = x_n^{(0)}}$$

is dilatation factor at the point $x_n^{(0)}$. The application of Eq. (6) N times yields

$$\begin{gathered} \xi_{N+1} = \varphi_N \varphi_{N-1} \ldots \varphi_1 \eta_0 \\ + \varphi_N \varphi_{N-1} \ldots \varphi_2 \eta_1 + \ldots + \varphi_N \eta_{N-1} + \eta_N. \end{gathered} \tag{7}$$

At the stage of passing through the cascade of period-doubling bifurcations, all factors φ_k are not greater than unity in absolute value; then, the mean value $\langle|\varphi_k|\rangle$, which we will denote by ν, does not exceed unity either:

$$\nu \equiv \langle|\varphi_k|\rangle \le 1. \tag{8}$$

In the domain of chaos (i.e., for $r > r_\infty = 3.659\ldots$), the mean value of the dilatation factor $\mu \equiv \langle|\varphi_k|\rangle$, which is exponentially related to the Lyapunov exponent λ, $\mu \equiv \langle|\varphi_k|\rangle \sim e^{\lambda}$, is greater than unity. If we retain only the dominating first term in (7), then we obtain the following estimate for the mean squared perturbation $\sigma_{N+1}^2 \equiv \langle \xi_{N+1}^2 \rangle$:

$$\sigma_{N+1}^2 \approx \mu^{2N_e} \nu^{2(N-N_e)} \sigma_\eta^2. \tag{9}$$

Here, N is the total number of steps, N_e is the effective number of step in the zone of chaos, and $N - N_e$ is the effective number of steps within the cascade of period-doubling bifurcations. The factor $\nu^{2(N-N_e)}$ in (8) is responsible for the decrease in fluctuations within the cascade of the period-doubling bifurcations, while the factor μ^{2N_e} is responsible for the increase in fluctuations in the domain of chaos.

The effective equalizing process of probabilities of regimes 1, 2, and 3 begins when the standard deviation σ_{N+1} becomes approximately equal to the characteristic interval $\Delta X \approx |X_i - X_j|$ between the sequential values of x on the closed trajectory of period three:

$$\sigma_{N+1}^2 \sim \mu^{2N_e} \nu^{2(N-N_e)} \sigma_\eta^2 \sim |\Delta X|^2. \tag{10}$$

This coarse estimate adequately describes the critical number of steps N_c that leads to equalizing the probabilities of final states at a given level of noise or, conversely, the critical value of the noise level $(\sigma_\eta^2)_c$ at a given N.

The quantity N_e can be estimated as the fraction of the total number of steps N falling within the domain of the chaotic regime:

$$N_e = N\chi, \quad \chi = \frac{r_f - r_\infty}{r_f - r_0}.$$

Then, it follows from (10) that, at the fixed noise level σ_η,

$$N_c \sim \frac{\ln(|\Delta X|/\sigma_\eta)}{\chi \ln\mu + (1-\chi)\ln\nu}, \tag{11}$$

whereas, at fixed N,

$$(\sigma_\eta)_c \sim \frac{\Delta X}{[\mu^{\chi}\nu^{(1-\chi)}]^N}. \tag{12}$$

In the particular case $\chi = 1/2$ (when the system is in the chaotic state about half the time), these formulas take the form

$$N_c = \frac{2\ln(|\Delta X|/\sigma_\eta)}{\ln(\mu\nu)}, \tag{13}$$

$$(\sigma_\eta)_c = \frac{\Delta X}{(\mu\nu)^{N/2}}. \tag{14}$$

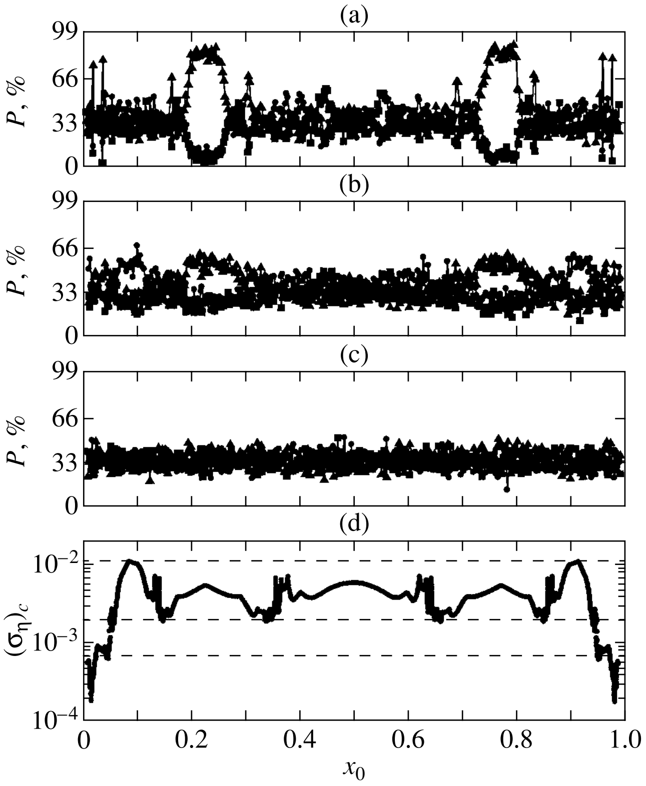


Fig. 4. The gradual equalization of the probabilities of settling in the states $x_n^{(1)}$ (■), $x_n^{(2)}$ (▲), and $x_n^{(3)}$ (●) with increasing noise: (a) $\sigma_\eta = 7.0 \times 10^{-4}$, (b) $\sigma_\eta = 2.0 \times 10^{-3}$, and (c) $(\sigma_\eta)_c = 1.1 \times 10^{-2}$ (for $N = 20$). Figure (d) shows the critical value of the noise standard deviation $(\sigma_\eta)_c$ calculated by formula (12) versus the initial value x_0. The horizontal dashed lines correspond to the noise used in the plots (a), (b), and (c), respectively.

This estimate is mainly qualitative. The numerical simulation showed that it agrees with the quantitative results for $N \in [20, 30]$.

Figure 4d presents the plot of $(\sigma_\eta)_c$ against x_0 for $N = 20$ (see formula (12)). ν was calculated by formula (8) for all initial values x_0 and r in the range from $r_0 = 2.84$ to $r_\infty < 3.659$; μ was similarly calculated for $r_\infty \geq 3.659$, taking into account the fact that $\max\Delta X = X_3 - X_1 = 0.81$. It is seen from the plot that the critical values of the noise heavily depend on the structure of the attraction domains. The horizontal dashed lines in Fig. 4d show the levels of noise corresponding to that used while constructing the plots in Figs. 4a, 4b, and 4c, respectively. The noise with $\sigma_\eta = 7.0 \times 10^{-4}$ (case *a*) is sufficient for equalizing the probabilities in mixed domains. The noise with $\sigma_\eta = 2.0 \times 10^{-3}$ (case *b*) decreases the probability within the attraction domains down to 50%, and for the noise with $(\sigma_\eta)_c = 1.1 \times 10^{-2}$ (case *c*), the probabilities P_1, P_2, and P_3 differ from 33% by no more than 5%.

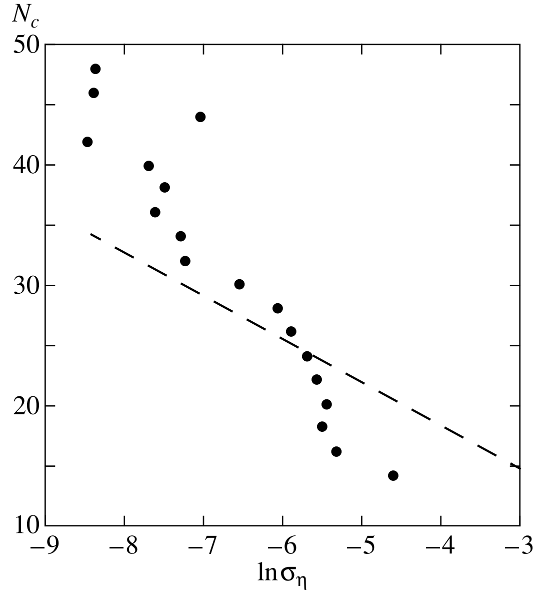


Fig. 5. The results of numerical calculations (dots) of the boundary $N_c(\sigma_\eta)$ dividing the plane (N, σ_η) into domains corresponding to the probability symmetry regime ($N > N_c$) and the regime of the violated probability symmetry ($N < N_c$) of periodic oscillations. The dashed line corresponds to N_c obtained by formula (13).

The plot $N_c(\sigma_\eta)$ is shown in Fig. 5. Every dot in the plot corresponds to the maximum value of the noise $(\sigma_\eta)_c$ for each N on the entire axis of the initial values; this corresponds to the dashed line (*c*) in Fig. 4d. Generally, N in Fig. 5 are smaller than those corresponding to the power law (1). This feature is the result of the fact that the system crosses the zone of chaos. Estimates show that formulas (11) and (12) satisfactorily agree with the numerical results for $N \in [20, 30]$; for $N < 20$, the calculated noise exceeds the one required for equalizing the probabilities; for $N > 30$, the calculated level of noise is less than that obtained experimentally.

6. NOISE-DEPENDENT HYSTERESIS

It was shown in [14, 15] that dynamic bifurcations are accompanied by the hysteresis phenomenon; more precisely, in the process of bifurcation transitions with changing parameters, the system delays in the unstable state and goes to one of the three possible stable states only after a while. This phenomenon was also observed in the system under study. It manifested itself in a delay of each period-doubling bifurcation and finally resulted in a delay of the onset of the chaotic regime. The greater the rate of the bifurcation transition, the greater the delay time at every critical value. As a consequence, the transparency windows are also shifted to greater values of the control parameter r if the transition rate s is positive and, conversely, to smaller values of r if s is negative. The presence of noise in the system results in the hysteresis breakdown. The critical level of noise (Fig. 5) breaks the hysteresis almost completely and equalizes the probabilities of the system settling in one of the three possible states.

One can judge the bifurcation transition rate by the shift of the transparency window, and the width of the transparency window yields an estimate of the noise level in the system. The hysteresis phenomenon can be used to measure noise in chaotic systems with variable parameters.

7. CONCLUSIONS

In this paper, using the logistics map as an example, we investigated the transition of a noisy dynamical system through the chaos to the oscillation regime of period three.

It is shown that, under certain relations between the external noise and the rate of the control parameter variation, the dynamic transition through the chaos domain can be predicted. The existence of attraction domains of final states determined by the initial conditions and the transition rate is established. It is found that wide stable attraction domains are intermittent with a series of short intervals of initial values of length 10^{-6} or less in which very small variations of the initial state x_0 result in the change of the final state.

Under the influence of noise or at small rates, the attraction domains are blurred, which leads to the equalizing of probabilities along the axis of initial values.

The boundary between the dynamic and stochastic scenarios of the bifurcation transition on the critical noise level–transition rate plane is found. It is shown that this boundary is shifted to greater values of noise than in the case of period-doubling bifurcations.

Analytical estimates are obtained for the critical noise level $(\sigma_\eta^2)_c$ at which the probabilities of settling in each of the states become equal for a given number of steps N or, which is the same, for a given rate of variation of the control parameter. The analytical estimates are in agreement with the results of numerical simulation.

The possibility of measuring the level of noise in nonlinear systems on the basis of the hysteresis phenomenon of transparency windows is indicated.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17418 and 00-02-17741), and by the Max Planck Institute for Physics of Complex Systems (Dresden, Germany).

REFERENCES

1. *Physical Encyclopedia* (Bol'shaya Rossiĭskaya Éntsiklopediya, Moscow, 1994), Vol. 4, p. 652.
2. H. G. Schuster, *Deterministic Chaos* (Physik-Verlag, Weinheim, 1984; Mir, Moscow, 1988).
3. A. E. Kaplan, Yu. A. Kravtsov, and V. A. Rytov, *Parametric Oscillators and Scalers* (Sov. Radio, Moscow, 1966).
4. I. N. Zheludev, Usp. Fiz. Nauk **157**, 683 (1989) [Sov. Phys. Usp. **32**, 357 (1989)].
5. V. I. Gol'danskiĭ and V. V. Kuz'min, Usp. Fiz. Nauk **157**, 3 (1989) [Sov. Phys. Usp. **32**, 1 (1989)].
6. L. L. Morozov and V. I. Gol'danskiĭ, Vestn. Akad. Nauk SSSR, No. 6, 54 (1984).
7. M. A. Shishkova, Dokl. Akad. Nauk SSSR **209**, 576 (1973).
8. A. I. Neĭshtadt, Usp. Mat. Nauk **40**, 300 (1985).
9. O. Ya. Butkovskiĭ, J. S. Brush, and Yu. A. Kravtsov, in *Predictability of Complex Dynamical System* (Springer-Verlag, Berlin, 1995).
10. O. Ya. Butkovskiĭ, J. S. Brush, Yu. A. Kravtsov, and E. D. Surovyatkina, Zh. Éksp. Teor. Fiz. **109** (6), 2201 (1996) [JETP **82**, 1186 (1996)].
11. O. Ya. Butkovskiĭ, Yu. A. Kravtsov, and E. D. Surovyatkina, Zh. Éksp. Teor. Fiz. **113**, 369 (1998).
12. S. A. Astakhov, B. P. Bezruchko, E. P. Seleznev, and D. A. Smirnov, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelineĭnaya Din. **5** (2/3), 87 (1997).
13. B. P. Bezruchko, R. N. Ivanov, V. I. Ponomarenko, and Ye. P. Seleznev, in *Proceedings of the 8th International Specialist Workshop on Nonlinear Dynamics of Electronic Systems, NDES 2001, Delft, The Netherlands, 2001*, p. 231.
14. C. Baesens, Physica D (Amsterdam) **53**, 319 (1991).
15. O. Ya. Butkovskiĭ, Yu. A. Kravtsov, and E. D. Surovyatkina, Zh. Tekh. Fiz. **67** (9), 128 (1997) [Tech. Phys. **42**, 1099 (1997)].

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 26–41.
From Zhurnal Éksperimental'noœ i Teoreticheskoœ Fiziki, Vol. 122, No. 1, 2002, pp. 33–48.
Original English Text Copyright © 2002 by Konchatnij, Merenkov.

NUCLEI, PARTICLES,
AND THEIR INTERACTION

Scanning of the $e^+e^- \longrightarrow \pi^+\pi^-$ Cross Section below 1 GeV at DAΦNE by Radiative Events[¶]

M. I. Konchatnij and N. P. Merenkov*
National Science Center Kharkov Institute of Physics and Technology, Kharkov, 61108 Ukraine
**e-mail: merenkov@kipt.kharkov.ua*

Received February 21, 2002

Abstract—The tagged photon events for the measurement of the $e^+e^- \longrightarrow \pi^+\pi^-$ total cross section by the radiative return method at DAΦNE is discussed. The effects caused by the not exactly head-on collision of beams and by the QED radiative corrections are investigated. The essential part consists of the analysis of the event selection rules that ensure the rejection of the 3-pion hadronic state and take the main properties of the multiple purpose KLOE detector into account. The study of the non-head-on effect is performed in the Born approximation by integrating over the tagged photon angles, whereas the radiative corrections are calculated neglecting this effect. Together with the quasireal electron approach, this allows us to derive analytical formulas for the correction to the cross section of the initial-state radiative process. Some numerical calculations illustrate our analytical results. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The recent measurement of the muon anomalous magnetic moment $a_\mu = (g - 2)_\mu/2$ performed in the Brookhaven E861 experiment with electroweak precision [1] has boosted the interest in a renewed theoretical calculation of this quantity [2]. The reported new world average has shown the discrepancy of 2.6 standard deviations with respect to the theoretical value based on the Standard Model calculation [3], and this may open a window into possible new physics beyond the Standard Model. On the other hand, the conclusion about a significant discrepancy between the reported data and the Standard Model prediction may be somewhat premature.

Theoretical estimations of a_μ include several contributions involving the nonperturbative hadronic sector of the Standard Model: vacuum polarization, light-by-light scattering, and higher-order electroweak corrections. Hadronic effects in the two-loop electroweak contribution are small, of the order of the experimental error, and the associated theoretical uncertainty can be brought under safe control [4].

The situation with hadronic effects in the light-by-light scattering radically changed in recent months (after the E861 data were reported) due to works cited in [5]. In these works, the authors used the description of the $\pi^0\gamma^*\gamma^*$ transition formfactor based on a large-N_C expansion and short-distance properties of QCD to calculate the pseudoscalar channel contribution (in the $\gamma^*\gamma^*$ system). The corresponding result disagrees by only its overall sign with the latest previous calculations of two different groups [6, 7]. It is interesting to note that the result in [5] forced both these groups to carefully check their programs, and they recently found their own (different) sources of the wrong sign for the pseudoscalar channel [8, 9].

The main ingredient of the theoretical prediction of a_μ, which is responsible for the bulk of the theoretical error, is the contribution of the hadron vacuum polarization. The problem is that it cannot be computed analytically because perturbative QCD loses its prediction power at low and intermediate energies, where, on the other hand, the corresponding effect is maximum. But this contribution can be calculated from the data on the total hadronic cross section σ_h for the process $e^+e^- \longrightarrow$ hadrons using a dispersion relation [10]. Because the existing data about σ_h come from different sources and do not always meet the required accuracy, they are supplemented with a theoretical input. Therefore, different estimations give different results, which either strengthen the difference between theoretical and experimental values of a_μ or make it only marginal [11–14].

The cross section σ_h also plays an important role in the evolution of the running electromagnetic coupling α_{QED} from low to high energies. This means that the interpretation of measurements at high-energy electron–positron and electron–proton colliders depends on the precise knowledge of σ_h, with one percent accuracy or even better.

The updated hadronic light-by-light contribution [5, 8, 9] decreases the discrepancy between the theory and the experiment for a_μ up to 1.5 standard deviations, and the disagreement between the Standard Model and the reported experimental value becomes not so sharp. Nevertheless, when the full set of data at the BNL col-

---

[¶] This article was submitted by the authors in English.

laboration is analyzed, the experimental error bars are expected to decrease by the additional factor three at least, and this calls for a new test of the Standard Model. The high-precision data about σ_h will play the key role in this test.

We note that the data recently derived in the direct scanning of σ_h by the CMD-2 [15] and BES II [16] collaborations were included in a new analysis [17]. This significantly reduces the error in the hadronic contribution to the shift of α_{QED} but does not remove the discrepancy in a_μ. Therefore, there exists an eminent physical reason for new measurements to accumulate high-precision data about σ_h at the total center-of-mass frame energies below 1 GeV.

The old idea to use the initial-state radiative events in the electron–positron annihilation process

$$e^-(p_1) + e^+(p_2) \longrightarrow \gamma(k) + \text{hadrons}(q) \qquad (1)$$

for the scanning of the total hadronic cross section σ_h has become quite attractive recently [18–22]. This radiative return approach allows performing the scanning measurements at the accelerators running at a fixed energy, and this circumstance is the main advantage compared to the traditional direct scanning. The reason is that the most important physical parameters, the luminosity and the beam energy, remain the same during the entire scanning at fixed-energy colliders. They must therefore be determined only once, which can be done with a very high accuracy. The drawback is of course a loss in the event number, and it is obvious that only high-luminosity accelerators can be competitive when the radiative return method is used.

It is a general opinion that the high-luminosity DAΦNE machine operating in the Φ resonance region with the multiple-purpose KLOE detector is the ideal collider to scan $\sigma_h(q^2)$ with the center-of-mass energy $\sqrt{q^2}$ varying from the threshold up to 1 GeV just by radiative events. It is now well understood that, in this energy region, the total hadronic cross section is mainly fulfilled by the contribution of the ρ resonance. This in turn implies that the dominant hadronic final state is that of the charged pion pair $\pi^+\pi^-$, and the KLOE detector allows measuring both the photon energy deposited in calorimeters and the 3-momenta of pions running through the drift chamber [20, 23].

Such a wide range of experimental possibilities of the KLOE detector can provide a realization of two approaches to scanning the $\pi^+\pi^-$ channel contribution to $\sigma_h(q^2)$: with tagged photon events [19, 20, 22] and without photon tagging [22, 24, 25]. The latter method has some advantages because it allows including the events with collinear initial-state radiative photons, which leads to the increase in the cross section by the energy logarithm enhancement factor [26].

On the other hand, the first data about the traditional tagged photon scanning of $\sigma_h(q^2)$ at DAΦNE are reported [26] (we also note that large radiative event rates were observed by the BaBar Collaboration [27]). To extract $\sigma_h(q^2)$ at different squared dipion invariant masses with one percent accuracy, one must precisely analyze the initial-state radiative events and take the final-state radiative events and the initial-state radiative interference as a background. Moreover, the radiative corrections must be calculated for all these contributions [21]. For a realistic experimental event selection, this task can usually be solved by means of Monte Carlo event generators. But for some ideal conditions, analytical calculations may be performed with high accuracy, and this is a very important test of the required one percent accuracy produced by the Monte Carlo generators.

The high-precision analysis of the initial-state radiative events is the main attribute of the radiative return method. The corresponding radiation corrections were considered in a number of papers by both the Monte Carlo generators [19, 22] and analytical calculations [18, 22, 24]. In the present paper, we derive analytical formulas for different contributions to the initial-state radiative cross section including the first-order radiative corrections. We use the same rules for the event selection as given in [24]. These selection rules are maximum draw near the experimental ones [20, 23] except the dipion angular phase space, for which we use the entire 4π opening angle. This is not the case with the realistic measurements at DAΦNE because there exists a so-called blind zone in the KLOE detector with the opening angle about 15° along the electron and positron beam directions. Any particle inside this blind zone cannot be detected either by the KLOE calorimeters or by the KLOE drift chamber.

In Section 2, we briefly recall the selection rules used here and analyze the Born cross section by numerically integrating the sufficiently complicated analytical formulas given in [24], which take the non-head-on beam collision into account. In Section 3, the contributions to radiative corrections due to the virtual and soft photon emission are obtained by analytically integrating over the angular phase space of the photon that deposits its energy in calorimeters. In Section 4, the effect of an additional hard photon emission inside the blind zone is investigated and the corresponding contribution to radiative corrections is derived. To perform the analytical calculation, we use the quasireal electron approximation [28] for both the differential cross section and the underlying kinematics. In Section 5, the total radiative correction is derived and some numerical estimates are given. The elimination of the auxiliary infrared parameter is demonstrated and the dependence of the radiative corrections to the initial-state radiative cross section on the squared dipion invariant mass q^2 and physical parameters defining the selection rules is investigated. We briefly summarize our results in the Conclusion. In the Appendices, we give some formulas that are useful in our intermediate calculations.

2. SELECTION RULES AND THE INITIAL-STATE BORN RADIATIVE CROSS SECTION

As mentioned above, the multiple purpose KLOE detector allows independently measuring the photon energy with two calorimeters QCAL and EMCAL and the 3-momenta of the charged pions with the drift chamber. The selection rules that we consider here can be formulated as follows: any event is included if only one hard photon with the energy $\omega > \omega_m$, $\omega_m = 50$ MeV, hits the calorimeters and if the difference between the lost energy Ω and the lost 3-momentum modulus $|\mathbf{K}|$ does not exceed a small value ηE, $\eta \ll 1$, where E is the beam energy. The lost energy is defined as the difference between the total initial energy and the sum of the charged pion energies, and the lost 3-momentum is defined similarly.

The first rule implies that, in addition to one hard photon, only soft photons that cannot be recorded by the KLOE detector can hit the calorimeters. The radiation of the additional hard photon is allowed inside the blind zone. The second rule ensures the removal of the 3-pion hadronic state arising due to possible $\Phi \longrightarrow \pi^+\pi^-\pi^0$ and $\omega \longrightarrow \pi^+\pi^-\pi^0$ decays. The neutral pion quickly decays into two γ quanta; one of these has time to light in the calorimeters, whereas the other can fly away into the blind zone. It is easy to see that this rule does not allow the lost invariant mass to be greater than $2E^2\eta$, and the 3-pion state is therefore forbidden if $\eta < 0.035$. Thus, the following event selection cuts are imposed [20]:

$$\begin{gathered}\Omega - |\mathbf{K}| \le \eta E, \quad \omega > \omega_m, \\ \eta = 0.02, \quad \omega_m = 50 \text{ MeV}.\end{gathered} \tag{2}$$

The first inequality in (2) is important in calculating the contribution caused by the emission of two hard photons (one inside the calorimeters and the other in the blind zone) because it only affects the phase space of two hard photons. At the Born level (with only one photon inside the calorimeters), we must therefore take the second restriction in (2) into account by introducing the trivial Θ function. The differential distribution over the dipion invariant mass can be written as [24]

$$\begin{aligned}\frac{d\sigma^B}{dq^2} &= \frac{\alpha}{2\pi^2}\sigma(q^2)\frac{(S-q^2)d\cos\theta d\varphi}{4S(2E-|\mathbf{P}_\Phi|\sin\theta\cos\varphi)^2} \\ &\times \frac{(S+T_1)^2+(S+T_2)^2}{T_1T_2}\Theta(\omega-\omega_m),\end{aligned} \tag{3}$$

$$\omega = \frac{S-q^2}{2(2E-|\mathbf{P}_\Phi|\sin\theta\cos\varphi)},$$

where θ and $\varphi(\omega)$ are the polar and azimuthal angles (energy) of the photon detected by the KLOE calorimeters. The approximation used here is valid if $E^2\theta_0^2 \gg m^2$, where $2\theta_0$ is the opening angle of the blind zone and m is the electron mass. The cross section $\sigma(q^2)$ of the process $e^+e^- \longrightarrow \pi^+\pi^-$ is expressed through the pion electromagnetic form factor $F_\pi(q^2)$ as

$$\sigma(q^2) = \frac{\pi\alpha^2|F_\pi(q^2)|^2}{3q^2}\left(1-\frac{4m_\pi^2}{q^2}\right)^{3/2},$$

where m_π is the pion mass. The invariants entering Eq. (3) are given by

$$T_1 = -2p_1k = -\omega(2E-2P_z\cos\theta-|\mathbf{P}_\Phi|\sin\theta\cos\varphi),$$

$$P_z = E\left(1-\frac{|\mathbf{P}_\Phi|^2}{8E^2}\right),$$

$$T_2 = -2p_2k = -\omega(2E+2P_z\cos\theta-|\mathbf{P}_\Phi|\sin\theta\cos\varphi),$$

$$S = 2p_1p_2 = 4E^2-|\mathbf{P}_\Phi|^2, \quad q^2 = S+T_1+T_2.$$

In writing these expressions, we took into account that, at DAΦNE, the electron and positron beams exercise not exactly a head-on collision at the interaction point, but there exists a small crossing angle between them that is equal to $|\mathbf{P}_\Phi|/E$, where $|\mathbf{P}_\Phi| = 12.5$ MeV. Because of a nonzero crossing angle, the energy of the tagged photon becomes dependent on its angular position, which complicates the exact analytical calculations. Thus, the question arises as to the magnitude of the corresponding effect. As shown in [24], there exist three regions,

$$D > 1, \quad 1 > D > \sin\theta_0, \quad \sin\theta_0 > D > -1,$$

where

$$D = \frac{4E(E-\omega_m)-q^2-|\mathbf{P}_\Phi|^2}{2\omega_m|\mathbf{P}_\Phi|},$$

and the form of the initial-state radiative cross section is different in each region. The analytical expressions for the distribution over the dipion invariant mass is simple in the first two regions, but it seems that only a numerical integration with respect to the photon polar angle is possible in the third region. We also note that, in the limiting case as $|\mathbf{P}_\Phi| \longrightarrow 0$, only the first region can occur with the obvious restriction

$$4E(E-\omega_m) > q^2.$$

The results of our calculations of the Born cross section are shown in Figs. 1 and 2. It follows from Fig. 1 that the contribution of the first region ($D > 1$) dominates in a wide interval of the dipion invariant masses. Within the approximation

$$4E^2(1-c_0) \gg \mathbf{P}_\Phi^2,$$

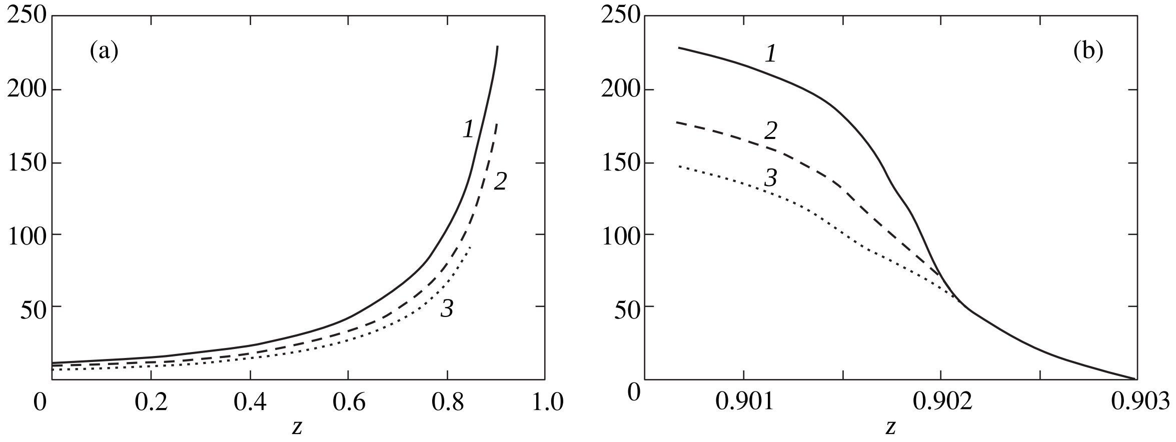


Fig. 1. The Born cross section $\frac{d\sigma}{dz}/\left[\frac{\alpha}{2\pi}\sigma(q^2)\right]$ of the initial-state radiative process $e^+e^- \longrightarrow \gamma + \pi^+\pi^-$ at different limiting angles of the blind zone versus variable z. Figure 1a corresponds to the contribution of the region $D > 1$ defined by Eq. (4). The sum of contributions of the regions $1 > D > s_0$ and $s_0 > D > -1$ is shown in Fig. 1b (see [24], Eqs. (19)–(21), for the corresponding analytical formulas). It depends on the minimum energy of the tagged photon ω_m, which we choose as 50 MeV. $\theta_0 = 5°$ (*1*), 10° (*2*), 15° (*3*).

the corresponding cross section can be written as [24]

$$\frac{d\sigma_0^B}{dq^2} = \frac{\alpha}{2\pi}\frac{\sigma(q^2)}{2E^2}F_0\Theta(D-1),$$

$$F_0 = \left(\frac{S}{S-q^2} - 1 + \frac{S-q^2}{2S}\right)$$

$$\times\left[2\ln\frac{1+c_0}{1-c_0} + \frac{|\mathbf{P}_\Phi|^2}{2E^2}\left(\ln\frac{1+c_0}{1-c_0} + \frac{c_0}{s_0^2}\right)\right] \tag{4}$$

$$-\frac{S-q^2}{S}c_0\left[1 + \frac{|\mathbf{P}_\Phi|^2}{2E^2}(3-c_0^2)\right],$$

where $c_0 = \cos\theta_0$, $s_0 = \cos\theta_0$, and the Θ function defines the maximum possible value of q^2.

The form of F_0 and D in Eq. (4) is valid if all polar angles between θ_0 and $\pi - \theta_0$ are permitted for the tagged photon. If large-angle photons radiated between θ_l and $\pi - \theta_l$ are not recorded, we must write

$$\frac{d\sigma_l^B}{dq^2} = \frac{\alpha}{2\pi}\frac{\alpha(q^2)}{2E^2}(F_0 - F_l)\Theta(D_l - 1), \tag{5}$$

where F_l can be derived from F_0 by simply replacing θ_0 with θ_l and

$$D_l = \frac{4E(E-\omega_m) - q^2 - |P_\Phi|^2}{2\omega_m|\mathbf{P}_\Phi|s_l}, \quad s_l = \sin\theta_l.$$

In Section 3, we consider this case with $\theta_l = 40°$ and call it the modified EMCAL setup.

To indicate the effect of nonzero $\mathbf{P}_\Phi$ in the first region, we show in Fig. 2 the ratio

$$R_\Phi = \frac{F_0 - F(z, c_0)}{F(z, c_0)},$$

$$F(z, c_0) = F_0(\mathbf{P}_\Phi = 0)$$

$$= \left[\frac{1+z^2}{1-z}\ln\frac{1+c_0}{1-c_0} - (1-z)c_0\right], \tag{6}$$

$$z = \frac{q^2}{4E^2}.$$

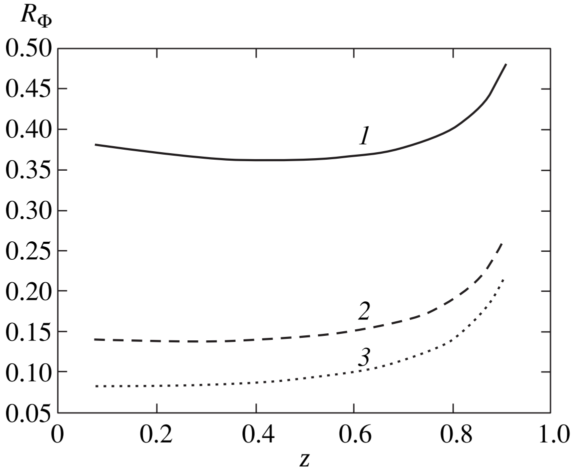


Fig. 2. The effect of a non-head-on collision of the electron and positron beams at DAΦNE is shown for the contribution of the region $D > 1$ in terms of the ratio defined by Eq. (6) at different $\theta_0 = 5°$ (*1*), 10° (*2*), 15° (*3*). The approximation $4E^2(1 - c_0) \gg |\mathbf{P}_\Phi|^2$ is used in calculating this ratio. The quantity R_Φ is given percent.

For the modified EMCAL setup, the corresponding ratio is given by

$$R^l_\Phi = \frac{F_0 - F_l - [F(z, c_0) - F(z, c_l)]}{F(z, c_0) - F(z, c_l)}. \tag{7}$$

As can be seen from Fig. 2, this effect does not exceed five per mille at $\theta_0 = 5°$ and decreases as the angle θ_0 grows.

The contribution of the third region ($\sin\theta_0 > D > -1$) is negligible everywhere, and the second region ($1 > D > \sin\theta_0$) contributes only inside a very narrow interval of the order of 2×10^{-3} near the maximum possible dipion invariant mass squared (see Fig. 1b). We therefore conclude that, for restrictions (2) considered here, the effect of non-head-on collisions on the event selection at DAΦNE is about several per mille in the most important ρ resonance region and is under control where it cannot be neglected. In this region, the corresponding cross section can therefore be given by Eq. (4) with $F(z, c_0)$ instead of F_0 with the required accuracy.

3. VIRTUAL AND SOFT CORRECTIONS

High-precision theoretical predictions are necessary in order to reach the accuracy of one percent in the measurement of the pion contribution to the hadronic cross section at DAΦNE by radiative events. These predictions must at least include the first-order radiative corrections that account for the virtual and real soft photon contribution in the overall phase space and an additional contribution due to a hard photon emission inside the blind zone. In calculating radiative corrections, we neglect $\mathbf{P}_\Phi$ at the very beginning and set s, t_1, and t_2 equal to S, T_1, and T_2 at $\mathbf{P}_\Phi = 0$, respectively.

To calculate the virtual and soft corrections, we start from the corresponding expression derived in [24] (Eq. (30)), perform the trivial azimuthal angle integration, and write the result in the convenient form

$$\frac{d\sigma^{V+S}}{dq^2} = \left(\frac{\alpha}{2\pi}\right)^2 \sigma(q^2)\frac{(1-z)d\cos\theta}{2s} \times\left[\rho\frac{(q^2-t_1)^2 + (q^2-t_2)^2}{t_1 t_2} + T\right]\Theta(1 - x_m - z), \tag{8}$$

where

$$T = \frac{3}{2}T_g - \frac{1}{8q^2}[T_{11}(q^2-t_1)^2 + T_{22}(q^2-t_2)^2 + (T_{12} + T_{21})(sq^2 - t_1 t_2)], \quad s = 4E^2,$$

$$\rho = 4(L_s - 1)\ln\Delta + 3L_q + \frac{2\pi^2}{3} - \frac{9}{2} + 4\left(\ln\frac{\Delta_1}{\Delta}\ln\frac{1+c_0}{1-c_0} + \ln\frac{\Delta_2}{\Delta_1}\ln\frac{1+c_2}{1-c_1}\right),$$

$$L_s = \ln\frac{s}{m^2}, \quad L_q = \ln\frac{q^2}{m^2}, \quad x_m = \frac{\omega_m}{E}.$$

The above expression for ρ contains three soft parameters Δ, Δ_1, and Δ_2. The first restricts the soft photon energy inside the blind zone with the value ΔE. It is auxiliary and cancels when the contribution caused by the hard photon emission is added (see Section 5). The parameters Δ_1 and Δ_2 are physical. They are defined by the sensitivity $\Delta_1 E$ of the QCAL calorimeter that surrounds the blind zone and covers polar angles of the detected photon from θ_0 to $\theta_1 = 20°$ with respect to both the electron and the positron beam directions ($c_1 = \cos\theta_1$) and by the sensitivity $\Delta_2 E$ of the EMCAL calorimeter that covers the photon angles between θ_1 and $\pi - \theta_1$. The case of a slightly modified geometry of EMCAL (with the polar angles from $\theta_l = 40°$ to $\pi - \theta_l$ not covered [20]) is considered in Appendix A. The coefficients T_g and T_{ik} on the right-hand side of Eq. (5) are calculated as functions of the invariants s, t_1, and t_2 in [24] (see also [29, 22]).

Our aim is to analytically integrate differential distribution (8) with respect to the tagged photon polar angle. The integration of the term containing ρ is trivial and yields

$$\frac{d\sigma^{S+V}_\rho}{dq^2} = \left(\frac{\alpha}{2\pi}\right)^2 \frac{\sigma(q^2)}{2E^2}\rho F(z, c_0)\Theta(1 - x_m - z). \tag{9}$$

To perform the remaining integrations, it is convenient to represent the quantity T as

$$\begin{aligned} T = {} & L_q^2 - L_s^2 + \frac{2s}{q^2 - s} - 2L_{qs} + \left[\frac{2s(s-2q^2)}{(q^2-s)^2} + 1\right]L_q \\ & + \frac{2(2q^2-s)s}{(q^2-s)^2}L_s + L_{1q}\left[-\frac{4s^2}{(q^2-s)t_2} - \frac{2(q^4+s^2)}{(q^2-s)t_1} + 2\right] \\ & + \ln\frac{-t_1}{m^2}\left[\frac{sq^2}{(s+t_2)^2} + \frac{q^2+3s}{s+t_2} - 1 - 2(L_q - L_s) \right. \\ & \left. \times\left(1 - \left(\frac{2s^2}{t_2} + \frac{q^4+s^2}{t_1}\right)\frac{1}{q^2-s}\right)\right] \\ & + \frac{2(q^4+3s^2)[6L_{qs} + 3(L_s^2 - L_q^2) + \pi^2] - 3(q^4 - 2sq^2 - s^2)}{6(q^2-s)t_1} \\ & - \left[\frac{sq^2}{(s+t_1)^2} - \frac{(q^2+3s)}{s+t_1}\right]L_q - \frac{\pi^2}{3} \\ & + \frac{s}{s+t_1} + \frac{1}{2} + (t_1 \longleftrightarrow t_2), \end{aligned} \tag{10}$$

where

$$L_{qs} = \mathrm{Li}_2\left(1-\frac{q^2}{s}\right), \quad L_{1q} = \mathrm{Li}_2\left(1-\frac{t_1}{q^2}\right).$$

The Spence function L_{1q} has a nonzero imaginary part, but we must take into account only its real part, in our calculations

$$\mathrm{Re}L_{1q} = \frac{\pi^2}{6} - \ln\left(-\frac{t_1}{q^2}\right)\ln\left(1-\frac{t_1}{q^2}\right) - \mathrm{Li}_2\left(\frac{t_1}{q^2}\right).$$

We also note that the coefficients T_g and $T_{i,k}$ contain the terms involving $t_{1,2}^{-2}$ and $t_{1,2}^{-3}$ [22, 24, 29], but these vanish in the quantity T.

Integrating the piece of cross section (5) that contains the quantity T with respect to the tagged photon polar angles, we obtain

$$\frac{d\sigma_T^{S+V}}{dq^2} = \left(\frac{\alpha}{2\pi}\right)^2\frac{\sigma(q^2)}{4E^2} \times (1-z)F_T(z, c_0)\Theta(1-x_m-z), \qquad (11)$$

where the function $F_T(z, c_m)$ is given in Appendix A for arbitrary values of the limiting angle θ_m. Here, we consider the case where $\theta_m = \theta_0$ and use the approximation $1 - c_0 \ll 1$, which is sufficiently good for $\theta_0 \le 10°$ (precisely this case is suitable for the blind zone of the KLOE detector),

$$\begin{aligned}(1-z)F_T(z, c_0 \longrightarrow 1) &= \frac{4\ln^3 z}{3(1-z)} \\ &+ \frac{2(1+z^2)}{1-z}\ln\frac{1-z}{z}\ln^2 z - (3-z)\ln\frac{(1-z)^2}{z}\ln z \\ &+ \frac{4(1-2z)}{1-z}\ln z + 4(1-z)\ln(1-z) \\ &- 2\left[3-z-2(1+z)\ln\frac{1-z}{z}\right]\mathrm{Li}_2(1-z) \\ &+ \frac{8}{1-z}\left[\mathrm{Li}_3(1-z) + z^2\mathrm{Li}_3\left(\frac{z-1}{z}\right)\right] - 5 + z \\ &- \frac{2(1+z^2)}{1-z}\ln z\ln^2\frac{1-c_0}{2} + \left[1-z-\frac{2}{1-z}-\frac{2(1+z^2)}{1-z}\right. \\ &\times\left.\left(\ln\frac{(1-z)^2}{z}\ln z + 2\mathrm{Li}_2(1-z)\right)\right]\ln\frac{1-c_0}{2}.\end{aligned} \qquad (12)$$

The total virtual and soft correction is the sum of (9) and (11).

For the modified form of the EMCAL calorimeter, the expressions for ρ and $F(z, c_0)$ in Eq. (9) and $F_T(z, c_0)$ in Eq. (11) must be changed as

$$\begin{aligned}\rho &\longrightarrow \rho + 4\ln\frac{\Delta}{\Delta_2}\ln\frac{1+c_l}{1-c_l}, \\ F(z, c_0) &\longrightarrow F(z, c_0) - F(z, c_l), \\ F_T(z, c_0) &\longrightarrow F_T(z, c_0) - F_T(z, c_l),\end{aligned} \qquad (13)$$

where the expression given in Appendix A must be used for the function $F_T(z, c_l)$ at $c_m = c_l$.

In calculating the virtual and soft corrections, we neglected terms of the order Δ_i, $i = 1, 2$ compared to unity. This accuracy implies the same relation between the tagged photon energy ω and the squared dipion invariant mass q^2 as in the Born approximation,

$$\omega = E(1-z) \qquad (14)$$

(provided that $\mathbf{P}_\Phi = 0$), and it suffices to guarantee the one percent precision.

4. HARD PHOTON EMISSION INSIDE THE BLIND ZONE

Selection rules (2) used here permit the radiation of an additional invisible photon inside the blind zone. For the events

$$e^-(p_1) + e^+(p_2) \qquad \gamma(k_1) + \gamma(k_2) + \pi^+\pi^-(q), \qquad (15)$$

one photon with the 4-momentum k_2 hits the photon detector and the other photon (with the 4-momentum k_1) is collinear and escapes it. It is obvious that relation (14) between the tagged photon energy and the squared dipion invariant mass is violated in this case.

To calculate the corresponding contribution into radiative corrections analytically, we start with using the quasireal electron approximation for both the form of the cross section and the underlying kinematics. Physically, this implies that we neglect terms of order $1 - c_0 \approx \theta_0^2/2$ and $m^2/E^2\theta_0^2$ compared to unity. We recall that $\theta_0^2/2 \le 0.02$ for the KLOE detector. In accordance with the quasireal electron approximation, the differential cross section of process (15) can be written in the same form as for the inclusive (untagged photon) event selection [25],

$$\begin{aligned}d\sigma^H &= 2\left(\frac{\alpha}{2\pi}\right)^2\frac{\sigma(q^2)}{4E^2x}P(x, L_0)dx \\ &\times\left[\frac{(q^2-xu_1)^2+(q^2-u_2)^2}{xu_1u_2}\right]\omega_2 d\omega_2 dc_2\Theta(\omega_2-\omega_m),\end{aligned} \qquad (16)$$

$$P(x, L_0) = \frac{1+x^2}{1-x}L_0 - \frac{2x}{1-x},$$

$$L_0 = \ln\frac{E^2\theta_0^2}{m^2}, \quad c_2 = \cos\theta_2,$$

$$x = 1 - \frac{\omega_1}{E}, \quad u_1 = -2k_2p_1, \quad u_2 = -2k_2p_2,$$

where $\omega_2(\theta_2)$ is the energy (the polar angle) of the tagged photon and ω_1 is the energy of the invisible collinear photon.

The factor $(\alpha/2\pi)P(x, L_0)dx$ describes the radiation probability of the collinear photon by the initial electron, the factor 2 accounts for the same contribution caused by the initial positron collinear radiation, and the rest is in fact the cross section of process (1) with the reduced electron 4-momentum ($p_1 \longrightarrow xp_1$) at $\mathbf{P}_\Phi = 0$.

Our aim is now to derive the differential distribution over the squared dipion invariant mass q^2, and it is convenient to use the relation between q^2 and c_2 in order to avoid the integration over c_2 on the right-hand side of Eq. (16). To disentangle the selection rules and obtain the integration region, it is also useful to introduce the total photon energy $\Omega = \omega_1 + \omega_2$ instead of ω_2,

$$\begin{gathered} q^2 = 4E(E-\Omega) + 2\omega_1\omega_2(1-c_2), \\ dc_2 \longrightarrow \frac{dq^2}{2\omega_1\omega_2}, \quad d\omega_2 \longrightarrow d\Omega_2. \end{gathered} \tag{17}$$

Taking into account that, in terms of new variables,

$$\begin{gathered} u_1 = -\frac{4E^2\Omega_z}{\omega_1}, \quad u_2 = -\frac{4E}{\omega_1}[\omega_1(\Omega-\omega_1) - E\Omega_z], \\ \Omega_z = \Omega - E(1-z), \end{gathered} \tag{18}$$

we can rewrite Eq. (16) as

$$\begin{gathered} \frac{d\sigma^H}{dq^2} = 2\left(\frac{\alpha}{2\pi}\right)^2\frac{\sigma(q^2)}{4E^2}M(z, L_0, \omega_1, \Omega) \\ \times\frac{d\omega_1 d\Omega}{E\Omega_z}\Theta(\Omega - \omega_1 - \omega_m), \end{gathered} \tag{19}$$

where

$$\begin{gathered} M(z, L_0, \omega_1, \Omega) = -L_0 - \frac{4(L_0-1)E\Omega_z}{\omega_1^2} \\ - \frac{zE^2L_0}{(E-\omega_1)^2} + \frac{[2z-(1+z)L_0]E}{E-\omega_1} \\ + \frac{[2(1+z^2)L_0 - 4z - (1-z)^2]E^2 - (1-z)(\Omega-2\omega_1)E}{\omega_1(\Omega-\omega_1) - E\Omega_z}. \end{gathered} \tag{20}$$

We note that although the term containing ω_1^2 in the denominator can be neglected in the case of the inclusive event selection [25], it is now important and even contributes to the cancellation of the auxiliary infrared parameter Δ.

We now find the integration region for the variables ω_1 and Ω determined by restrictions (2) and by the inequalities

$$-c_m < c_2 < c_m, \quad E\Delta < \omega_1 < \Omega - Ex_m \tag{21}$$

limiting the possible angles for the tagged noncollinear photon and the energies of the invisible collinear photon. Here, we do not require $1 - c_m$ to be small, bearing in mind a further application to the modified EMCAL setup. The first restriction in (2) defines the maximum value of Ω; the minimum value of Ω can be obtained from relation (17) at $\omega_1 = \Delta E$ and $c_2 = c_m$,

$$\begin{gathered} \Omega_{\max} = E(1-z)\left(1 + \frac{\eta}{2}\right), \\ \Omega_{\min} = E(1-z)\left(1 + \frac{\Delta(1-c_m)}{2}\right). \end{gathered} \tag{22}$$

The condition $c_2 > -c_m$ implies that

$$\begin{gathered} \omega^- < \omega_1 < \omega^+, \\ \omega^\pm = \frac{\Omega}{2}\left[1 \pm \sqrt{1 - \frac{8E\Omega_z}{\Omega^2(1+c_m)}}\right]. \end{gathered} \tag{23}$$

Finally, the inequality $c_2 < c_m$ can be formulated as follows: if the values of Ω are such that

$$\Omega < \Omega_c, \quad \Omega_c = E(1-z)\left(1 + \frac{(1-c_m)(1-z)}{8}\right),$$

then

$$\begin{gathered} \omega_1 > \omega_+ \text{ or } \omega_1 < \omega_-, \\ \omega_\pm = \frac{\Omega}{2}\left[1 \pm \sqrt{1 - \frac{8E\Omega_z}{\Omega^2(1-c_m)}}\right]. \end{gathered} \tag{24}$$

The consistent combination of the set of inequalities (21)–(24) for ω_1 and Ω defines the integration region. In general, this region depends on the dipion invariant mass through z; it is shown in Fig. 3, where we use the notation

$$\begin{gathered} \Omega_\Delta = E(1-z)(1+\Delta), \\ \Omega_x = E(1-z)\left[1 + \frac{(1-c_0)x_m y}{2(1-z)}\right], \\ y = 1 - z - x_m. \end{gathered} \tag{25}$$

This region differs from the corresponding region for the inclusive event selection [25].

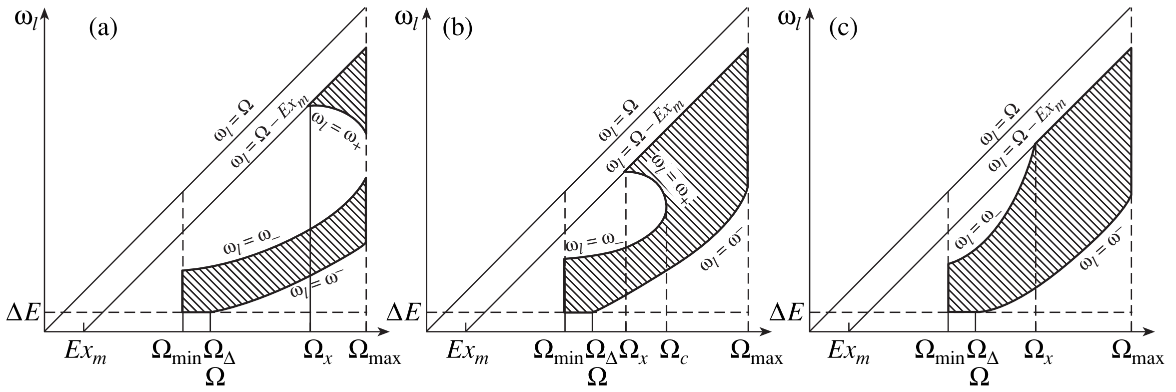


Fig. 3. The integration region for the contribution of the hard photon emission inside the blind zone in the regions $z < z_c$ (a), $z_c < z < 1 - 2x_m$ (b), and $1 - 2x_m < z < 1 - x_m$ (c). We neglect the contribution of the top piece in Fig. 3a, which is justified by numerical control.

We note that, in the calculation, we controlled our analytical integration by means of the numerical one. This allowed us to conclude that, in the case where

$$z < z_c = 1 - \frac{4\eta}{1 - c_m},$$

the contribution of the top region in Fig. 3a is small, and we excluded it from consideration. Although the regions in Figs. 3b and 3c are different, the respective contributions to cross section (19) have the same analytical form.

The list of the integrals that are required in both cases, $z < z_c$ and $z > z_c$, is given in Appendix B. Using these integrals, we write the contribution of the additional hard photon emission inside the blind zone to the radiative corrections as

$$\frac{d\sigma^H}{dq^2} = \left(\frac{\alpha}{2\pi}\right)^2 \frac{\sigma(q^2)}{2E^2} \times [G_\Delta \ln\Delta + P(z, L_0)G_p + L_0 G_1 + G_0]. \quad (26)$$

If the tagged photon is detected in the angular region $\pi - \theta_0 > \theta > \theta_0$, the coefficients G_i ($i = \Delta, p, 1, 0$) are given by

$$G_\Delta = 4(L_0 - 1)\left[1 - z + \frac{1 + z^2}{1 - z} l_0\right], \quad l_0 = \ln\frac{\theta_0^2}{4}, \quad (27)$$

$$\begin{aligned} G_p = {} & -\ln^2\frac{(1-z)\eta}{2(1-z-x_m)x_m} \\ & - 2\ln\frac{1-z}{x_m}\ln\frac{4(1-z)x_m}{\eta^2} - 4\mathrm{Li}_2\left(\frac{x_m}{1-z}\right) + \pi^2 \\ & - 2l_0\ln\frac{(1-z)(1-z-x_m)\eta}{2x_m}, \end{aligned} \quad (28)$$

$$\begin{aligned} G_1 = {} & x_m\left[2 + \left(1 + \frac{1}{z + x_m}\right)\ln\frac{1-z}{x_m}\right] \\ & - 2(1-z)\left(2\ln\frac{\eta}{2} + 3\right) + (2z + x_m)\left(\frac{1}{z + x_m} - 1\right) \\ & \times\left(l_0 + \ln\frac{2(1-z-x_m)}{\eta}\right) \\ & - (1+z)\left[\frac{1}{2}\ln^2 z - \ln z\ln\frac{x_m}{1-z}\right. \\ & - \ln(z + x_m)\left(1 - l_0 - \ln\frac{2x_m}{(1-z)\eta}\right) + \mathrm{Li}_2(1-z) \\ & \left. + \mathrm{Li}_2\left(-\frac{x_m}{z}\right) - \mathrm{Li}_2(z + x_m) + \frac{\pi^2}{6}\right], \end{aligned} \quad (29)$$

$$\begin{aligned} G_0 = {} & -\frac{\pi^2}{3}(2 - 3z) + 4(1-z)\left(1 + \ln\frac{\eta}{2}\right) \\ & + 2z\left[\frac{1}{2}\ln^2 z - \ln z\ln\frac{x_m}{1-z} + \ln(z + x_m)\left(l_0 + \ln\frac{2x_m}{(1-z)\eta}\right)\right. \\ & \left. + \mathrm{Li}_2(1-z) + \mathrm{Li}_2\left(-\frac{x_m}{z}\right) - \mathrm{Li}_2(z + x_m)\right] \\ & + (1-z)\left[\ln^2\frac{2(1-z-x_m)x_m}{(1-z)\eta}\right. \\ & + \ln\frac{1-z}{x_m}\ln\frac{4(1-z)x_m}{\eta^2} + 2l_0\ln\frac{(1-z-x_m)\eta}{2} \\ & \left. + 2\mathrm{Li}_2\left(\frac{x_m}{1-z}\right) + \mathrm{Li}_2\left(\frac{(1-z)\eta}{2(1-z-x_m)x_m}\right)\right], \end{aligned} \quad (30)$$

where we pass to the limit $1 - c_0 \ll 1$ and take into account that only the regions in Figs. 3b and 3c contribute in this limiting case.

To describe the contribution to radiative corrections caused by the double hard photon emission with the tagged photon in the range $\pi - \theta_0 > \theta > \pi - \theta_l$, $\theta_l > \theta > \theta_0$, which corresponds to the modified EMCAL setup, we must evaluate the difference

$$\frac{d\sigma^H}{dq^2}(\theta_m = \theta_0) - \frac{d\sigma^H}{dq^2}(\theta_m = \theta_l). \tag{31}$$

We have

$$\frac{d\sigma^H_M}{dq^2} = \left(\frac{\alpha}{2\pi}\right)^2 \frac{\sigma(q^2)}{2E^2}[(G_\Delta - G^l_\Delta)\ln\Delta + P(z, L_0)(G_p - G^l_p) + L_0(G_1 - G^l_1) + G_0 - G_l], \tag{32}$$

where the expressions for the functions G^l_i are given in Appendix C.

In this case, there also exists an additional region where the radiation of the hard photon at large angles can contribute. It covers polar angles from θ_l up to $\pi - \theta_l$. We take only one piece of the corresponding contribution into account (the one that is proportional to $\ln\Delta$) and write it as [24]

$$\frac{d\sigma^L_M}{dq^2} = \frac{d\sigma^B_l}{dq^2}\frac{\alpha}{2\pi}4\ln\frac{1}{\Delta}\ln\frac{1 + c_l}{1 - c_l}, \tag{33}$$

where $d\sigma^B_l/dq^2$ is defined by Eq. (5). The remaining contribution is small because of restriction (2), and we expect that it is parameterically equal to $-\alpha/2\pi l_0$ relative to the Born cross section.

5. THE TOTAL RADIATIVE CORRECTION

The total radiative correction to the cross section of the initial-state radiation process (1) with the $\pi^+\pi^-$ hadronic final sate is defined by the sum of the contributions caused by the virtual and real soft photon emission and by the radiation of the hard collinear photon inside the blind zone of the KLOE detector. In calculating the radiative correction, we suppose that $\mathbf{P}_\Phi = 0$, because the corresponding effect due to the non-head-on collision of beams cannot be greater than 10^{-3} at the radiative correction level. It is easy to see that the auxiliary infrared cutoff parameter Δ vanishes for both forms of the EMCAL calorimeter. If $\pi - \theta_0 > \theta > \theta_0$, it enters this sum in the combination

$$\left(\frac{\alpha}{2\pi}\right)^2\frac{\sigma(q^2)}{2E^2} \times \ln\Delta\left\{\left[4(L_s - 1) + 4\ln\frac{1 - c_0}{1 + c_0}\right]F(z, c_0) + G_\Delta\right\}, \tag{34}$$

where the expression inside the curly brackets vanishes in the limiting case where $1 - c_0 \ll 1$, which was used in calculating G_Δ. We can therefore write the analytical expression for the derived radiative correction as

$$\frac{d\sigma^{RC}}{dq^2} = \frac{d\sigma^B}{dq^2}\delta^{RC}, \quad \delta^{RC} = \frac{\alpha}{2\pi}\frac{V}{F(z, c_0)}, \tag{35}$$

$$V = \tilde{\rho}F(z, c_0) + \frac{1}{2}(1 - z)F_T(z, c_0) + P(z, L_0)G_p + L_0G_1 + G_0, \tag{36}$$

where $d\sigma^B/dq^2$ is defined by Eq. (4) at $\mathbf{P}_\Phi = 0$, $\tilde{\rho}$ is ρ without the terms containing $\ln\Delta$, and the limit $1 - c_0 \ll 1$ must be taken for the functions $F(z, c_0)$ and $F_T(z, c_0)$.

For the modified EMCAL calorimeter, the expression in the curly brackets in (34) is replaced by

$$4\left\{\left[L_s - 1 + \ln\frac{(1 - c_0)(1 + c_l)}{(1 + c_0)(1 - c_l)}\right] - \ln\frac{1 + c_l}{1 - c_l} - L_0 + 1\right\}[F(z, c_0) - F(z, c_l)], \tag{37}$$

where the terms in the square brackets correspond to the contribution of the virtual and soft photon emission and the remaining terms are caused by the large-angle (larger than θ_l) and small-angle (smaller than θ_0) hard photon radiation.

The total radiative correction can then be written as

$$\delta^{RC}_l = \frac{\alpha}{2\pi}\frac{V_M}{[F(z, c_0) - F(z, c_l)]}, \tag{38}$$

$$\begin{aligned} V_M &= \tilde{\rho}_M[F(z, c_0) - F(z, c_l)] \\ &+ \frac{1 - z}{2}[F_T(z, c_0) - F_T(z, c_l)] \\ &+ P(z, L_0)(G_p - G^l_p) + L_0(G_1 - G^l_1) + G_0 - G^l_0, \\ \tilde{\rho}_M &= \tilde{\rho} + 4\ln\Delta_2\ln\frac{1 - c_l}{1 + c_l}. \end{aligned} \tag{39}$$

To identify trends in the behavior of the radiative correction, we study its dependence on the physical parameters that define event selection rules (2), namely η and x_m, and the dependence on the opening angle of the blind zone θ_0 and the respective sensitiveness Δ_1 and Δ_2 of the QCAL and EMCAL calorimeters.

The results for δ^{RC} given by Eq. (35) are shown in Figs. 4 and 5. As was expected, the radiative correction is large and negative because the positive contribution caused by the real photon radiation cannot compensate the negative one-loop correction. This effect intensifies because the first inequality in (2) decreases the phase space of the additional invisible real photon. The absolute value of the radiative correction depends on z and changes from 14% near the $\pi^+\pi^-$ pair production threshold to 25% at the maximum possible squared dipion invariant mass. In the more interesting region of the

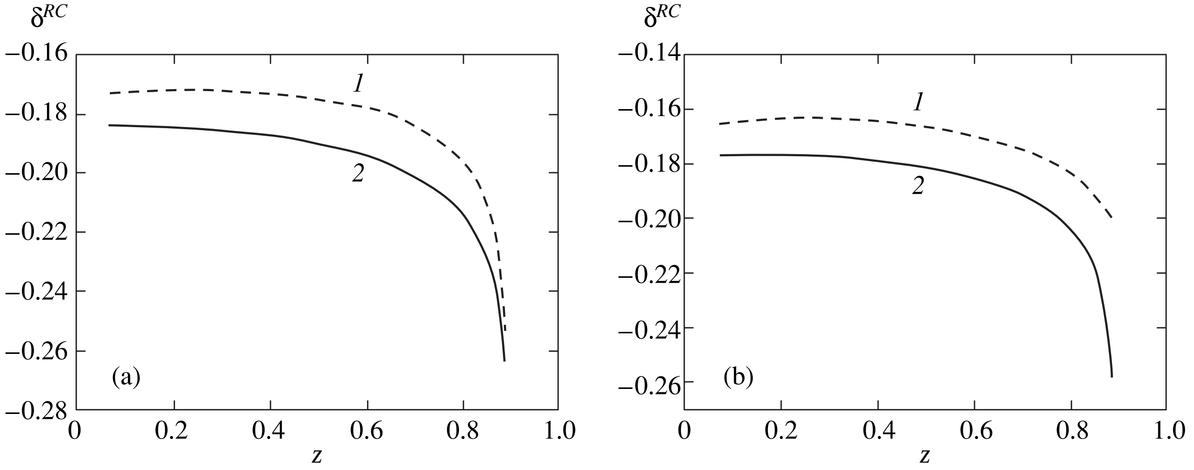


Fig. 4. Trends in the z dependence of the quantity δ^{RC} defined by Eq. (35) under the variation of the physical parameters θ_0 and x_m: (a) $\theta_0 = 10°$, $x_m = 0.098$ (*1*); $\theta_0 = 5°$, $x_m = 0.098$ (*2*); (b) $\theta_0 = 7.5°$, $x_m = 0.039$ (*1*); $\theta_0 = 7.5°$, $x_m = 0.098$ (*2*). All curves are calculated at $\eta = 0.02$, $\Delta_1 = 0.002$, and $\Delta_2 = 0.01$.

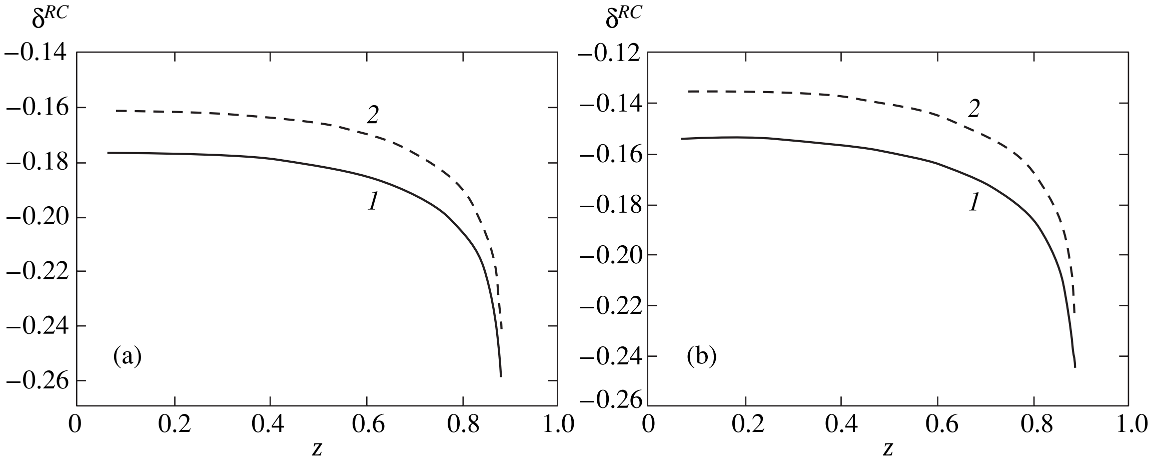


Fig. 5. Influence of the physical parameters η, Δ_1, and Δ_2 on the z dependence of δ^{RC}. The minimum energy of the tagged photon is 50 MeV ($x_m = 0.098$) (a) $\theta_0 = 7.5°$, $\eta = 0.02$, $\Delta_2 = 0.01$, $\Delta_1 = 0.002$ (*1*), 0.01 (*2*); (b) $\theta_0 = 7.5°$, $\eta = 0.03$, $\Delta_2 = 0.015$, $\Delta_1 = 0.002$ (*1*), 0.015 (*2*).

ρ resonance ($0.5 < z < 0.7$), it amounts to about 14–20%.

The main peculiarities in the behavior of the radiative correction are related to the change of the positive contribution caused by the radiation of an additional invisible hard photon. If the limiting angle θ_0 decreases, the absolute value increases because the invisible photon angular phase space is then compressed. Conversely, the decrease in the minimal energy of the tagged photon leads to an expansion of the energy phase space of the invisible photon at a fixed value of Ω (see Fig. 3) and, therefore, to a decrease of the absolute value. A similar effect occurs as the parameter η grows. But the total energy Ω of both the tagged and the invisible photons then increases, and the absolute value decreases as in the previous case.

The change in the parameters Δ_1 and Δ_2 affects the energy phase space of an additional real invisible soft photon inside the KLOE calorimeters. If these parameters are increased, the corresponding phase space expands and the absolute value decreases.

The total first-order radiative correction δ_l^{RC} for the modified EMCAL setup is shown in Fig. 6. Near the threshold, it is somewhat smaller than δ^{RC} in the absolute value, but it grows more rapidly with the increase in z.

Our calculations are restricted by considering only the first-order correction to the Born cross section. But a large absolute value requires evaluating the effects of higher order QED corrections to clarify the question of whether our approximation suffices to provide the one

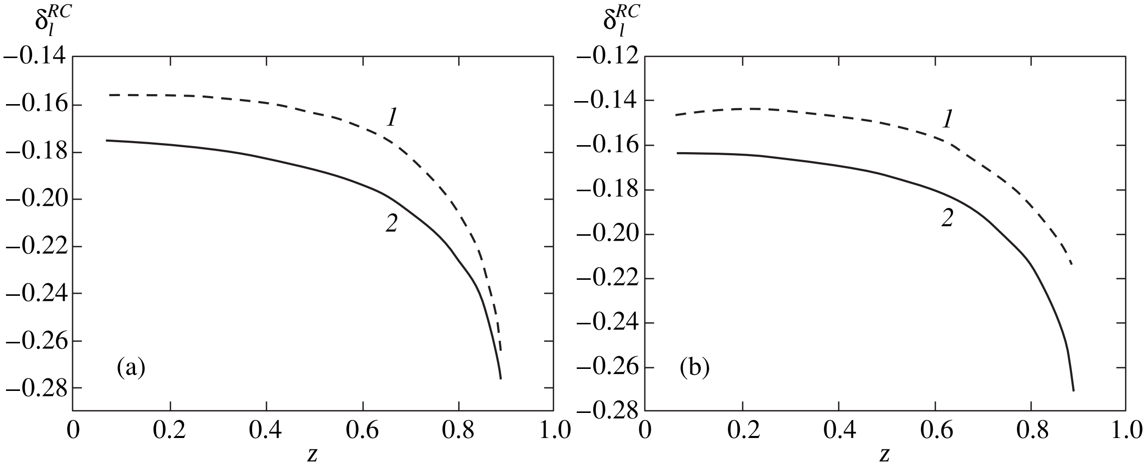


Fig. 6. The z dependence of δ_l^{RC} defined by Eq. (38) at different values of θ_0 and x_m: (a) $\theta_l = 40°$; $x_m = 0.098$; $\theta_0 = 10°$ (*1*), 5° (*2*); (b) $\theta_l = 40°$; $x_m = 0.039$ (*1*), 0.098 (*2*); $\theta_0 = 7.5°$.

percent accuracy even in the region of the ρ resonance. We hope to calculate these effects elsewhere.

6. CONCLUSIONS

The radiative return method with tagged photons offers a unique opportunity for a measurement of the total hadronic cross section $\sigma(e^+e^- \longrightarrow \text{hadrons})$ over a wide range of energies. The decrease in the event number is easily compensated by a high luminosity of the new electron–positron colliders. Of a particular interest are the experimental efforts at low and intermediate energies because they are mandatory for the future of the electroweak precision physics.

Success of the precision studies of the hadronic cross section through the measurement of radiative events relies on the matching level of reliability of the theoretical expectation. The principal problem is the analysis of radiative corrections to the initial-state radiative cross section at realistic conditions as regards the event selection. In the present work, we have developed the approach proposed in [24] for a high-precision analytical calculation of the $e^+e^- \longrightarrow \pi^+\pi^-$ channel contribution to the hadronic cross section. This channel dominates in the range below 1 GeV because of the radiative return on the ρ resonance, and the corresponding contribution can be measured with a high precision at the DAΦNE accelerator with the multiple-purpose KLOE detector [26].

Our calculations include the analysis of the effects related to a non-head-on collision of beams and the first-order radiative correction. We have demonstrated that, at the Born level, the non-head-on effects do not exceed several per mille. To derive the radiative correction, we neglected these effects and also applied the quasireal electron method [28] to describe events with two hard photons, one tagged by the KLOE calorimeters and the other invisible inside the blind zone. This approach has allowed us to analytically disentangle realistic restrictions related to the event selection rules and the KLOE detector geometry. The first-order QED radiative correction obtained in this way is negative and large in absolute value. We investigated the main trends in its behavior at the variation of the physical parameters that define experimental restrictions on event selection; we conclude that the higher order corrections must be evaluated in order to ensure the one percent accuracy required for the theoretical predictions.

APPENDIX A

Here, we give the exact result of the analytical angular integration of the quantity T (see Eq. (8)) with respect to the tagged photon polar angles at arbitrary values of the limiting angle θ_m and the squared dipion invariant mass,

$$\int\limits_{-c_m}^{c_m} T dc = 2F_T(z, c_m), \tag{A.1}$$

$$(1-z)F_T(z, c_m) = -2\frac{1+z^2}{1-z}\ln z \ln^2\frac{1-c_m}{2}$$
$$+ T_1(z, c_m)\ln\frac{1-c_m}{2} + T_0(z, c_m),$$

where

$$T_0(z, c_m) = \frac{4}{3(1-z)}\ln^3 z + 2c_m\Bigg[\frac{z-5}{2}$$

$$+ \frac{(1-z)[-4z\ln z + ((1-c_m^2)(1-z^2) + 8z)\ln(1-z)]}{[1+c_m+(1-c_m)z][1-c_m+(1+c_m)z]}$$

$$+\frac{3-6z+z^2}{1-z}\ln z\Big] + 2(2-c_m(1-z))\Big[\Big(1+\ln\frac{1+c_m}{2}\Big)$$

$$\times\ln\Big(1-\frac{(1-c_m)(1-z)}{2}\Big) + \mathrm{Li}_2\Big(-\frac{(1+c_m)(1-z)}{2z}\Big)\Big]$$

$$-2[2+c_m(1-z)]\Big[\ln\Big(1-\frac{(1+c_m)(1-z)}{2}\Big)\ln\frac{1-z}{z}$$

$$+\mathrm{Li}_2\Big(-\frac{(1-c_m)(1-z)}{2z}\Big)\Big] + 4(1-z)c_m\mathrm{Li}_2\Big(-\frac{1-z}{z}\Big)$$

$$+\frac{4(1+z^2)}{1-z}\ln\frac{1-z}{z}\Big[\mathrm{Li}_2\Big(-\frac{(1-c_m)(1-z)}{2z}\Big)$$

$$-\mathrm{Li}_2\Big(-\frac{(1+c_m)(1-z)}{2z}\Big)\Big]$$

$$+\frac{8}{1-z}\Big\{2\ln z\,\mathrm{Li}_2\Big(\frac{1-c_m}{2}\Big)+\ln\frac{1-z}{z} \tag{A.2}$$

$$\times\Big[\mathrm{Li}_2\Big(\frac{(1-c_m)(1-z)}{2}\Big)-\mathrm{Li}_2\Big(\frac{(1+c_m)(1-z)}{2}\Big)\Big]$$

$$-z^2\Big[\mathrm{Li}_3\Big(-\frac{(1-c_m)(1-z)}{2z}\Big)-\mathrm{Li}_3\Big(-\frac{(1+c_m)(1-z)}{2z}\Big)\Big]$$

$$-\mathrm{Li}_3\Big(-\frac{1-c_m}{(1+c_m)z}\Big)+\mathrm{Li}_3\Big(-\frac{(1-c_m)z}{1+c_m}\Big)$$

$$-\mathrm{Li}_3\Big(\frac{(1-c_m)(1-z)}{2}\Big)+\mathrm{Li}_3\Big(\frac{(1+c_m)(1-z)}{2}\Big)$$

$$+\frac{5+z^2}{4}\ln z\ln^2\frac{1+c_m}{2}\Big\}+\frac{2}{1-z}\ln\frac{1+c_m}{2}$$

$$\times\Big\{-2z^2\ln^2 z-(3-z)(1-z)\ln z+\frac{1+2z-z^2}{2}$$

$$+2(1+z^2)\Big[\ln z\ln(1-z)-\mathrm{Li}_2\Big(-\frac{(1+c_m)(1-z)}{2z}\Big)\Big]$$

$$+4\mathrm{Li}_2\Big(\frac{(1-c_m)(1-z)}{2}\Big)+2(1-z^2)\mathrm{Li}_2\Big(-\frac{1-z}{z}\Big)$$

$$+(1-z)^2(1+c_m)\frac{3+c_m+(1-c_m)z}{1+c_m+(1-c_m)z}\Big\},$$

$$T_1(z, c_m) = \frac{2}{1-z}\Big\{2z^2\ln^2 z+(3-z)(1-z)\ln z$$

$$-\frac{1+2z-z^2}{2}-2(1+z^2)$$

$$\times\Big[\ln z\ln(1-z)-\mathrm{Li}_2\Big(-\frac{(1-c_m)(1-z)}{2z}\Big)\Big]$$

$$-4\mathrm{Li}_2\Big(\frac{(1+c_m)(1-z)}{2}\Big)-2(1-z^2)\mathrm{Li}_2\Big(-\frac{1-z}{z}\Big) \tag{A.3}$$

$$-(1-z)[2+c_m(1-z)]\ln\Big(1-\frac{(1+c_m)(1-z)}{2}\Big)$$

$$-(1-z)^2(1-c_m)\frac{3-c_m+(1+c_m)z}{1-c_m+(1+c_m)z}\Big\}.$$

Here, we use the standard notation for the Spence functions

$$\mathrm{Li}_2(x) = -\int_0^1\frac{dt}{t}\ln(1-xt), \quad \mathrm{Li}_3(x) = \int_0^1\frac{dt}{t}\ln t\ln(1-xt).$$

If we assume that $c_m = c_0$, $1 - c_0 \ll 1$, the result in Eq. (12) is recovered.

APPENDIX B

In the case where $z > z_c$, the integration region for the double hard photon emission is shown in Figs. 3b and 3c and the corresponding differential cross section is defined by Eq. (19). The list of the necessary integrals is defined by expansion (20) of the quantity $M(z, L_0, \omega_1, \Omega)$ and is given by

$$I_1 = \int\frac{d\omega_1 d\Omega}{E\Omega_z} = y\Big[2-\ln\frac{y(1-c_m)}{\eta}\Big]-x_m\ln\frac{1-z}{x_m},$$

$$y = 1-z-x_m, \tag{B.1}$$

$$I_2 = \int\frac{d\omega_1 d\Omega}{(E-\omega_1)\Omega_z} = \frac{\pi^2}{6}+\frac{1}{2}\ln^2 z$$

$$+\ln(z+x_m)\ln\frac{(1-c_m)x_m}{\eta(1-z)}+\ln z\ln\frac{1-z}{x_m} \tag{B.2}$$

$$+\mathrm{Li}_2\Big(-\frac{x_m}{z}\Big)+\mathrm{Li}_2(1-z)-\mathrm{Li}_2(z+x_m),$$

$$I_3 = \int\frac{E d\omega_1 d\Omega}{(E-\omega_1)^2\Omega_z} = -\frac{x_m}{z(z+x_m)}\ln\frac{1-z}{x_m}$$

$$-\frac{1+z}{z}\ln(z+x_m)-\frac{y}{z+x_m}\ln\frac{(1-c_m)y}{\eta}, \tag{B.3}$$

$$\begin{aligned} I_4 &= \int\frac{(\Omega-2\omega_1)d\omega_1 d\Omega}{[\omega_1(\Omega-\omega_1)-E\Omega_z]\Omega_z} = \frac{\pi^2}{6}-\frac{1}{2}\ln^2\frac{1+c_m}{2} \\ &- \ln\frac{1-c_m}{1+c_m}\ln\frac{x_m y\eta}{2(1-z)\Delta^2}-\frac{1}{2}\ln^2\frac{2x_m y}{(1-z)\eta} \\ &- \mathrm{Li}_2\left(\frac{1-c_m}{2}\right)-\mathrm{Li}_2\left(\frac{(1-z)\eta}{2x_m y}\right), \end{aligned} \tag{B.4}$$

$$\begin{aligned} I_5 &= \int\frac{(1-z)Ed\omega_1 d\Omega}{[\omega_1(\Omega-\omega_1)-E\Omega_z]\Omega_z} = \frac{\pi^2}{2}-\frac{4y}{1-z} \\ &- 2\mathrm{Li}_2\left(\frac{x_m}{1-z}\right)-\frac{2(1+c_m)}{1-c_m}\left[\ln\frac{1+c_m}{2}\right. \\ &\left.- \ln\left(\frac{2}{1-c_m}-\frac{x_m}{1-z}\right)+\ln\left(\frac{1+c_m}{1-c_m}+\frac{x_m}{1-z}\right)\right] \\ &+ \ln\frac{1-c_m}{1+c_m}\ln\frac{2\Delta^2 x_m}{y(1-z)\eta}-\frac{1}{2}\ln^2\frac{2y}{\eta} \\ &- \frac{3}{2}\ln^2\frac{1-z}{x_m}+\ln\frac{1-x}{x_m}\ln\frac{y\eta}{2x_m^2} \\ &+ \mathrm{Li}_2\left(\frac{y(1-c_m)}{2(1-z)}\right)-\mathrm{Li}_2\left(-\frac{y(1-c_m)}{(1+c_m)(1-z)}\right) \\ &+ \mathrm{Li}_2\left(-\frac{x_m(1-c_m)}{(1+c_m)(1-z)}\right)-\mathrm{Li}_2\left(\frac{x_m(1-c_m)}{2(1-z)}\right), \end{aligned} \tag{B.5}$$

$$\begin{aligned} I_6 &= \int\frac{d\omega_1 d\Omega}{\omega_1^2} = \frac{(1-z)}{2}\Big[2(1-c_m\ln\Delta) \\ &- (1+c_m)\ln\frac{1+c_m}{\eta}-(1-c_m)\left(\ln y+\frac{x_m}{1-z}\right)\Big]. \end{aligned} \tag{B.6}$$

In calculating these integrals, we neglected terms of order x_m^2 and $(1-c_m)x_m$ compared to unity; these terms are of the same order as the parameter η.

In the cases where

$$z < 1-\frac{4\eta}{1-c_m},$$

we must integrate over the region shown in Fig. 3a. As mentioned above, the contribution of the top piece of this region, where

$$\Omega - Ex_m > \omega_1 > \omega^+,$$

is small (about 1–2%) compared to the bottom one and can be neglected. This approximation is sufficient to provide the one percent accuracy of radiative correction. We use the notation J_i, similarly to I_i, to label separate integrals over this region,

$$\begin{aligned} J_1 &= I_1 + L_K - 2(K_+ - K_-), \\ K_\pm &= \frac{1}{2}\left(1\pm\sqrt{1-\frac{4\eta}{(1-c_m)(1-z)}}\right), \\ L_K &= \ln\frac{K_+}{K_-}, \end{aligned} \tag{B.7}$$

$$\begin{aligned} J_2 &= I_2 - \ln z L_K + \mathrm{Li}_2((1-z)K_-) - \mathrm{Li}_2((1-z)K_+) \\ &- \mathrm{Li}_2\left(-\frac{(1-z)K_-}{z}\right)+\mathrm{Li}_2\left(-\frac{(1-z)K_+}{z}\right), \end{aligned} \tag{B.8}$$

$$J_3 = I_3 + \frac{1-z}{z}L_K - \frac{1+z}{z}\ln\frac{z+(1-z)K_+}{z+(1-z)K_-}, \tag{B.9}$$

$$J_4 = I_4 + \ln\frac{1-c_m}{2}\left(\ln\frac{1-c_m}{2}-\frac{1}{2}\ln\frac{1-c_m}{1+c_m}\right), \tag{B.10}$$

$$\begin{aligned} J_5 &= 2\ln\frac{1-c_m}{1+c_m}\ln\Delta+\frac{\pi^2}{6}+2\mathrm{Li}_2(K_-)-4K_- \\ &+ \frac{1}{2}\ln^2\frac{1-c_m}{1+c_m}-2\ln(1-z)\ln\frac{1-c_m}{1+c_m} \\ &- \frac{1}{2}\ln^2\frac{(1-c_m)\eta}{2(1+c_m)(1-z)} \\ &- \frac{1}{2}\ln\frac{(1+c_m)^2\eta}{4(1-c_m)(1-z)}L_K+\ln K_-\ln K_+ \\ &+ \frac{2(1+c_m)}{1-c_m}\ln\frac{2(1+c_m+(1-c_m)K_-)}{(1+c_m)(1+c_m+(1-c_m)K_+)} \\ &+ \mathrm{Li}_2\left(-\frac{(1-c_m)K_+}{1+c_m}\right)-\mathrm{Li}_2\left(-\frac{(1-c_m)K_-}{1+c_m}\right) \\ &+ \mathrm{Li}_2\left(-\frac{(1-c_m)K_-}{2}\right)-\mathrm{Li}_2\left(-\frac{(1-c_m)K_+}{2}\right). \end{aligned} \tag{B.11}$$

$$\begin{aligned} J_6 &= I_6 + \frac{(1-z)(1-c_m)}{2}\left(\ln\frac{(1-z)(1-c_m)}{\eta}\right. \\ &\left.- \frac{x_m}{1-z}-2\ln K_+ - 2(K_+ - K_-)\right). \end{aligned} \tag{B.12}$$

APPENDIX C

In this appendix, we give the analytical form of the functions G_i^l for arbitrary values of the limiting tagged photon angle θ_l. The only condition on θ_l used in Appendix B is that

$$(1-c_l)x_m \ll 1.$$

This restricts θ_l by the values about 45°. With the exception of G^l_Δ, the functions G^l_i are different for $z > z_c$ and $z < z_c$. In the first case, we have

$$\begin{aligned} G^l_\Delta(z > z_c) &= 4(L_0 - 1) \\ &\times\Bigl[c_l(1-z) + \frac{1+z^2}{1-z}\ln\frac{1-c_l}{1+c_l}\Bigr]\ln\Delta, \end{aligned} \tag{C.1}$$

$$\begin{aligned} G^l_p(z > z_c) &= -2\ln\frac{1-c_l}{1+c_l}\ln\frac{(1-z)y\eta}{2x_m} - \ln^2\frac{x_m\eta}{2(1-z)y} \\ &- 2\ln\frac{1-z}{x_m}\ln\frac{(1-z)x_m}{y^2} + \pi^2 - \frac{8y}{1-z} \\ &- \frac{4(1+c_l)}{1-c_l}\ln\frac{(1+c_l)[(1-c_l)(1-z) + x_m(1-c_l)]}{2[2(1-z) - x_m(1-c_l)]} \\ &- 4\mathrm{Li}_2\Bigl(\frac{x_m}{1-z}\Bigr) + 2\mathrm{Li}_2\Bigl(\frac{(1-c_l)y}{2(1-z)}\Bigr) - 2\mathrm{Li}_2\Bigl(\frac{(1-c_l)x_m}{2(1-z)}\Bigr) \\ &+ 2\mathrm{Li}_2\Bigl(-\frac{(1-c_l)x_m}{(1+c_l)(1-z)}\Bigr) - 2\mathrm{Li}_2\Bigl(\frac{(1-c_l)y}{(1+c_l)(1-z)}\Bigr), \end{aligned} \tag{C.2}$$

$$\begin{aligned} G^l_1(z > z_c) &= x_m\Bigl(2(2-c_l) + \frac{1+z+x_m}{z+x_m}\ln\frac{1-z}{x_m}\Bigr) \\ &+ 2(1-z)\Bigl[(1-c_l)\ln y - (1+c_l)\ln\frac{1+c_l}{\eta} - 3\Bigr] \\ &+ \frac{y(2z+x_m)}{z+x_m}\ln\frac{(1-c_l)y}{\eta} \\ &+ (1+z)\Bigl[\ln\frac{1-z}{x_m}\ln\frac{z+x_m}{z} \\ &+ \ln(z+x_m)\Bigl(1 - \ln\frac{1-c_l}{\eta}\Bigr) + \mathrm{Li}_2\Bigl(\frac{z-1}{z}\Bigr) \\ &- \mathrm{Li}_2\Bigl(-\frac{x_m}{z}\Bigr) + \mathrm{Li}_2(z+x_m) - \frac{\pi^2}{6}\Bigr], \end{aligned} \tag{C.3}$$

$$\begin{aligned} G^l_0(z > z_c) &= \frac{\pi^2}{3}(3z-2) - 2x_m(3-c_l) \\ &- 2z\Bigl[\ln\frac{1-z}{x_m}\ln\frac{z+x_m}{z} - \ln(z+x_m)\ln\frac{1-c_l}{\eta} \\ &- \mathrm{Li}_2\Bigl(\frac{z-1}{z}\Bigr) + \mathrm{Li}_2\Bigl(-\frac{x_m}{z}\Bigr) + \mathrm{Li}_2(z+x_m)\Bigr] \\ &+ (1-z)\Bigl[8 + \frac{1}{2}\ln^2\frac{1+c_l}{2} - \ln^2\frac{1-c_l}{1+c_l} + \ln^2\frac{(1-c_l)\eta}{2(1+c_l)y} \\ &+ 4y\ln\frac{1-c_l}{1+c_l} + 2\ln\frac{1-z}{x_m}\ln\frac{1-z}{z} - 2(1-c_l)\ln y \\ &+ \mathrm{Li}_2\Bigl(\frac{1-c_l}{2}\Bigr) + 2\mathrm{Li}_2\Bigl(\frac{x_m}{1-z}\Bigr) + \mathrm{Li}_2\Bigl(\frac{(1-z)\eta}{2yx_m}\Bigr) \\ &+ \mathrm{Li}_2\Bigl(\frac{(1-c_l)x_m}{2(1-z)}\Bigr) - \mathrm{Li}_2\Bigl(\frac{(1-c_l)y}{2(1-z)}\Bigr) \\ &- \mathrm{Li}_2\Bigl(-\frac{(1-c_l)x_m}{(1+c_l)(1-z)}\Bigr) + \mathrm{Li}_2\Bigl(-\frac{(1-c_l)y}{(1+c_l)(1-z)}\Bigr) \\ &+ 2\frac{1+c_l}{1-c_l}\Bigl(\ln\frac{(1+c_l)[(1-c_l)(1-z) + x_m(1-c_l)]}{2[2(1-z) - x_m(1-c_l)]} \\ &- \ln\frac{1+c_l}{\eta}\Bigr)\Bigr]. \end{aligned} \tag{C.4}$$

In the case where $z < z_c$, the corresponding functions G^l_i are given by

$$\begin{aligned} G^l_p(z < z_c) &= \frac{\pi^2}{3} - 8K_- - \ln^2\frac{\eta}{2(1-z)} \\ &- 2\ln\frac{1-c_l}{1+c_l}\ln\frac{\eta(1-z)}{2} + 2\ln K_+\ln K_- \\ &- \ln\frac{(1+c_l)^2\eta}{4(1-c_l)(1-z)}L_K \\ &+ \frac{4(1+c_l)}{1-c_l}\ln\frac{2[1+c_l+(1-c_l)K_-]}{(1+c_l)[1+c_l+(1-c_l)K_+]} \\ &+ 2\mathrm{Li}_2\Bigl(\frac{(1-c_l)K_-}{2}\Bigr) - 4\mathrm{Li}_2(K_-) - 2\mathrm{Li}_2\Bigl(\frac{(1-c_l)K_+}{2}\Bigr) \\ &+ 2\mathrm{Li}_2\Bigl(-\frac{(1-c_l)K_+}{1+c_l}\Bigr) - 2\mathrm{Li}_2\Bigl(-\frac{(1-c_l)K_-}{1+c_l}\Bigr), \end{aligned} \tag{C.5}$$

$$\begin{aligned} G^l_1(z < z_c) &= 2x_m(3-2c_l) \\ &+ \frac{2-y}{1-y}\Bigl[x_m\ln\frac{1-z}{x_m} + y\ln\frac{y(1-c_l)}{\eta}\Bigr] \\ &+ 2(1-z)\Bigl[(3-2c_l)(K_+ - K_-) - 3 - (2-c_l)L_K \\ &+ (1-c_l)\ln y + (1+c_l)\ln\frac{1+c_l}{\eta}\Bigr] \\ &+ (1+z)\Bigl[-\frac{\pi^2}{6} + \Bigl(1 + \ln\frac{(1-z)\eta}{(1-c_l)x_m}\Bigr)\ln(z+x_m) \\ &+ \ln z\Bigl(L_k - \ln\frac{1-z}{x_m}\Bigr) + \ln\frac{z+(1-z)K_+}{z+(1-z)K_-} + \mathrm{Li}_2\Bigl(\frac{z-1}{z}\Bigr) \end{aligned} \tag{C.6}$$

$$+\mathrm{Li}_2(z+x_m)-\mathrm{Li}_2\left(-\frac{x_m}{z}\right)+\mathrm{Li}_2\left(\frac{(z-1)K_-}{z}\right)$$

$$-\mathrm{Li}_2\left(\frac{(z-1)K_-}{z}\right)+\mathrm{Li}_2((1-z)K_+)-\mathrm{Li}_2((1-z)K_-)\Bigg],$$

$$G_0^l(z<z_c) \;=\; -4x_m(1-c_l)+(1-z)\Bigg[6-\frac{\pi^2}{3}$$

$$-2(3-2c_l)(K_+-K_-)+\frac{1}{2}\ln^2\frac{(1-z)^2}{yx_m}-2(1-c_l)\ln y$$

$$-2(1+c_l)\ln\frac{1+c_l}{\eta}+\ln^2\frac{1-c_l}{2}-2\ln\frac{1-c_l}{1+c_l}\ln\frac{1-c_l}{2yx_m}$$

$$+\ln^2\frac{2(1-z)(1+c_l)}{\eta(1-c_l)}$$

$$+\left(2(1-c_l)+\frac{1}{2}\ln\frac{(1+c_l)^2\eta}{4(1-c_l)(1-z)}\right)L_K-\ln K_+\ln K_-$$

$$-\ln\frac{(1-z)^2}{yx_m}\ln\frac{2(1+c_l)(1-z)}{(1-c_l)\eta}-2\mathrm{Li}_2(K_-)$$

$$-\mathrm{Li}_2\left(\frac{(1-c_l)K_-}{2}\right)+\mathrm{Li}_2\left(\frac{(1-c_l)K_+}{2}\right) \qquad \text{(C.7)}$$

$$+\mathrm{Li}_2\left(-\frac{(1-c_l)K_-}{1+c_l}\right)-\mathrm{Li}_2\left(-\frac{(1-c_l)K_+}{1+c_l}\right)$$

$$+\mathrm{Li}_2\left(\frac{1-c_l}{2}\right)+\mathrm{Li}_2\left(\frac{(1-z)\eta}{2yx_m}\right)$$

$$+\frac{2(1+c_l)}{1-c_l}\ln\frac{(1+c_l)[1+c_l+(1-c_l)K_+]}{2[1+c_l+(1-c_l)K_-]}\Bigg]$$

$$+2z\Bigg[\frac{\pi^2}{6}+\ln\frac{(1-c_l)x_m}{(1-z)\eta}\ln(z+x_m)-\ln z\left(L_K-\ln\frac{1-z}{x_m}\right)$$

$$+\mathrm{Li}_2\left(-\frac{x_m}{z}\right)-\mathrm{Li}_2\left(\frac{z-1}{z}\right)-\mathrm{Li}_2(z+x_m)$$

$$+\mathrm{Li}_2((1-z)K_-)-\mathrm{Li}_2((1-z)K_+)$$

$$+\mathrm{Li}_2\left(\frac{(z-1)K_+}{z}\right)-\mathrm{Li}_2\left(\frac{(z-1)K_+}{z}\right)\Bigg].$$

REFERENCES

1. R. M. Carey *et al.* (g-2 Collab.), Phys. Rev. Lett. **82**, 1632 (1999); H. M. Brown *et al.* (g-2 Collab.), Phys. Rev. D **62**, 091101 (2000); Phys. Rev. Lett. **86**, 2227 (2000); B. Lee Roberts *et al.* (g-2 Collab.), hep-ex/0111046.
2. A. H. Höcker, hep-ph/0111243; J. F. de Trocóniz and F. J. Yndurain, hep-ph/0111258.
3. V. M. Hughes and T. Kinoshita, Rev. Mod. Phys. **71**, S133 (1999).
4. S. Peris, M. Perrottet, and E. de Rafael, Phys. Lett. B **355**, 523 (1995); A. Czarneski, B. Krause, and W. J. Marciano, Phys. Rev. D **52**, R2619 (1995); E. A. Kuraev, T. V. Kuchto, and A. Shiller, Yad. Fiz. **51**, 1631 (1990) [Sov. J. Nucl. Phys. **51**, 1031 (1990)]; E. A. Kuraev, T. V. Kuchto, Z. K. Silagadze, and A. Shiller, Preprint INP 90-66 (Budker Institute of Nuclear Physics, Siberian Division, Academy of Sciences of USSR, Novosibirsk, 1990).
5. M. Knecht and A. Nyffeler, hep-ph/0111058; M. Knecht, A. Nyffeler, M. Perrottet, and E. de Rafael, hep-ph/0111059.
6. J. Bijnens, E. Pallante, and J. Prades, Phys. Rev. Lett. **75**, 3781 (1995); Nucl. Phys. B **474**, 379 (1996).
7. M. Hayakawa, T. Kinoshita, and A. I. Sanda, Phys. Rev. Lett. **75**, 790 (1995); Phys. Rev. D **54**, 3137 (1996); M. Hayakawa and T. Kinoshita, Phys. Rev. D **57**, 465 (1998).
8. J. Bijnens, E. Pallante, and J. Prades, Preprint LUTP 01-37, hep-ph/0112255.
9. M. Hayakawa and T. Kinoshita, Preprint KEK-TH-993, hep-ph/0112102.
10. N. Cabibbo and R. Gatto, Phys. Rev. **124**, 1577 (1961); M. Gourdin and E. de Rafael, Nucl. Phys. B **10**, 667 (1967).
11. A. Charnecki and W. J. Marciano, hep-ph/0102122; W. J. Marchiano and B. L. Roberts, hep-ph/0105056; K. Melnikov, SLUC-PUB-8844, hep-ph/0105267; F. J. Yndurain, hep-ph/0102312; F. Jegerlehner, hep-ph/9901386.
12. M. Davier and H. Höcker, Phys. Lett. B **419**, 419 (1998); **435**, 427 (1998).
13. S. Narison, Phys. Lett. B **513**, 53 (2001); V. Cirigliano, G. Ecker, and H. Neufeld, Phys. Lett. B **513**, 361 (2001).
14. J. F. Trocóniz and F. J. Yndurain, hep-ph/0106025; J. Prades, hep-ph/0108192.
15. R. R. Akhmetshin *et al.*, hep-ex/9904027; Nucl. Phys. A **675**, 424 (2000); Phys. Lett. B **475**, 190 (2000).
16. D. Kong, hep-ph/9903521; J. Z. Bai *et al.* (BES Collab.), Phys. Rev. Lett. **84**, 594 (2000); hep-ph/0102003.
17. F. Jegerlehner, hep-ph/0104304; hep-ph/0105283.
18. A. B. Arbuzov *et al.*, JHEP 12 (1998) 009; M. Konchatnij and N. P. Merenkov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 769 (1999) [JETP Lett. **69**, 811 (1999)].
19. S. Binner, J. H. Kũhn, and K. Melnikov, Phys. Lett. B **459**, 279 (1999); K. Melnikov *et al.*, Phys. Lett. B **477**, 114 (2000); H. Czyź and J. H. Kũhn, Eur. Phys. J. C **18**, 497 (2001).
20. S. Spagnolo, Eur. Phys. J. C **6**, 637 (1999); G. Cataldi *et al.*, KLOE MEMO 195 (August 13, 1999).
21. A. Hoefer, J. Gluza, F. Jegerlehner, hep-ph/0107154.
22. G. Rodrigo *et al.*, Eur. Phys. J. C **22**, 81 (2001); hep-ph/0106132; G. Rodrigo, hep-ph/0111151; G. Rodrigo *et al.*, hep-ph/0112184.

23. G. Cataldi *et al.*, in *Physics and Detectors at DAΦNE, 1999*, p. 569.
24. V. A. Khoze *et al.*, Eur. Phys. J. C **18**, 481 (2001).
25. N. P. Merenkov and O. N. Shekhovtsova, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 69 (2001) [JETP Lett. **74**, 65 (2001)]; V. A. Khoze *et al.*, hep-ph/0202021, submitted to Eur. Phys. J. C.
26. M. Adinolfi *et al.* (KLOE Collab.), hep-ex/0006036; A. Aloisio *et al.* (KLOE Collab.), hep-ex/0107023; A. Denig (on behalf of KLOE Collab.), hep-ex/0106100.
27. E. P. Solodov (BABAR Collab.), hep-ex/0107027.
28. V. N. Baier, V. S. Fadin, and V. A. Khoze, Nucl. Phys. B **65**, 381 (1973).
29. E. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **45**, 782 (1987) [Sov. J. Nucl. Phys. **45**, 486 (1987)].

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 42–47.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 49–56.
Original Russian Text Copyright © 2002 by Tseskis.

FLUIDS

The Hydrodynamics of Two-Dimensional Turbulence: Relation between Self-Similar Spectral Ranges and Anomaly of Third Moments

A. L. Tseskis
Institute of High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia
e-mail: tseskis@yandex.ru

Received May 16, 2001

Abstract—Self-similar spectra of two-dimensional turbulence are discussed, as well as their correlation with the conservation laws and with the form of the function characterizing the energy transfer in the wave vector space. It is demonstrated how the form of this function affects the direction of fluxes of energy and enstrophy in the **k** space. The assumptions of the relationship of these fluxes with the time derivative of the correlation function (which are in fact similar to Kolmogorov's hypotheses for three-dimensional turbulence) enable one to derive relations which demonstrate the anomalous behavior of the third moments of two-dimensional turbulence, observed in recent experiments. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Two-dimensional turbulence provides one of few examples of physical problems whose solution does not become simpler as the space dimensionality is reduced. The presence of "redundant" (with respect to the three-dimensional case) conservation laws enables one to prove, in two-dimensional hydrodynamics of incompressible fluid, the existence and uniqueness of solutions for both the Euler and the Navier–Stokes equation [1–3]; however, it gives no way of constructing a solution fit for describing the turbulence. Therefore, for providing such a description, similarly to the three-dimensional case, one has to turn to the respective relations of statistical hydrodynamics, i.e., the Karman–Howarth equation (and its spectral analog). The latter relations, however, have one and the same form for two and three dimensions; accordingly, they differ only by the value of a single integral parameter contained in them (see [4]) and have the same number of unknown functions. The correlations between the components of correlation tensors in the two- and three-dimensional cases are also similar in this sense; consequently, the two-dimensional problem is not simpler than the three-dimensional one. Moreover, the presence of an additional quadratic integral of motion makes difficult a direct application of the known hypotheses of universality which are valid in three-dimensional statistical hydrodynamics; by virtue of this, it turns out that it is not obvious that various relations similar to those observed in the three-dimensional case [5, 6] may be found for the correlation and structure functions.

As to the possibilities of experimental studies into two-dimensional turbulence, they are also quite limited, which is due to the need for employing external stimulation in order to provide for a two-dimensional structure of turbulent motion. Such stimulation may be provided by an external uniform magnetic field (the fluid must be conducting) or by the rotation of a system as a whole with a constant angular velocity (see the discussion of these problems in [7]). It proves rather difficult to obtain the values of appropriate parameters (for example, magnetic field) that would provide, with adequate accuracy, the two-dimensionality of motion at high values of the Reynolds number. New possibilities of experimental observation of two-dimensional turbulence were indicated in [8]: they are associated with the generation of turbulence in a thin soap film by passing this film through a two-dimensional comb; in doing so, the two-dimensionality of motion is ensured by the smallness of the ratio of the film thickness to the characteristic scales of turbulence. However, even in such experiments, one still cannot attain high values of the Reynolds number for technical reasons. Numerous recent studies are devoted to direct numerical force of two-dimensional turbulence. In these studies, because of the limited computer capabilities, no values of the Reynolds number above 10^3 could be attained either [9, 10].

In spite of the above-identified difficulties, the basic qualitative distinctions of two-dimensional turbulent motion from three-dimensional one have now been found and the reasons for these distinctions understood. These distinctions reside primarily in the fact that, in the case of two-dimensional turbulent motion, the energy transfer from components with higher wave numbers to those with lower wave numbers proves to be possible, i.e., the energy transfer from small-scale to large-scale motions. This possibility was first suggested by Onsager [11], who also noted that the final stage of

1063-7761/02/9501-0042$22.00 © 2002 MAIK "Nauka/Interperiodica"

decay of two-dimensional turbulence may reduce to the formation of a coherent vortex structure. Note that Onsager [11] treated the turbulent motion as a motion of a collection of point vortexes (the appropriate representation of the vortex field in the form of a set of discrete components enables one to impart the Hamiltonian form to the equations of motion and, after that, apply the methods of statistical mechanics); the thus obtained results are also valid for the so-called guiding-center plasma [12] and for two-dimensional problems associated with superfluid motion [13]. It was later demonstrated that these results could be obtained purely hydrodynamically [4, 14]; in particular, it turned out that the spectrum of two-dimensional turbulence, which reduced to the δ function at the final stage of decay, followed directly from some self-similar solution of the Karman–Howarth equation (the respective properties of the correlation functions were additionally discussed in [15]). The formation of a coherent vortex system at the final stage of decay of two-dimensional turbulence was demonstrated both experimentally [8] and by way of direct numerical force [9].

As to the intermediate stages of decay, it is customary to assume that, along with the Kolmogorov spectrum (the energy density in the wave number space **k**) $E(k) \propto k^{-5/3}$, the spectrum $E \propto k^{-3}$ is observed in the two-dimensional case as well; this latter spectrum is due to the flux of mean-square vorticity (enstrophy) in the **k** space (the most recent detailed review of related papers is found in [6]). Sometimes, the so-called canonical spectrum is also treated (a and b are constants),

$$E(k) \propto \frac{k}{a + bk^2},$$

which may be obtained, for example, by representing turbulent motion as a system of the hydrodynamic type [16] (for such systems, the Liouville theorem is valid) with subsequent application of statistical methods. Note, however, that such systems, generally speaking, cannot be treated as a combination of noninteracting subsystems, so that the integrals of motion are not additive, and therefore it is not quite correct to apply the methods of statistical physics to this system; in the case of hydrodynamic turbulence, the use of these methods often brings about absurd results [17]. Moreover, one can apparently state with assurance that any force of turbulence (which reduces naturally to the force of energy transfer, i.e., of nonlinear terms in the hydrodynamic equations) in the two-dimensional case is even less justified than in the three-dimensional case. Indeed, by virtue of the absence of the smoothing effect of viscosity under conditions of two-dimensional turbulent motion (see below), the nonlinear effects predominate during its evolution, and any distortion of the hydrodynamic equations will eventually have a significant effect on the results obtained. By way of example, we will cite the well-known results of [18], where the realization of Millionshchikov's hypothesis results in negative values of the energy density in some finite range of wave numbers. Naturally, one has to eliminate such physically meaningless results by adding new empirical constants and functions to the respective models, which takes these models far beyond the limits of the input exact hydrodynamic equations. On the other hand, the experience in studying common (three-dimensional) turbulence tells us that all of the relevant results were obtained purely hydrodynamically with the additional use of the considerations of dimensionality and similarity; of course, this called for a detailed experimental investigation of turbulent motion. For these reasons, it appears to be of importance that such an approach should be realized in the two-dimensional case as well, although no detailed experimental data (that would be comparable with those for three-dimensional turbulence) are available so far. In what follows, we will discuss the question of how the self-similar spectra of two-dimensional turbulence are related to the laws of conservation and give our considerations concerning the direction of energy and enstrophy flows in the **k** space. Exact relations are written for the third moments, and the correlation is given between these relations, and the parameters pertaining to the ranges of self-similarity. The results are compared with recently obtained experimental data.

2. SELF-SIMILAR SPECTRA OF TWO-DIMENSIONAL TURBULENCE

The description of isotropic turbulence is based on the Karman–Howarth equation (which is an exact corollary of the Navier–Stokes equations) or on an equivalent equation for the energy spectrum,

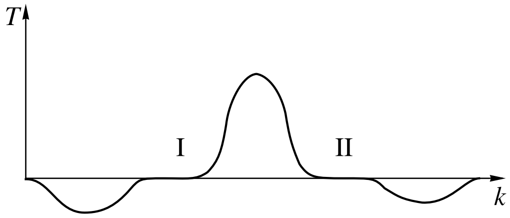
$$\frac{\partial E(k, t)}{\partial t} = T(k, t) - 2\nu k^2 E(k, t), \tag{1}$$

where E is the energy spectrum, $\int E(k, t)dk = \langle \mathbf{v}^2 \rangle/2$ (the fluid density is taken to be equal to unity, and the angle brackets correspond to statistical averaging), $T(k, t)$ is the function characterizing the energy flux in the **k** space and related to the nonlinear terms in the equations of motion, and ν is the viscosity. Equation (1) is valid for the two- and three-dimensional cases, while the difference between the two and three-dimensional problems in this formulation consists in that, in addition to the energy conservation law expressed in the form

$$\int_0^\infty T(k, t)dk = 0, \tag{2}$$

the two-dimensional case is characterized (with zero viscosity) by the law of conservation of enstrophy (mean-square vorticity),

$$\int_0^\infty k^2 T(k, t)dk = 0 \tag{3}$$



Diagrammatic view of the function $T(k)$.

(here we use the universally accepted terminology; what actually happens is that relations (2) and (3) (in the two-dimensional case) hold irrespective of the value of viscosity—at $\nu = 0$, they express the respective laws of conservation). It is obvious from Eq. (2) that $T(k, t)$ as a function of k cannot have one and the same sign for all values of k; in the three-dimensional case, it is arranged as follows: there is some κ such that $T(k, t) \le 0$ at $k \le \kappa$ and $T(k) > 0$ at $k > \kappa$, and it is obvious that

$$\int_0^\infty k^2 T dk > 0. \tag{4}$$

The latter inequality corresponds to the regular concepts of energy transfer from large- to small-scale components, although its proof or a rigorous definition of the conditions at which it is valid is still unknown and represents as complex a problem as the integration of the basic hydrodynamic equations (of course, the reasoning on this subject, based on the concepts of stretching of vortex tubes [19], is not rigorous). It is this form of $T(k)$ that is supported by all known experimental data pertaining both to laboratory turbulence (flows in pipes and channels, decay of turbulence behind the grid, and so on) and to turbulent motions in the atmosphere and in the ocean.

Turning to the two-dimensional case, one can readily see that condition (3) rules out the possibility of such a form of spectrum. Indeed, if we assume that $T(k)$ changes sign at some point κ, we have, by virtue of the theorem of the mean,

$$\begin{gathered}\int_0^\kappa T dk + \int_\kappa^\infty T dk = 0, \\ p^2 \int_0^\kappa T dk + q^2 \int_\kappa^\infty T dk = 0,\end{gathered} \tag{5}$$

where $p < \kappa$ and $q > \kappa$. Because the set of equations (5) is inconsistent in the case of nonzero values of the integrals, such a form of spectrum is impossible, and the function $T(k)$ changes sign more than once. Therefore, in this case, the simplest choice corresponds to a double change of sign of the function $T(k)$; this function is shown diagrammatically in the figure. One can readily see that conditions (2) and (3) are satisfied also by the function symmetric to that shown in the figure relative to the axis k, i.e., $-T(k)$. As in the three-dimensional case, the theoretical choice between $T(k)$ and $-T(k)$ does not appear possible: it may be realized only when analyzing the experimental data.

Next, because solutions of two-dimensional hydrodynamic equations with $\nu \longrightarrow 0$ change to solutions with $\nu = 0$ [2], the last term in the right-hand side of Eq. (1) may be omitted for simplicity. One can readily see that the integrals

$$\varepsilon(z) = \int_0^z T dk, \quad \varepsilon_\omega(z) = \int_0^z k^2 T dk \tag{6}$$

give the variation of the energy and enstrophy, respectively, per unit time in the $[0, z]$ range in the **k** space. In regions I and II identified in the figure, these integrals are independent of z and, therefore, the self-similarity hypothesis may be used for the respective regions of the $E(k)$ spectrum (note that the presence of such horizontal regions, along which $T(k) = 0$, is the necessary condition of the Kolmogorov self-similarity [6]). According to this hypothesis, the $E(k)$ spectrum is defined by the combination of one of the quantities in Eqs. (6) and the wave number k, which has the desired dimension; here, the choice of ε or ε_ω depends on which one of these quantities is dominant. Unlike the conventional approach [20], in the case of which arbitrary assumptions are made as to the end of the inertial interval at which one or another parameter is significant, this choice will be uniquely defined. Of course, one cannot directly compare the quantities ε and ε_ω which have different dimensionalities. Therefore, we will treat the dimensionless relation

$$\varphi = |\varepsilon_\omega|\lambda^2/|\varepsilon|, \tag{7}$$

where λ is the scale of turbulent motion (defined by the behavior of correlation functions over short distances),

$$\lambda^2 = \frac{\int_0^\infty E dk}{\int_0^\infty k^2 E dk};$$

one can see from the equation of motion (1) with $\nu = 0$ and relations (2) and (3) that this quantity is time-independent. It is now obvious from Eq. (7) that, at $\varphi \ll 1$, the spectrum is mainly affected by the quantity ε and, at $\varphi \gg 1$, by ε_ω. It follows from the very form of the function $T(k)$ that it is only

$$\varphi \ll 1 \text{ in region I,}$$

$$\varphi \gg 1 \text{ in region II}$$

which is possible (owing to the presence of the factor k^2 in the integrand for ε_ω).

We make up, as usually, dimensionless combinations containing k and, accordingly, ε and ε_ω to obtain

$$\begin{aligned} E(k) &\propto |\varepsilon|^{2/3}k^{-5/3} \quad \text{in region I,} \\ E(k) &\propto |\varepsilon_\omega|^{2/3}k^{-3} \quad \text{in region II.} \end{aligned} \tag{8}$$

The fluxes of energy and enstrophy have the same direction: for the form of $T(k)$ selected by us, they are both positive in region I (energy and enstrophy are transferred from large to small scales) and negative in region II; one can readily see that this pattern corresponds to the concentration of energy of turbulent motion in the range between regions I and II and to the growth of the $E(k)$ peak in this range. Of course, the change of sign of $T(k)$ would bring about the opposite result; i.e., the energy in the regions left of I and right of II would increase. By virtue of the fact that the former one of these regions is limited on the left by the point $k = 0$, this may bring about the formation of an $E(k)$ peak even in the neighborhood of $k = 0$, so that the inference about the formation of a narrow peak on the $E(k)$ curve, generally speaking, does not depend on whether we take the function $T(k)$ shown in the figure or a function symmetric to $T(k)$ relative to the axis k.

Note that disregard of viscous damping must in fact be expressed by the inequalities

$$2\nu\int k^2 E dk \ll |\varepsilon|, \quad 2\nu\int k^4 E dk \ll |\varepsilon_\omega|,$$

which, for the reasons identified above, may be valid in the two-dimensional case, but are never valid in the three-dimensional case because of the so-called enstrophy catastrophe,

$$\lim_{\nu \to 0} \nu\int_0^\infty k^2 E dk \neq 0;$$

the latter fact makes the energy dissipation the only quantity defining the self-similarity of the inertial region of the spectrum of three-dimensional turbulence.

In conclusion of this section, we will point out that no full experimental proof is available of the existence of the k^{-3} and $k^{-5/3}$ spectra in the two-dimensional case. Evidence for their possible realization may be found in [21], where two-dimensional motion initiated by a magnetic field was investigated. Direct numerical force of two-dimensional turbulence, based on exact hydrodynamic equations, also fails to produce these spectra, at any rate, simultaneously; the reason for this is apparently the inadequacy of the values of the Reynolds number at which the calculations may be realized. On the other hand, these spectra are often realized (simultaneously) in various model calculations in which the respective Reynolds numbers are of the order of 10^7 and higher (see the review of relevant studies in [6]). Therefore, one can state that the presence of the self-similar spectra discussed above in the case of two-dimensional turbulence calls for further serious experimental proof.

3. ANOMALIES OF THIRD MOMENTS OF TWO-DIMENSIONAL TURBULENCE

It is known that a further application of the Kolmogorov local structure hypotheses in the three-dimensional case enables one to obtain, in addition to the spectrum, both qualitative and quantitative characteristics of the second and third moments of the velocity field [5, 6], which agree quite adequately with experimental data. This possibility is associated with the existence of a quantity (the mean value of dissipation of the kinetic energy of turbulent motion) defining the dynamics of decay as a whole. Formally, it is expressed by the fact that the difference of the quantities $\partial B_{LL}/\partial t$ and $(1/2)\partial D_{LL}/\partial t$ appears in the left-hand side of the equation for the structure function of the velocity field D_{LL} (in what follows, standard notation is used: the correlation functions

$$B_{LL}(r) = \langle u_L(\mathbf{x})u_L(\mathbf{x}+\mathbf{r})\rangle,$$

$$B_{LL,L}(r) = \langle u_L(\mathbf{x})u_L(\mathbf{x})u_L(\mathbf{x}+\mathbf{r})\rangle,$$

where u_L is the velocity component along the straight line connecting the points $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$; and the structure functions $D_{LL}(r) = 2[B_{LL}(0) - B_{LL}(r)]$ and $D_{LLL}(r) = 6B_{LL,L}(r)$); according to the Kolmogorov hypothesis, the quantity $(1/2)\partial D_{LL}/\partial t$ in the inertial range is zero. To be more exact, it must be stated [5] that

$$\frac{1}{2}\left|\frac{\partial D_{LL}(r,t)}{\partial t}\right| \ll \left|\frac{\partial B_{LL}(0,t)}{\partial t}\right| = \frac{2}{3}|\varepsilon|. \tag{9}$$

In the two-dimensional case, by virtue of the reasons given above, the value of dissipation cannot define the dynamics of decay, and the possibility of ignoring the dissipation (along with viscosity) results in that the time derivative of D_{LL} still remains in the left-hand side of the respective equation. Consequently, in this situation, we can treat right away the equation for B_{LL}, which has the form [7]

$$\frac{\partial B_{LL}}{\partial t} = \left(\frac{\partial}{\partial r} + \frac{3}{r}\right)B_{LL,L}. \tag{10}$$

Each of two regions of the inertial range (corresponding to the $k^{-5/3}$ and k^{-3}) is characterized by the quantities ε and ε_ω, respectively; we combine these quantities with the quantity r to obtain the same dimension as $\partial B_{LL}/\partial t$ and change over from Eq. (10) to the following equations (C_1 and C_2 are positive dimensionless constants):

$$\begin{aligned} C_1\varepsilon &= \left(\frac{\partial}{\partial r} + \frac{3}{r}\right)B_{LL,L} \quad \text{in region I,} \\ C_2\varepsilon_\omega r^2 &= \left(\frac{\partial}{\partial r} + \frac{3}{r}\right)B_{LL,L} \quad \text{in region II.} \end{aligned} \tag{11}$$

We multiply (11) by r^3 and integrate in view of the condition $B_{LL,L}(0, t) = 0$ to derive

$$\begin{aligned} B_{LL,L} &= \frac{C_1}{4}\varepsilon r \quad \text{in region} \quad \text{I}, \\ B_{LL,L} &= \frac{C_2}{6}\varepsilon_\omega r^3 \quad \text{in region} \quad \text{II}. \end{aligned} \tag{12}$$

So, the dependence of $B_{LL,L}$ on r (as well as $D_{LLL}(r)$) in the region defined by the energy flux reduces to linear, as in the three-dimensional case. In the range of return flux of enstrophy, the respective dependence turns out to be cubic—the result obtained in [22] by a different method. However, unlike the three-dimensional case, it follows from Eqs. (12) that the opposite signs of the third moment correspond to ranges I and II (it will be recalled that the quantities ε and ε_ω have different signs, as follows from Eqs. (2) and (3)). In particular, relations (12) demonstrate that, for the form of $T(k)$ adopted by us, we have $B_{LL,L} < 0$ in region I and $B_{LL,L} > 0$ in region II. Because lower values of r correspond to higher values of the wave vector, the function $B_{LL,L}$ must be positive in some range and becomes negative right of this range; the positive values of $B_{LL,L}$ correspond to the k^{-3} spectrum, and the negative values, to the $k^{-5/3}$ spectrum. This pattern differs markedly from the behavior of the third moment in the three-dimensional case, where it is negative in the entire inertial range.

The above-described anomaly was indeed observed in recent experiments [23] with a thin fluid film passed through a two-dimensional comb. In this case, a spectrum close to k^{-3} corresponds to high positive values of the third moment. However, no spectral region of $k^{-5/3}$ was observed in these investigations, which is apparently due both to the inadequacy of the values of the Reynolds number (of the order of several hundred) and to the fact that the largest scales of length in the k^{-3} range were of the order of the external scale of turbulence.

The foregoing results are closely related to the assumption made as to the pattern of variation of sign of the function $T(k)$. Indeed, if we used the function $-T(k)$ which also satisfies the laws of conservation expressed by Eqs. (2) and (3), we would at any rate obtain $B_{LL,L} < 0$ in the range with the k^{-3} spectrum. Therefore, we must make sure that the choice made (as well as the similarity considerations employed) corresponds to the real situation. Because Belmonte *et al.* [23] obtained data for the third moment for the smallest values of r outside of the inertial range, one can compare these data with D_{LLL} (or, which is the same, $B_{LL,L}$) related to $T(k)$ in the case of small values of r by some exact relation. Before going to derivation of this relation, note that, with the adopted alternation of signs of the function $T(k)$, the inequality

$$\int_0^\infty k^4 T dk < 0$$

is valid by virtue of Eqs. (2) and (3); this inequality will be required below. In order to prove this inequality, consider the function

$$S(z) = \int_0^z T dk,$$

for which we have the following statements from its definition and relations (2) and (3):

$$S(0) = S(\infty) = 0, \quad \int_0^\infty kS dk = 0,$$

$$\int_0^\infty k^4 T dk = -4\int_0^\infty k^3 S dk.$$

One can readily see that S changes its sign only once, from negative to positive; therefore, in accordance with the theorem of the mean, the last integral is positive, so that the statement being proven follows from the last one of the foregoing equalities.

Further, in order to derive the sought relation, one must relate the correlation tensor $B_{lm,n}(r)$ to its Fourier transform expressed in terms of the only scalar function $F_3(k)$ [6],

$$B_{lm,n} = i\int e^{i\mathbf{k}\cdot\mathbf{r}} F_3\left(\frac{\delta_{mn}k_l}{k} + \frac{\delta_{ln}k_m}{k} - \frac{2k_l k_m k_n}{k^3}\right) d\mathbf{k}.$$

We will now use the identity

$$\int e^{i\mathbf{k}\cdot\mathbf{r}} k_l F(k) d\mathbf{k} = -\frac{i\partial}{\partial r_l}\int e^{i\mathbf{k}\cdot\mathbf{r}} F(k) d\mathbf{k},$$

express the tensor $B_{lm,n}$ in terms of the component $B_{LL,L}$ (the respective relation for the two-dimensional case is given in [4]), and perform integration with respect to the angular variable in view of the identity

$$\int_0^\pi e^{ikr\cos\theta} d\theta = \pi J_0(kr)$$

to derive, after relatively simple but cumbersome computations,

$$B_{LL,L} = \frac{1}{2}\int_0^\infty \left[\frac{J_0''}{kr} - \frac{J_0'}{(kr)^2}\right] \frac{T(k)}{k} dk \tag{13}$$

(in the two-dimensional case, $T(k) = 4\pi k^2 F_3$), where the prime corresponds to the derivative of the Bessel function J_0 with respect to its argument. We expand J_0 into a series in powers of kr to derive, from Eq. (13),

$$B_{LL,L} = \frac{1}{2}\int_0^\infty \left[\frac{kr}{8} - \frac{(kr)^3}{96} + \frac{3(kr)^5}{16(4!)^2} - \dots\right]\frac{T(k)}{k}dk,$$

which, because of the presence of two (rather than one, as in three-dimensional hydrodynamics) laws of conservation given by Eqs. (2) and (3), results in that the first nonvanishing expansion term $B_{LL,L}$ (which here is of the fifth order with respect to r) takes the form

$$B_{LL,L} = \frac{r^5}{3\times 2^{11}}\int_0^\infty k^4 T dk. \tag{14}$$

Because the integral in the right-hand side of Eq. (14) is negative, the function $B_{LL,L}(r)$ at small values of r must likewise be negative. The latter fact is indeed observed experimentally [23]. Therefore, the foregoing results both agree with the known experimental data and enable one to use the latter data to find the possible form of the function $T(k)$ characterizing the most important property of turbulent motion, that of energy transfer in the wave number space.

4. CONCLUSION

The foregoing results demonstrate that the notions of the local structure of the velocity field of two-dimensional turbulent motion, based on the Kolmogorov hypotheses, fully agree with the presently available experimental data. However, the direct use of these hypotheses is accompanied by the need to make a certain assumption about the alternation of signs of the function $T(k)$ (in perfect analogy with the three-dimensional case). In view of this, note that the solution of this problem realized above (this problem is not usually discussed in the literature) did not at all appear to be obvious. In particular, in [24], where a detailed numerical force of two-dimensional turbulence was performed, the form of the function $T(k)$, both of the model function and of that obtained as a result of direct numerical solution of hydrodynamic equations, was such that its first extremum was positive. No ranges of self-similarity of the spectrum were observed (neither $k^{-5/3}$ nor k^{-3}), and the return energy flux was naturally observed in the neighborhood of $k = 0$. However, as we made sure above, such a pattern does in no way agree with the recent experimental data. It remains unclear, in particular, why it is the numerical integration of hydrodynamic equations that brings about such a form of $T(k)$: Does this happen because of the inaccuracy of calculations or because of the use of a special form of boundary conditions (periodic) which are not realized experimentally? Therefore, the question remains open whether it is possible in the two-dimensional case to observe corollaries (for example, the negativity of the third moment) of such a form of the function $T(k)$. The very statement of such questions pertaining to two-dimensional turbulence (to which quite definite answers are available in the three-dimensional case) indicates that the relevant results, including those obtained by me, still cannot be regarded as exhaustive. They may come to be exhaustive in the case of comparison with the data of experiments in which, possibly, much higher values of the Reynolds number will be attained than those attained at present.

REFERENCES

1. O. A. Ladyzhenskaya, *Mathematical Problems of Viscous Incompressible Fluid Dynamics* (Nauka, Moscow, 1961).
2. D. G. Ebin and J. Marsden, Ann. Math. **92** (1), 102 (1970).
3. V. I. Arnold (Arnol'd), *Mathematical Methods of Classical Mechanics* (Nauka, Moscow, 1974; Springer-Verlag, New York, 1989).
4. A. L. Tseskis, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 5, 19 (1987).
5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
6. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (Gidrometeoizdat, St. Petersburg, 1996).
7. V. A. Bityurin and A. L. Tseskis, Zh. Éksp. Teor. Fiz. **111**, 528 (1997) [JETP **84**, 289 (1997)].
8. Y. Couder, J. Phys. Lett. **45**, 353 (1984).
9. P. Dmitruk, D. Gomes, *et al.*, Phys. Rev. E **54**, 2555 (1996).
10. H. Brands, S. R. Maassen, and H. J. H. Clercx, Phys. Rev. E **60**, 2864 (1999).
11. L. Onsager, Nuovo Cimento Suppl. **6**, 279 (1949).
12. C. E. Seiler, Y. Salu, *et al.*, Phys. Fluids **18**, 803 (1975).
13. D. R. Nelson and J. M. Kosterlitz, Phys. Rev. Lett. **39**, 1201 (1977).
14. A. L. Tseskis, Zh. Éksp. Teor. Fiz. **83**, 176 (1982) [Sov. Phys. JETP **56**, 95 (1982)].
15. A. L. Tseskis, Prog. Astronaut. Aeronaut. **182**, 183 (1998).
16. A. M. Obukhov, Dokl. Akad. Nauk SSSR **184**, 309 (1969) [Sov. Phys. Dokl. **14**, 32 (1969)].
17. G. F. Carnevale, J. Fluid Mech. **122**, 143 (1982).
18. Y. Ogura, J. Fluid Mech. **16** (1), 33 (1963).
19. G. I. Taylor, Proc. R. Soc. London, Ser. A **164** (918), 15 (1938).
20. G. K. Batchelor, Phys. Fluids Suppl. **12**, 233 (1969).
21. Yu. B. Kolesnikov and A. B. Tsinober, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 4, 146 (1974).
22. G. Falkovich and V. Lebedev, Phys. Rev. E **49**, R1800 (1994).
23. A. Belmonte, W. I. Goldburg, *et al.*, Phys. Fluids **11**, 1196 (1999).
24. J. R. Herring, S. A. Orszag, *et al.*, J. Fluid Mech. **66**, 417 (1974).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 48–56.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 57–67.
Original Russian Text Copyright © 2002 by Danilov, Dovzhenko, Dolzhanskiĭ, Kochina.

Decaying Quasi-Two-Dimensional Turbulence in a Thin Liquid Layer

S. V. Danilov, V. A. Dovzhenko*, F. V. Dolzhanskiĭ, and V. G. Kochina
Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Pyzhevskiĭ per. 3, Moscow, 109017 Russia
**e-mail: fourvort@aport.ru*

Received November 30, 2001

Abstract—The results of laboratory measurements of decaying quasi-two-dimensional turbulence in a thin liquid layer are considered. It is shown that the enstrophy-to-energy ratio decreases according to a power law on a certain decay interval. The exponent in the power law is a function of the Reynolds number. The enstrophy decay is found to be anomalously slow as predicted in a number of numerical studies. It is shown that the anomalous decay in the quasi-two-dimensional flow under investigation is not due to intense vortex formation as in the numerical decaying turbulence, but due to the limited range of scales on which a flow can be regarded as two-dimensional. *© 2002 MAIK “Nauka/Interperiodica”*.

1. INTRODUCTION

It is well known that the energy flux of a 2D turbulent flow induced on small scales in an incompressible liquid is directed to small wave numbers in the spectrum [1]. Steady-state conditions in such a system can be attained only for additional dissipation at small wave numbers since, otherwise, the scale of the spectral energy peak and the energy itself increase. For this reason, a large number of publications are devoted to an analysis of the decay of 2D turbulent flows whose energy remains virtually unchanged in the limit of large Reynolds numbers (although the scale of the energy peak increases).

In accordance with Batchelor’s prediction [2], the enstrophy of a decaying 2D turbulent flow decreases with time in proportion to t^{-2} (in the limit of large Reynolds numbers), while the typical size of vortices increases linearly with time. Numerous numerical experiments [3–9] on decaying 2D turbulence did not confirm this prediction and indicate a much slower decay as a rule. This can be due to the fact that the decay of 2D turbulence is associated with the formation of vortices [10] in which almost the entire enstrophy of the flow is concentrated. These vortices are found to be stable to displacements induced by the background vorticity field, and, hence, filamentation (which is the main process through which the enstrophy of the flow is transferred over small distances) is suppressed in this case. Different scenarios of the decay were proposed in the literature; our task is to determine which scenario takes place in actual practice.

The number of experimental studies devoted to an analysis of decaying 2D turbulence is not large. A series of experiments were made with thin soap films flowing down under the action of the gravity force [11–15]. In these experiments, turbulent perturbations decay downstream, which complicates the investigation of decay regularities. In experiments described in [11, 12, 14, 15], the velocity field fluctuations are measured at a point, and the 3D spectra are calculated by using the Taylor hypothesis. In [16, 17], a turbulent flow is induced by the MHD method in a thin layer of a dilute electrolyte placed in a magnetic field created by a biperiodic set of magnets. In these experiments, as well as in the experiments described in [13], the velocity field is measured simultaneously over the entire cross section of the flow with the help of digital processing of video images of tracer particles placed on the surface of the liquid. The experimentally observed enstrophy decay occurs according to an anomalously slow power law with an exponent of –0.47. In the latter work, it is concluded that the decay follows the scenario predicted by Carnevale *et al.* [4]. According to this scenario based on the results of numerical experiments, not only the energy, but also the turbulence extrema are preserved during the decay due to the vortex formation. If the number of vortices decreases with time in proportion to $t^{-\xi}$, the enstrophy decreases with time, according to [4], following a power law with an exponent equal to $-\xi/2$. It was found that the value of ξ is 0.72.

Turbulence decay is associated with viscosity and, according to the results of recent studies [8, 9], is sensitive to the dissipation mechanism. For this reason, the matching of the results obtained in numerical experiments with hyperviscosity [4] and the results obtained in the laboratory experiments [17] appears unexpected. It is not only due to the fact that hyperviscosity cannot be realized in a laboratory experiment. It is more important that slow horizontal displacements of liquid in thin layers are controlled by friction at the bottom, which is much stronger than conventional dissipation on any scale that can be regarded as two-dimensional.

This research aims at experimental investigation of the decay mechanisms of 2D turbulence induced by the MHD method in a thin layer of an electrolyte in a wide range of the initial Reynolds numbers. We will demonstrate that, indeed, a power segment exists in the enstrophy decay law, but not for all Reynolds numbers. The exponent in the power law depends on the way in which the energy decay is compensated. Our conclusion is that the enstrophy decay is anomalously slow not as a result of intense vortex formation, as in the turbulence decay obtained in numerical experiments, but due to a limited range of scales on which a laboratory flow can be regarded as two-dimensional. In order to demonstrate this, we will compare the experimental results with the results of numerical integration of a quasi-two-dimensional equation of vorticity transformation taking into account the bottom friction.

2. EXPERIMENTAL SETUP AND EXTERNAL PARAMETERS

2.1. Experimental Setup

The experimental setup consists of a rectangular cell having a height of 20 mm and a thin (1 mm) bottom, which is placed in a rectangular biperiodic array formed by round magnets (of diameter 14 mm) with alternating polarities. The cell length is 660 mm. Copper electrodes having a thickness of 10 mm each are arranged along the cell at a distance of 140 mm from one another. The cell was filled with copper sulfate solution of density $\rho = 1.070$ g/cm^3. The depth of the liquid layer h varied from 2 to 6 mm. Current I between the electrode was varied from 0.1 to 3 A. When the electric current is passed in the presence of a space-periodic magnetic field, a vortex lattice with alternating directions of rotation is formed in the liquid. For $I > I_{\text{cr}}$, this stationary vortex structure loses its stability and becomes a flow varying in space and time, which has no initial spatial periodicity. In a certain time, this flow can be regarded as a steady-state or “quasi-stationary” flow in the sense that the average value of energy determined on different time intervals is independent of the length of the interval. With increasing supercriticality, this quasi-stationary flow becomes more and more chaotic. At a certain initial instant $t = t_0$, the current is switched off and the flow velocity starts decaying.

The attainment of the “quasi-stationary” stage of the flow was initially controlled visually and was then confirmed or refined after the processing of experimental results. If necessary, the experiment was repeated with an increased time interval between the current switching on and off. The criterion was the existence of a clearly manifested constant mean value of energy over a sufficiently long time interval on the curve describing the time dependence of energy. Naturally, this dependence itself is pulsating.

2.2. Measuring Technique

The velocity field on the liquid surface was measured with the help of correlation processing of consecutive video images of tracer particles on the liquid surface. We used aluminum powder as a tracer. The working region of the flow had an area of 187 × 140 mm. Clusters of aluminum powder particles make it possible to obtain video images with a good contrast for appropriate illumination and a “black” bottom. Video images were converted into a digital format on a mesh of 720 × 540 pixels. The first of two consecutive frames was divided into square elements with a side of 9 pixels. The mean displacement of each square element upon a transition to the next frame was determined from the minimum of the difference in the intensities, emerging as a result of displacements of squares relative to each other. This allowed us to reconstruct the velocity field on a regular mesh of 76 × 57 points. The step between the points corresponds approximately to 2.5 mm, the approximate number of mesh spacings per magnet being equal to six. Such a resolution is sufficient for formation of vortices at the initial stage of the flow, but leads to certain errors in the calculated values of the velocity field derivatives (including vorticity and divergence) on scales smaller than the pumping scale. As the flow decays, the characteristic sizes of the regions with vorticity of the same polarity increase, and the accuracy of calculated values of vorticity increases. However, the fraction of subpixel displacements increases upon a decrease in the flow velocity field amplitude, which might become a new source of errors. In order to avoid this, we increased the interval between the frames chosen for processing. The initial time interval between the pair of frames being processed was 1/25 s, while the processing was terminated at a time interval equal to the time of energy decay by a factor of e (in experiments with the thickest layer, the final interval was larger than 1 s). This allowed us to extend considerably the dynamic range of measuring the velocity field and energy. For example, in most cases, we were able to trace a decrease in the flow energy by more than three orders of magnitude.

It was found that the divergence of the 2D velocity field differs from zero, but its root-mean-square value is usually smaller than 1/6 of the rms vorticity at the initial stage of the decay and decreases relative to the rms vorticity in the course of the decay of the flow. The relative amplitude of divergence was found to be a good indicator of the accuracy of reconstruction of the velocity field: an increase in the fraction of subpixel displacements primarily affects the growth of this quantity. The relative error of measurements of the velocity field is estimated at 5%, while the relative error of measurements of vorticity and divergence is at least twice as large. The smallness of the divergence leads to the rough assumption that the 2D flow is incompressible.

The flows induced in thin layers are slow (the Froude number is small) and quasi-two-dimensional (the velocity field is close to horizontal). The vorticity field on the free surface of such a flow can be approximately described by the equation [18, 19]

$$\partial\zeta + [\psi, \zeta] = -\lambda\zeta + \nu\Delta\zeta, \tag{1}$$

where ψ is the stream function on the surface of the flow, $\zeta = \Delta\psi$ is the vorticity, $[\psi, \zeta]$ is the Jacobian of ψ and ζ, ν is the kinematic viscosity, and λ is the bottom friction coefficient. The term containing the bottom friction coefficient is the result of parametrization of the effect of the bottom; its introduction can be justified by the fact the flow must be locally close to a Poiseuille flow. In this case, the bottom friction coefficient can be estimated as $\lambda = 2\nu\kappa/h^2$, where κ is the fitting parameter of the order of unity. This parameter is to be measured [18, 19]. For a forced flow, its value depends on the magnetic field distribution over the height. For decaying 2D flows under the conditions when viscosity dominates and the velocity field is two-dimensional (Stokes flow), the value of coefficient λ must coincide with the decrement of the first Stokes mode, which gives $\kappa = \pi^2/8$ [16, 17, 20]. In actual practice, weak vertical displacements are always present, and parameter κ has a larger value.

The flow in a thin layer is characterized by two Reynolds numbers [21] corresponding to two types of dissipation on the right-hand side of Eq. (1). We introduce the standard Reynolds number $\mathrm{Re} = UL/\nu$, where U is the characteristic velocity and L is the characteristic size. We assume that $U = E^{1/2}$ and $L = (E/Z)^{1/2}$, whence $\mathrm{Re} = E/Z^{1/2}\nu$. Here, E and Z are the energy and the enstrophy per unit mass. The assumed value of the size is equal to the wavelength divided by 2π; consequently, the initial size in our case must exceed the size of a magnet, divided by $\pi\sqrt{2}$ (or 0.31 cm). Since the maximum rms velocity attainable in the present experiment amounts to a few centimeters per second, the maximum initial Reynolds number $\mathrm{Re} \approx 100$. The Reynolds number determined from the size of the working region attains values up to 3500, which is larger than in the experiments described in [16, 17].

The second Reynolds number corresponds to bottom friction and can be defined as $\mathrm{Re}_\lambda = U/L\lambda = Z^{1/2}/\lambda$ [18, 19]. Since dissipation due to bottom friction in quasi-two-dimensional flows is always larger, it defines the true supercriticality of the flow. A detailed discussion of the applicability of the Reynolds number to the bottom friction in various real quasi-two-dimensional flows for moderate values of supercriticality is presented in [19]. In [22–24], Eq. (1) was successfully verified directly in experiments. However, the validity and peculiarities of application of the bottom friction approximation for flows with a complex profile of exciting force [25, 26], time-varying flows [27], and, the more so, turbulent flows [28, 29] are still to be investigated.

Since most theoretical results are formulated in terms of the standard Reynolds number, we will use this criterion for describing experimental flows, taking into account the fact that the inner and outer Reynolds numbers are connected through the relation

$$\mathrm{Re} \propto \frac{E}{Zh^2}\mathrm{Re}_\lambda.$$

3. MEASUREMENTS OF THE BOTTOM FRICTION COEFFICIENT

Measurements were made at depths h = 2, 3, 4, 5, and 6 mm for currents I = 100, 200, 300, 400, 500, and 600 mA. At small depths, conventional instrumental methods for ensuring the horizontality of the cell bottom are insufficiently accurate; for this reason, exact fitting was carried out through an analysis of the flow parameters at the opposite ends of the cell (for this reason, measurements were not made, in particular, at a depth of 1 mm). For the depth used, the uniformity of the velocity field in the entire cell was reliably ensured. All the flows were supercritical.

The method of determining the value of λ was based on the balance equation for kinetic energy,

$$E = \int(\nabla\psi)^2 dxdy,$$

which, in accordance with Eq. (1), can be written in the form

$$\frac{dE}{dt} = -2\lambda E - 2\nu Z \tag{2}$$

for a region with rigid boundaries. In our case, the boundaries of the working region are open, but the fluxes through them are very small, which was verified directly.

It follows hence that the value of λ can be determined from the logarithmic derivative of energy with respect to time:

$$2\lambda = -\frac{d\ln E}{dt} - \frac{2\nu Z}{E}. \tag{3}$$

The second term on the right-hand side of this equation makes a noticeable contribution only for large depths of the liquid layer. For depths h = 2, 3, 4, 5, and 6 mm, we obtain the following mean values of λ: 1.09, 0.49, 0.25, 0.165, and 0.105 s^{-1}, respectively. This gives the value of the fitting parameter $\kappa \approx 1.7 \pm 0.1$ on the average in all experiments (it should be noted, however, that the fitting coefficient decreases with the depth). The experimentally determined value noticeably differs from the theoretical value of $\kappa = \pi^2/8$ determined from the decrement of the first Stokes mode (for a strictly horizontal motion). This discrepancy can apparently be attributed to the effect of vertical displacements which were disregarded. Although the 3D nature of the flow is still manifested, the flow structure is such that, for not very high velocities (current Reynolds numbers), Eq. (1) holds if we assume the experimental value for κ.

Figure 1 shows the time dependences (a) of energy for three realizations of decay and (b) of individual components of Eq. (3) for the first of these realizations (the flow was excited by a current of 1.5 A, the initial value of the Reynolds number was ≈ 70, and the liquid layer depth was $h = 5$ mm). The time from the beginning of decay is laid along the abscissa axis. The three realizations shown in Fig. 1a differ insignificantly, and the measurements corresponding to each realization lead to close values of the bottom friction coefficient. The maximum difference does not exceed 4%. This also applies to measurements made for other values of the initial Reynolds number. Figure 1b shows that, after the initial period of adjustment, the flow indeed attains conditions which are approximately characterized by a constant value of the bottom friction coefficient. The deviations in the logarithmic derivative of energy on the initial segment indicate that the overall dynamics of the flow includes 3D effects. The adaptation is associated with a rearrangement of the vertical structure of the flow. This leads to an enhanced dissipation and to a considerable decrease in the Reynolds number.

4. CHARACTERISTICS OF DECAYING TURBULENCE

Most of measurements for large initial Reynolds numbers were made for a layer of liquid with $h = 5$ mm. We will consider below only such cases. The measurements were carried out for the initial current varying from 0.5 to 3 A with a step of 0.5 A.

The energy of the experimental flow decreases with time due to the effect of the bottom friction. The decrease in enstrophy is associated not only with the effect of bottom friction, but also with the rearrangement of the flow structure. In order to single out the effects associated with the rearrangement of the flow structure, we can use the following two approaches. The first is associated with the analysis of the decrease in the ratio Z/E. The second is based on the introduction of an exponentially compressed element of time

$$d\tau = \exp[-\lambda(t-t_0)]dt$$

or

$$\tau = \frac{1-\exp[-\lambda(t-t_0)]}{\lambda},$$

as well as the compensated fields

$$(\psi', \zeta') = (\psi, \zeta)\exp[\lambda(t-t_0)].$$

Here, t_0 is the time corresponding to the onset of decay. In the new variables, the term responsible for the bottom friction vanishes, and the equation for the vorticity assumes the form

$$\partial_\tau \zeta' + [\psi', \zeta] = \nu'\Delta\zeta'.$$

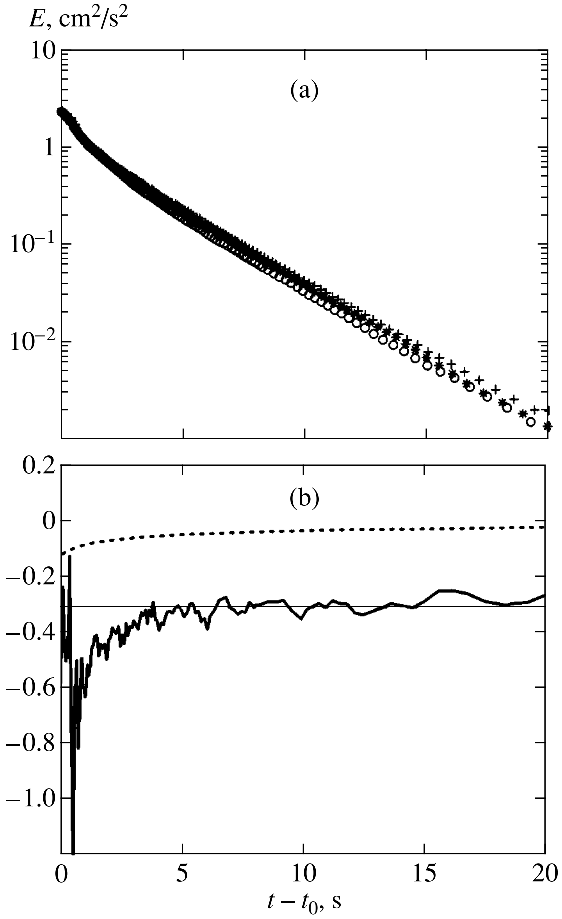


Fig. 1. (a) Energy decay in experiments with Re(0) $\approx$ 70 (a current of 1.5 A, three realizations) and $h = 5$ mm and (b) the contributions to energy decay for a realization (○): $d(\ln E)/dt + 2\nu Z/E$ (broken line), -2λ (solid straight line), and $-2\nu Z/E$ (dotted line). The values on the ordinate axis in (b) are in inverse seconds.

This equation is characterized by an exponentially increasing kinematic viscosity $\nu' = \nu\exp[\lambda(t-t_0)]$. In the new variables, decay occurs during a finite “condensed time.” Hansen *et al.* [17] mentioned that it is the analysis in these coordinates that leads to matching with the theory described in [4]. For large times, the compensated enstrophy and the ratio Z/E asymptotically follow the same law relative to the conventional time.

Figure 2 shows the results of processing of experimental data for three values of the initial current (0.5, 1.5, and 3 A). In Fig. 2a, the compensated enstrophy is presented as a function of compressed time. It can be seen that, as the initial current increases (the initial amplitude of the velocity field does not increase linearly with current since the flow is under strongly supercritical conditions), a power segment is formed on the decay curve. It is completely absent for lower values of the initial current and always vanishes for a large time interval. The straight line drawn in the figure has a

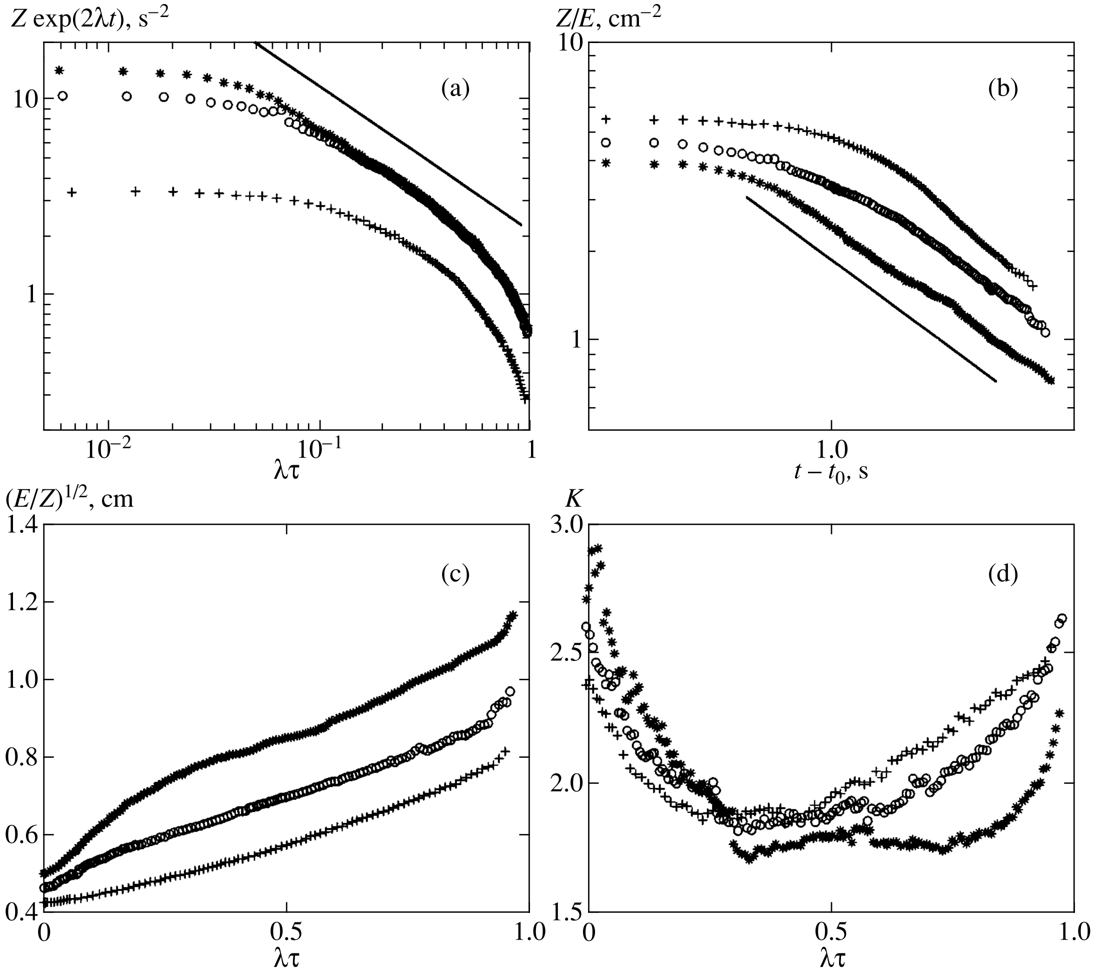


Fig. 2. (a) Compensated enstrophy $Z' = Z\exp(2\lambda t)$, (b) enstrophy-to-energy ratio, (c) characteristic size $l = (E/Z)^{1/2}$, and (d) excess as functions of time in decay realizations with different initial Reynolds numbers. Symbols + corresponds to the realization induced by a current of 0.5; ○, to 1.0; and ∗, to 3.0 A; the slope of the straight line is –0.75 (a) and –0.4 (b).

slope of –0.75. Thus, we can speak of the presence of a power segment on the compensated enstrophy vs. compressed time curve only conditionally to a certain extent. Thus, the conclusion drawn in [17] (the dynamic range of decay in our experiments is an order of magnitude larger) on the coincidence of the decay law with the predictions of the theory described in [4] is not confirmed.

It can be seen from Fig. 2b that the ratio Z/E as a function of conventional time follows a power law to a larger extent. However, here also the decay cannot be characterized by a single exponential for all cases. The straight line in this figure has a slope of –0.4. Such a strong difference between the two slopes is due to the following two factors. First, the segment with a power behavior of enstrophy begins in the region where the energy has not attained the asymptotic mode of decrease (see Fig. 1b). Second, the exponential compression of time in the former case must obviously lead to an increase in the steepness of the decay curve.

The above-mentioned decay law is close to that obtained in [4], but this does not indicate that our experiments confirm the theory formulated therein. Indeed, during the decay, the experimental vorticity field does not follow the scenarios observed in numerical experiments. Figures 3 and 4 show the velocity field and the vorticity field for the flow excited by a current of 1.5 A just before switching off the current, 1 s after the switching off, and after the time interval $1/\lambda$ following the switching off. The loops in Fig. 4 are drawn through the levels $\pm\zeta_{rms}$ and $\pm2\zeta_{rms}$. Both figures show that vortex structures increase in size during the decay, but the decay is not accompanied by the formation of small-scale structures and vorticity field filamentation. The vorticity field distribution is smooth, and the regions in which the vorticity is modulo larger than double the root-mean-square value are virtually absent. It can be seen that vorticity of the same polarity is indeed combined into coarse regions, but structures similar to vortices in the numerical 2D turbulence are not formed.

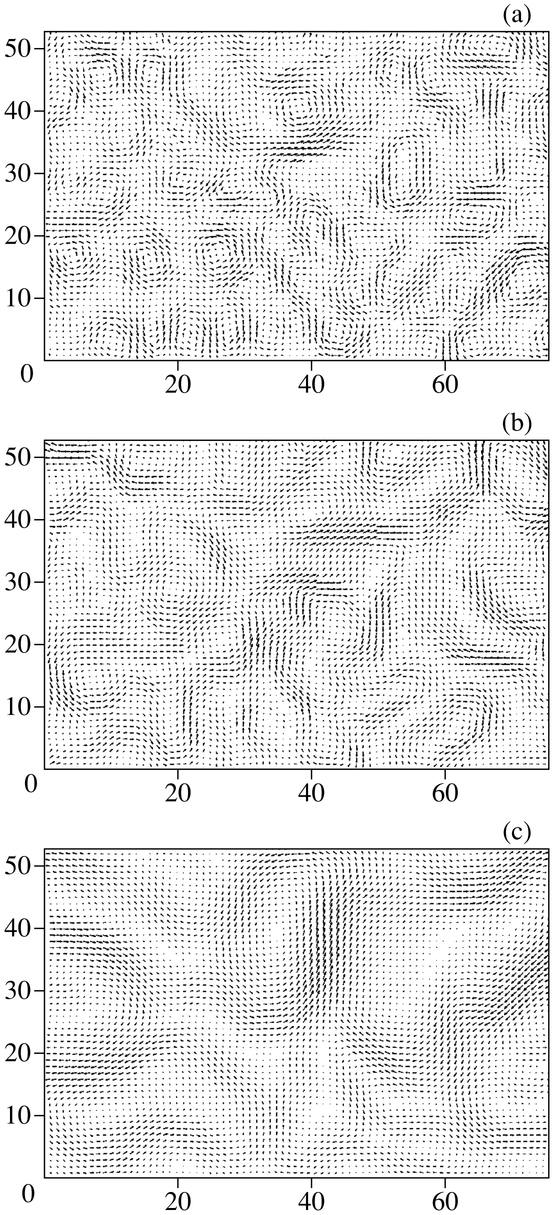


Fig. 3. Velocity field in the realization excited by a current of 1.5 A (a) at the instant of current switching off, (b) 1 s after switching off, and (c) $1/\lambda \approx 6$ s after switching off. The length of the vector is normalized to the maximum length at a given instant.

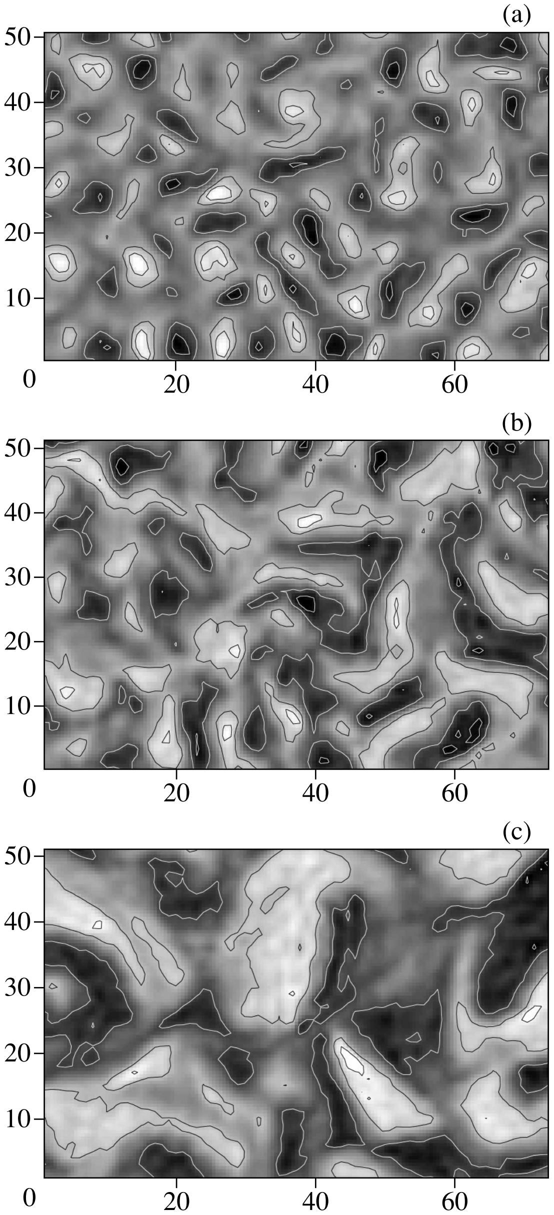


Fig. 4. Vorticity distribution corresponding to realizations of the velocity field presented in Fig. 3. Dark regions correspond to negative vorticity, and light regions, to positive vorticity. The loops are drawn for levels $\pm\zeta_{rms}$ and $\pm 2\zeta_{rms}$.

The boundaries of the regions become uneven in the course of decay.

The mean size of the vortex structures defined as $l = (E/Z)^{1/2}$ is presented in Fig. 2c. During the total decay time, this size increases approximately by a factor of two. By definition, the mean size exhibits a power behavior in the same cases as the enstrophy-to-energy ratio. Such a behavior is manifested weakly when compressed time is used. In [17], the characteristic size is determined from structures in the vorticity field. Hansen *et al.* [17] state that this size obeys a power law compatible with the decay law for enstrophy in the sense of the theory [4]. Since the vorticity field does not contain clearly outlined vortex structures, any determination of the size is inaccurate, and such an analysis will not be carried out here.

Figure 2d shows the behavior of the excess (kurtosis) of the vorticity field,

$$K = \frac{S\int \xi^4 dxdy}{\left(\int \xi^2 dxdy\right)^2},$$

where S is the area of the working region.

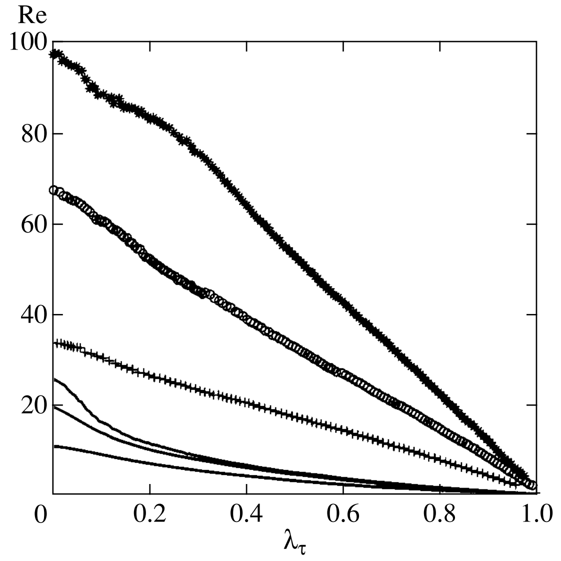


Fig. 5. Behavior of Reynolds numbers in realizations induced by a current of 0.5 (∗), 1.5 (○), and 3 A (+). Solid curves correspond to Re_λ for these realizations.

In all cases the excess behaves similarly and never exceeds the Gaussian value of 3 (the same fact also takes place in experiments [17]). The value of the excess amounts to several tens in typical numerical experiments on the 2D turbulence decay, while large values of excess are due to the presence of well-defined extrema in the vorticity field, which considerably exceed the root-mean-square values of vorticity (see, for example, [3]). Small values of excess in laboratory experiments indicate a very strong suppression of vorticity dynamics due to dissipative effects and necessitate much larger initial Reynolds numbers for attaining conditions similar to those in numerical experiments.

Figure 5 shows the time dependences of the Reynolds numbers during the decay. According to the results obtained by Chasnov [9], the decay of a 2D turbulent flow must be accompanied by an increase in the current Reynolds number due to rapid decay of enstrophy if the initial Reynolds number exceeds the critical value approximately equal to $15.7/\sqrt{2}$. This is not observed in laboratory experiments in view of additional decay introduced by bottom friction as well as the decay at the initial stage, which is associated with 3D effects. At the same time, this leads to a decrease in conventional Reynolds numbers, while the Reynolds numbers corresponding to bottom friction (solid curves in Fig. 5) are found to moderate.

Under the experimental conditions, the advection of the vorticity field has no time to cause the formation of thin extended vortex structures responsible for the enstrophy dissipation in the standard decaying turbulence. Accordingly, enstrophy decays anomalously slowly relative to the theoretical and numerical predictions. It was proved by Chasnov [9] that the Batchelor decay law in 2D turbulence with an ordinary (and not hyper) viscosity must be realized only for the critical Reynolds number, while the exponent in the decay law increases to –0.8 as the Reynolds number tends to infinity. If we follow this result formally, we should expect a steeper decay law for the experimental system under investigation, which is not observed in actual practice.

Although the velocity fields for the flows under investigation can be generally assumed to be two-dimensional (the ratio of divergence to the curl of the velocity field at the surface does not exceed 1/6 on the average), 3D effects cannot be neglected in an analysis of the experimental results. On scales smaller than the depth of the liquid and for times of variation shorter than $1/\lambda$, the rearrangement of the flow is of the 3D type. It was found that additional dissipation introduced by such a 3D adaptation at the initial stage of the decay leads to anomalous behavior of the enstrophy decay in the laboratory experiment. In order to prove this, we carried out numerical experiments for studying the decay of the flows which strictly obey Eq. (1) and whose parameters (the initial spectrum and Reynolds numbers) correspond to the experimental values.

5. NUMERICAL EXPERIMENTS

Numerical experiments were carried out by using the pseudospectral method. We integrated Eq. (1) for vorticity with bottom friction. The size of the integration region was $2\pi \times 2\pi$; calculations were made on a mesh with 128×128 points. Such a small resolution is sufficient since the initial Reynolds numbers are comparatively small and decrease most of the time during the evolution of the flow. The initial realization of the vorticity field was chosen so that the 3D energy spectrum of the flow was concentrated in the vicinity of the mode $(k) = (k_x, k_y) = (N, N)$, where $N = 5$, which approximately corresponds to the experimental conditions. The initial amplitude of the vorticity field was chosen so that the initial energy was equal to unity. Then, we calculated the initial enstrophy of the flow. Further, we chose the kinematic viscosity from the condition that the Reynolds number $\mathrm{Re} = E/Z^{1/2}\nu$ is equal to the preset value. The value of coefficient λ is determined by the depth of the layer of liquid. In order to preserve similarity with the laboratory experiment, we chose

$$\lambda = \frac{2\nu\kappa}{h^2} = \frac{2\nu k}{(\alpha d)^2},$$

where α is the ratio of the thickness of the liquid layer to the diameter of the magnet, which determines the initial size of the vortex. For the laboratory experiment, this ratio amounts to 0.33. The role of the initial size of vortices in the numerical experiment is played by the quantity π/N. Numerical calculations were made for several values of the Reynolds number in the interval from 50 to 500. In each case, integration was carried out over the time $3/\lambda$ during which the velocity in the liquid had decayed considerably.

A considerable difference between the numerical flow obeying Eq. (1) and its experimental "image" is that the logarithmic derivative of energy for the former flow exactly coincides with the combination $-2\lambda - 2\nu Z/E$. This leads to a noticeable discrepancy in the dynamics at the initial stage of the decay. The numerical flow decays considerably more slowly, and nonlinear interactions play a considerably more significant role at this stage. In should be recalled that, in the laboratory flow, the velocity field experiences a 3D adaptation due to which the true energy dissipation rate has a larger absolute value than $-2\lambda - 2\nu Z/E$.

For this reason, a considerably stronger filamentation of the background vorticity is observed at the initial stage of the decay in the numerical flow, on the one hand, and the vorticity is concentrated into vortices in the vicinity of local extrema, on the other hand. The latter is manifested in a certain increase in the excess in the case corresponding to a smaller Reynolds number and in its significant growth in the case of the largest Reynolds number (see Fig. 6a). Figure 6b shows Reynolds numbers as functions of time for the corresponding realizations. In the experiment with the initial Reynolds number Re = 50, the excess decay is observed all the time, as well as in the laboratory experiments. In the case with Reynolds number Re = 100, a tendency to the formation of a plateau is observed on the time interval $\lambda t < 0.5$. Finally, in the case with the initial Reynolds number equal to 200, the instantaneous Reynolds number increases for $\lambda t < 0.5$. A comparison with Fig. 6a shows that these time intervals coincide with the intervals on which the excess grows. It is only on these intervals that the behavior of decaying turbulence in a flow with bottom friction resembles the typical behavior of the numerical decaying turbulence without bottom friction. At subsequent stages, the flow dynamics is mainly determined by the decay, and nonlinear processes play an insignificant role.

Figure 7 shows the enstrophy-to-energy ratio as a function of dimensionless time λt. These curves contain power segments, but their slope increases with the Reynolds number (the slope is approximately equal to 1.0 for $\mathrm{Re}_\lambda(0) = 200$). This change in the slope indicates that the dissipation of enstrophy is suppressed in the case of smaller Reynolds numbers. This is due to the fact that mixing and filamentation of the vorticity field for such Reynolds numbers have no time to occur yet. Thus, the initial Reynolds numbers must be larger than 200 for the pattern familiar from numerical experiments without bottom friction to be observed at a certain initial segment of the decay. The latter statement applies the more so to the laboratory experiments in which additional energy decay is observed at the initial stage.

We also measured the time evolution of the average number of vortices and their mean size. We apply the term vortex to the part of the vorticity field in which it exceeds double the rms value in amplitude. In the course of decay, these characteristics approximately follow power laws, but exponents are functions of the Reynolds numbers. For example, the exponent in the

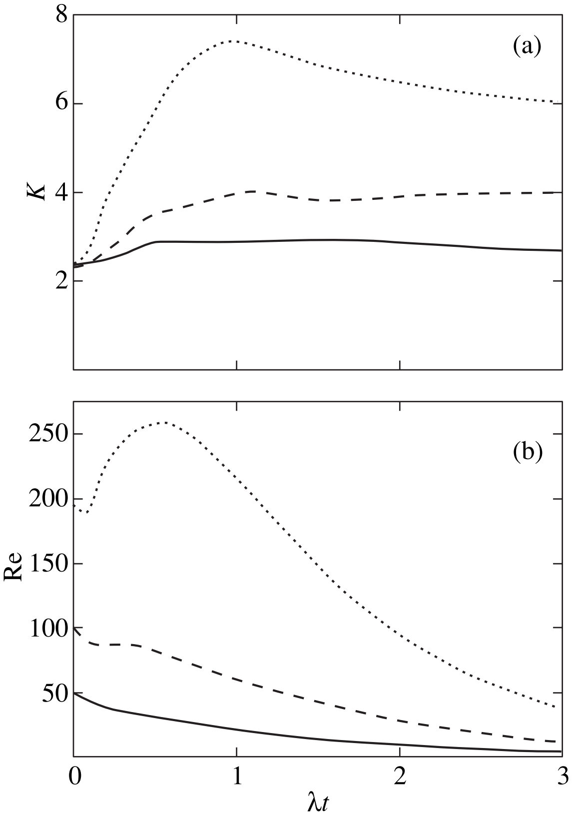


Fig. 6. (a) Excess for the vorticity field and (b) Reynolds numbers in realizations of numerical experiments with the initial Reynolds number equal to 50 (solid curve), 100 (dashed curve), and 200 (dotted curve) as functions of time.

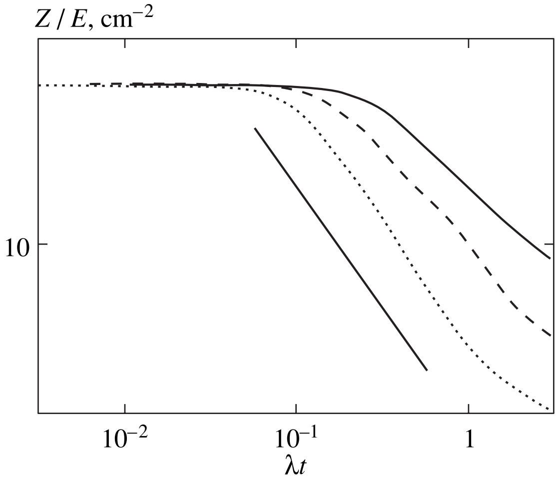


Fig. 7. Enstrophy-to-energy ratio as a function of the decay time in the same cases as in Fig. 6. The slope of the solid line is equal to 1.

decay law for the number of vortices changes from –0.5 to –0.9, but it is hardly expedient to compare these laws with the predictions of the theory [4] in view of different behavior of the excess.

Thus, the decay of the "numerical" turbulence obeying the vorticity equation with bottom friction differs from the decay of the laboratory turbulence and both phenomena differ from the pattern predicted in [4] for the Reynolds numbers under investigation.

6. CONCLUSIONS

Decaying quasi-two-dimensional turbulence in the laboratory flow exhibits some features predicted theoretically for decaying 2D turbulence, such as an increase in the size of turbulence region and an exponential decay of enstrophy. At the same time, bottom friction introduces its own peculiarities, and true dynamics strongly differs from theoretical predictions in spite of the conclusions drawn in [17]. Besides, according to the results of our numerical experiments, decaying turbulence in thin layers differs from the laboratory quasi-two-dimensional flow. This is due to the fact that Rayleigh friction for a strong supercriticality is just a convenient parametrization of the bottom effect. True dynamics in the laboratory flow is associated with 3D movements and 3D adaptation leading to an additional decay.

Bottom friction is an inseparable component of virtually all quasi-two-dimensional systems in nature and in the laboratory (it takes place even for soap films due to air friction [12]). It is proved in [19] that, on all scales over which the motion can be regarded as two-dimensional, bottom friction prevails over conventional friction. Over a small scale, 3D flows associated with the effect of the bottom, cannot be controlled by the conventional kinematic viscosity alone. The effect of friction against the bottom for such flows is also significant. For this reason, experimental quasi-two-dimensional systems not do necessarily follow the theory disregarding true dissipation. The problem is to construct a theory correctly describing the experimental results and based on the effect of the main factor reducing the system to the 2D case rather than to make attempts to match the behavior of flows mainly determined by the bottom friction to 2D hydrodynamics (see, for example, [16, 17, 30]).

ACKNOWLEDGMENTS

We are grateful to O.I. Karpilova for her help in carrying out experiments and in preparing the manuscript for publication.

This work was financed by the Russian Foundation for Basic Research (project nos. 99-05-64350 and 99-05-64351) and the federal program on "Investigation of Turbulence and Structure Formation in Layered Nonlinear Wave Systems."

REFERENCES

1. R. H. Kraichnan, Phys. Fluids **10**, 1417 (1967).
2. G. K. Batchelor, Phys. Fluids **12**, 233 (1969).
3. J. C. McWilliams, J. Fluid Mech. **146**, 21 (1984).
4. G. F. Carnevale, J. C. McWilliams, Y. Pomeau, *et al.*, Phys. Rev. Lett. **66**, 2735 (1991).
5. J. B. Weiss and J. C. McWilliams, Phys. Fluids A **5**, 608 (1993).
6. R. Benzi, S. Partanello, and P. Santangelo, J. Phys. A **21**, 1221 (1988).
7. R. Benzi, M. Colella, M. Briscolini, and P. Santangelo, Phys. Fluids A **4**, 1036 (1992).
8. P. Bartello and T. Warn, J. Fluid Mech. **326**, 357 (1996).
9. J. R. Chasnov, Phys. Fluids **9**, 171 (1997).
10. J. C. McWilliams, J. Fluid Mech. **219**, 361 (1990).
11. H. Kellay, X. L. Wu, and W. I. Goldburg, Phys. Rev. Lett. **74**, 3975 (1995).
12. B. K. Martin, X. L. Wu, W. I. Goldburg, and M. A. Rutgers, Phys. Rev. Lett. **80**, 3964 (1998).
13. M. Rivera, P. Vorobieff, and R. E. Recke, Phys. Rev. Lett. **81**, 1417 (1998).
14. A. Belmonte, W. I. Goldburg, H. Kellay, *et al.*, Phys. Fluids **11**, 1196 (1999).
15. H. Kellay, C. H. Bruneau, and X. L. Wu, Phys. Rev. Lett. **84**, 1696 (2000).
16. O. Cardoso, D. Marteau, and P. Tabeling, Phys. Rev. E **49**, 454 (1994).
17. A. E. Hansen, D. Marteau, and P. Tabeling, Phys. Rev. E **58**, 7261 (1998).
18. E. B. Gledzer, F. V. Dolzhanskiĭ, and A. M. Obukhov, *Hydrodynamic-Type Systems and Their Application* (Nauka, Moscow, 1981), p. 368.
19. F. V. Dolzhanskiĭ, V. A. Krymov, and D. Yu. Manin, Usp. Fiz. Nauk **160** (7), 1 (1990) [Sov. Phys. Usp. **33**, 495 (1990)].
20. B. Jüttner, D. Marteau, P. Tabeling, and A. Thess, Phys. Rev. E **55**, 5479 (1997).
21. F. V. Dolzhanskiĭ, Fiz. Atmos. Okeana **23**, 348 (1987).
22. F. V. Dolzhanskiĭ, V. A. Krymov, and D. Yu. Manin, J. Fluid Mech. **241**, 705 (1992).
23. F. V. Dolzhanskiĭ, V. A. Krymov, and D. Yu. Manin, Russ. J. Comp. Mech. **1**, 107 (1993).
24. V. A. Krymov and D. Yu. Manin, Fiz. Atmos. Okeana **28**, 129 (1992).
25. S. D. Danilov and V. A. Dovzhenko, Fiz. Atmos. Okeana **31**, 621 (1995).
26. S. D. Danilov and V. G. Kochina, Fiz. Atmos. Okeana **33**, 845 (1997).
27. S. D. Danilov, F. V. Dolzhanskii, V. A. Dovzhenko, and V. A. Krymov, Chaos **6**, 297 (1996).
28. S. D. Danilov and F. V. Dolzhanskiĭ, Fiz. Atmos. Okeana **36**, 35 (2000).
29. S. D. Danilov, V. A. Dovzhenko, and V. A. Krymov, Fiz. Atmos. Okeana **36**, 284 (2000).
30. D. Marteau, O. Cardoso, and P. Tabeling, Phys. Rev. E **51**, 5124 (1995).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 5–10.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 10–16.
Original Russian Text Copyright © 2002 by Alexeyev, Alexeyeva.

**GRAVITATION,
ASTROPHYSICS**

Twenty Years of Galactic Observations in Searching for Bursts of Collapse Neutrinos with the Baksan Underground Scintillation Telescope

E. N. Alexeyev* and L. N. Alexeyeva

Institute for Nuclear Research, Russian Academy of Sciences,
pr. Shestidesyatiletiya Oktyabrya 7a, Moscow, 117312 Russia
**e-mail: alexeyev@ms1.inr.ac.ru*

Received January 29, 2002

Abstract—The results of twenty-year-long Galactic observations in neutrino radiation are summarized. Except for the recording of a neutrino signal from the supernova SN 1987A, no Galactic bursts of collapse neutrinos have been detected. An upper bound on the mean frequency of gravitational collapses in our Galaxy was obtained, f_{collapse} (at 90% confidence) < 0.13 yr^{-1}. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In 1933, Wolfgang Pauli introduced a new neutral particle of low or zero mass to save the law of conservation of energy in nuclear beta decays [1]. This particle was experimentally observed only twenty-five years later by Reines and Cowan [2]. The interactions of neutrinos and their role in particle physics, astrophysics, and cosmology have been studied with an increasing intensity ever since. The discovered particle turned out to be so amazing that it allowed one not only to study nuclear processes on Earth and in its atmosphere but also to look into stellar objects, because it is highly penetrating due to the weakness of its interactions with matter.

Progress in theoretical and experimental research has led to the development of many detectors designed to search for and record terrestrial and extraterrestrial neutrinos. Since the energy sources in stars are nuclear reactions, they also emit neutrinos. A new line of research emerged—neutrino astronomy. However, a constant flux of low-energy neutrinos is very difficult to detect, as evidenced by long-term experiments to study neutrino fluxes from the Sun, the star closest to the Earth [3–5].

Back in 1934, Baade and Zwicky [6] suggested the existence of neutron stars and came up with the idea that these are formed during supernova explosions. Thirty years later, in 1965, Zel'dovich and Gusseĭnov [7] concluded that a short burst of high-energy neutrinos could arise when matter is neutronized in the gravitational core collapse of a massive star. In the same year, Domogatsky and Zatsepin [8] pointed out an experimental possibility of searching for such neutrino bursts. In 1966, Colgate and White [9] surmised that neutrinos could play a crucial role in supernova explosions. Almost concurrently, the discovery of pulsars in the Crab Nebula [10], which is the supernova remnant observed by Chinese astronomers in 1054, and in Vela [11] late in the fall of 1968 provided evidence for the formation of neutron stars during supernova explosions. Subsequent observations and data analysis confirmed this conclusion and gave an insight into the physical processes that underlie these phenomena [12, 13].

In the succeeding years, both a theory for the final stages of stellar evolution and experiments to search for bursts of such collapse neutrinos were intensively developed. The explosion of the supernova SN 1987A on February 23, 1987, was a special milestone in the development of the two lines of research.

The first experiments to search for neutrino bursts from gravitational stellar core collapses were begun in 1973 by the Pennsylvania–Texas–Turin collaboration. These were three small water Cherenkov facilities with target masses of about 20 t each and with particle detection energy thresholds of 20 MeV: the first was at the Homestake Mine at a depth of 4400 m [14], the second was at a mine in Ohayo at a depth of 1800 m [15], and the third was in the road tunnel under Mont Blanc at a depth of 4270 m [16]. They did not work long, only a few years. Although no expected burst of collapse neutrinos was detected, they made the first step on the long way to 1987.

The Baksan underground scintillation telescope (Northern Caucasus, Russia) [17], the LSD scintillation facility under Mont Blanc (Italy) [18], the scintillation detector at the salt mine in the town of Artemovsk (Ukraine) [18], and the water Cherenkov detectors Kamiokande II (Japan) [19] and IMB (USA) [20] belong to the second generation of collapse-neutrino detectors with target masses of several hundred tons, which began their operation in 1977–1980. The last two detectors were specially designed to search for proton decays, but they proved to be incapable of recording collapse neutrinos.

Before 1987, the theoretical models for the late evolutionary stages of massive stars were one-dimensional calculations of spherically symmetric nonmagnetic nonrotating configurations. The basic neutrino-radiation parameters expected during the gravitational core collapse of a massive star and the cooling of a newborn star were formulated: the total neutrino-radiation energy is (2–5) × 10^{53} erg, the mean electron-neutrino energy is 8–12 MeV, the fraction of electron antineutrinos in the total radiation is 0.16–0.25, and the burst duration is 10–20 s [21–24].

The recording of a neutrino signal from a type-II supernova exploded in the Large Magellanic Cloud, a neighboring galaxy, by three facilities, Kamiokande II [25], IMB [26], and the Baksan telescope [27], on February 23, 1987, first confirmed the basic ideas of the general theoretical pattern of gravitational collapse with the formation of a neutron star. At the same time, however, it raised many questions both for theory and for experiment.

Observations of the expanding envelope of the supernova remnant show that the explosion was asymmetric, with an unusual shape of the remnant, with matter mixing [28, 29], with an as yet undiscovered pulsar [30] or with a candidate for a pulsar with unusual properties [31, 32]. All of this led to rapid progress in the theoretical modeling of processes inside stars and in understanding the many processes in core collapse and in the stellar envelope, as well as to a complication of the collapse pattern [33–35], particularly when rotation and strong magnetic fields are included in the calculations [36]. However, not all of the questions have been answered as yet. In particular, the stellar-envelope ejection mechanism and, what is important for experimenters, the complete temporal structure of the neutrino luminosity are still problems to be solved. In addition, the general-relativity [37] and weak-magnetism [38] effects, which have not yet been included in the calculations and which may lead to a higher neutrino luminosity, as well as a possible nonzero neutrino mass, which would cause the neutrino spectra to change [39], are being discussed.

The development of methods for analyzing experimental data with small statistics and using them to study the neutrino signals recorded on February 23, 1987, by the three facilities also led us to conclude that the neutrino radiation has a more complex temporal structure [40].

An important lesson drawn from the history of experimental observation of the neutrino signal from SN 1987A is that the facilities capable of recording collapse neutrinos improved and adjusted their physical parameters and are combined into a single network, SNEWS (SuperNova Early Warning System), to form a coincidence trigger for their signals (when they appear) that will notify observatories of other types of radiation of the beginning of supernova observations [41].

In addition, it became clear that detectors with an even larger target mass sensitive to all types of neutrino are required to reliably and completely record the expected Galactic neutrino burst. This is how the third-generation neutrino detectors with a characteristic target mass of about 1 kt or higher and with low event detection energy thresholds (about 5 MeV) emerged. When a Galactic burst of collapse neutrinos occurs, the most powerful of these detectors will record over 8500 neutrinos of all types (SuperKamiokande, the target mass is 32 kt of water) [42], about 1000 neutrinos of all types in the SNO detector with heavy and light water (the total target mass is 2.4 kt) [43], and over 350 neutrinos of all types in the LVD scintillation detector [44]. Apart from the long-operating LSD [18] and MACRO [35] scintillation facilities, which will also record about 100 or more Galactic neutrinos each, the Amanda facility in the ice at the South Pole in Antarctica [46] was included in the network of collapse-neutrino observations in 1997. A number of new detectors are also prepared to be launched or are being designed [47].

Thus, of particular importance is the so unexpected phenomenon of a type-II or, possibly, type-Ib/c supernova [48] in our Galaxy. At least seven neutrino detectors, one of which is the Baksan underground scintillation telescope, were prepared for its recording.

2. THE BAKSAN UNDERGROUND SCINTILLATION TELESCOPE

The Baksan underground scintillation telescope is an instrument for a broad range of scientific research, with one of its experiments being the search for neutrino bursts from supernova explosions. The telescope is located in the Northern Caucasus, at the foot of the Mount Elbrus, in the tunnel under the Andyrchi Mountain at a depth of 850 m. A general view of the telescope is schematically shown in Fig. 1, and the facility is described in detail in [49]. Here, we briefly recall its basic parameters.

The telescope consists of 3150 standard detectors. These detectors form a closed configuration with two internal layers that comprise four vertical planes and four horizontal planes. Each standard detector is 70 × 70 × 30 cm^3 in size, is filled with an organic C_nH_{2n+2} ($n \approx 9$) scintillator, and is viewed by one photomultiplier with a photocathode diameter of 15 cm. The total target mass is 330 t, and the target mass enclosed in the three internal (starting from the lower horizontal plane) layers is 130 t (1200 standard detectors). The charged-particle detection energy thresholds are 8 and 10 MeV for the horizontal and vertical planes, respectively. The energy reference for standard-detector measurements is the amplitude of the most probable energy release when cosmic-ray muons pass through it (50 MeV). Accordingly, the detector thresholds are given in fractions of the amplitude of this energy release.

An intense burst of neutrinos of all types is expected during a type-II supernova explosion. However, the vast majority of the events recorded with the Baksan telescope will be produced in inverse beta-decay reactions,

$$\bar{\nu}_e + p \longrightarrow n + e^+.$$

The contribution from the remaining types of neutrinos to the total recorded signal will be minor [49].

Since the telescope is located at a relatively small depth, a special event-selection method is used to significantly reduce the background from the cosmic-ray muons that cross the facility in our experiment. This method is based on an enormous energy difference between the muons and the positrons formed from collapse electron antineutrinos. Passing through the telescope, the muons leave a track from several triggered detectors, while the positrons formed in a particular detector will lose all of their energy almost entirely in this detector. Thus, the main event-selection criterion in our program, one and only one detector from 3150, implies the selection of single standard-detector triggerings; i.e., at the program marker, data on the state of all devices in the facility at the triggering time of one and only one detector in it is written to the online computer memory and then to storage.

As a result, the neutrino signal from a supernova explosion on the facility will appear as a series of events from singly triggered detectors during the neutrino burst.

The count rate from single background events, the statistical clumps of which can imitate the expected signal, was not constant with time at the beginning of the experiment on the telescope, because operations were continuously performed to remove one or another background source, which caused a continuous reduction in the total count rate of single events. Figure 2 shows time variations in the yearly mean count rate of single detector triggerings summed over the three internal planes (the lower and the next two planes in Fig. 1) in which 130 t of scintillator are contained in 1200 detectors. The data from this part of the facility, i.e., the total number of single triggerings within a given time interval, act as a trigger in the offline analysis of information from the telescope: when an unusual signal was detected in this part of the telescope (i.e., when the number of events was larger than a given number), information from the entire facility was analyzed offline.

Figure 2 also indicates the time intervals when the particle detection energy threshold on the telescope was reduced. As we see from this figure, the background after 1985, when the main operations to remove the strongest noise were finished, was virtually constant within a certain detection energy threshold. Its subsequent modest reduction resulted from the continuous operations to improve the physical parameters of the facility.

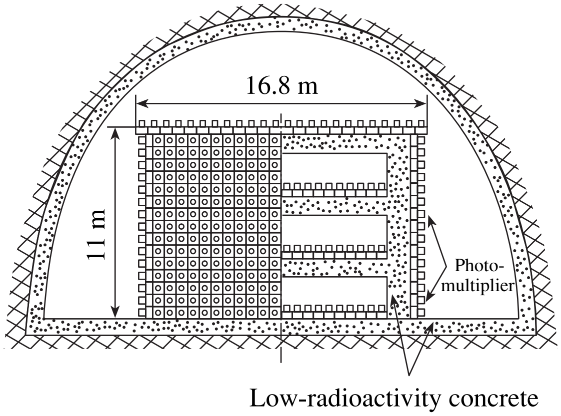


Fig. 1. A schematic view of the Baksan underground scintillation telescope.

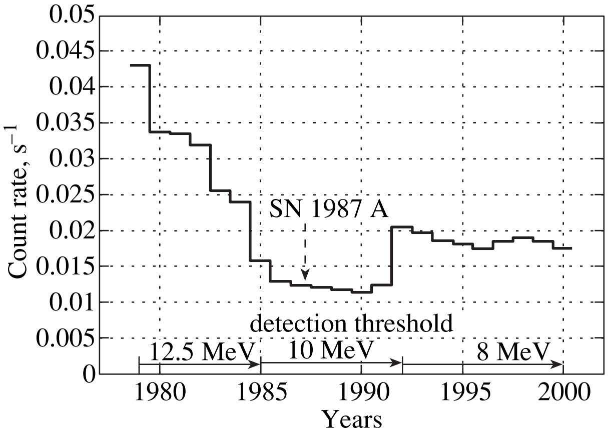


Fig. 2. Time variations in the yearly mean count rate of single triggerings of 1200 detectors in the three telescope internal planes (the target mass is 130 t). The particle detection energy thresholds in different years are shown. The arrow indicates the explosion of SN 1987A on February 23, 1987.

The time when the explosion of SN 1987A occurred is shown in the same figure.

If the core of a massive star collapses at the Galactic center to form a neutron star, then one might expect over 50 events on the telescope from the interactions of electron antineutrinos in the 130-t scintillator or over 100 events in the entire telescope. The current information is analyzed by the method of a 20-s-long time interval sliding from event to event with the total threshold number of events in the signal equal to 5. For detection of a signal with a given or larger number of events, 40 rather than 20 s of information with complete data on each event (coordinates on the telescope, time, energy amplitude, the number of detector triggerings in the signal over the current day, and photomultiplier-pulse duration) is analyzed. A signal would be considered to be a serious candidate for collapse detection if nine or more events were observed for 20 s in the telescope target mass of 130 t (internal layers). No such signal was detected over the entire period of observations with the telescope.

The telescope was constructed late in 1977. However, the physical experiments began later, because it took time for us to figure out how the instrument behaved. The facility has been operating almost continuously under the program of search for collapse neutrinos (the program was called Collapse) since the mid-1980. One maintenance day per week, when the necessary repair and maintenance operations are performed on the facility, and the emergency situations of seasonal power cuts or online computer replacement constitute an exception. So, the total time of Galactic observation accounts for 90% of the calendar time.

Thus, information from the telescope was accumulated under the program of search for collapse neutrinos, but there was no expected neutrino signal. The history of detecting the neutrino signal from SN 1987A in the Baksan data and the related dramatic searches for the cause of the then unexpected clock error have long been published [28, 49]. A new clock with a self-contained power supply has been in operation since February 1988, providing a 1-ms accuracy of determining the absolute time.

3. RESULTS OF THE OBSERVATIONS

When the Baksan experiment to search for collapse neutrinos began, the frequency of expected signals was of great importance. Observations of historical supernovae in our Galaxy, supernovae in other galaxies, stellar statistics, pulsar statistics, and even the thermoluminescence of samples of bottom sediments were used in the literature to determine the time interval between events in which supernovae, neutron stars, or black holes were expected to be formed.

At that time, the most optimistic estimate of this interval was obtained from an analysis based on observations of 149 pulsars, which predicted the birth of one pulsar approximately every six years [50]. Although a later reanalysis of the pulsar birth rate yielded a different estimate of this interval, about 30 years [51], we still hoped for a positive result at the beginning of our experiment, especially since the last supernova was observed on Earth 400 years ago.

Another optimistic result was obtained by analyzing stellar statistics: when calculating the total star death rate with the ultimate formation of pulsars, black holes, supernovae, and any other possibilities, except white dwarfs, the interval between such events was found to be 9^{+2}_{-3} [52]. The estimates for the rate of type-II supernovae in our Galaxy by other authors give a large spread in the predictions of its mean value. This appears to be attributable to a large number of assumptions used and approximations, for example, about the luminosity of the Galaxy [53], its morphological type [54], the initial mass function [55], the fraction of detected pulsars [56], and many others [57–60].

Subsequently, the relationships between supernova statistics and galactic evolution were analyzed and it was concluded that the supernova rates were functions of cosmic time [61, 62]. In addition, a study of the light curves for SNe Ib/c showed them to be similar to the light curves for type-II supernovae [48], led one to conclude that the SN Ib/c progenitor was a massive star, and suggested that neutron stars could also be formed during their explosions.

Recent estimates of the supernova rate from stellar statistics show that the following number of such events may be expected in our Galaxy [63, 64]: 2^{+1}_{-1} for SNe Ib/c and 12^{+6}_{-6} for SNe II in 1000 years. Recently, however, the same authors improved these Galactic values [65]: $1.5^{+1.0}_{-1.0}$ for SNe II + Ib/c in 100 years or about one supernova with the formation of a neutron star every 50 years. The latter value is almost equal to this interval previously estimated to be 10–50 years [66] when the above possible sources of discrepancies are taken into account. In general, it is clear that the sought-for rate estimated from stellar statistics lies within the range 10^{-2}–3×10^{-2} per year or with a mean estimate of the interval between supernovae equal to 47^{+12}_{-12} years [67].

In recent years, it has been found from the statistics of pulsars that they are formed, on average, once every 60–330 years [68], although it is still unclear in what systems the pulsars of a particular class are born [69, 70].

Thus, recent estimates for the rates of supernova explosions and pulsar formation lead us to conclude that this event is rare and that the spread in estimates is large. Therefore, it becomes all the more necessary to obtain the result from direct observations of these events with neutrino detectors.

The Baksan telescope has been observing the Galaxy since June 30, 1980. Because of all the operations performed to increase its sensitivity to the expected burst of collapse neutrinos, the telescope views the entire Galaxy [49]. The calendar time of the Collapse experiment is 19.75 years, while the total live observing time is 17.6 years. No signal, except SN 1987A in the Large Magellanic Cloud, that could be reliably interpreted as a burst of Galactic electron antineutrinos was detected with the facility over this period.

An upper bound on the mean frequency of gravitational collapses in the Galaxy can be obtained from the observing time. If we denote the mean frequency of collapses by f_{collapse} and if we assume that, first, their frequency (as rare events) obeys the Poisson law and, second, the probability of missing the signal is less than 10% at 90% confidence, then we derive the following inequality for the total live observing time T = 17.6 years:

$$\exp(-f_{\text{collapse}}T) < 0.1,$$

whose solution for f_{collapse} yields a bound on the frequency of collapses in the Galaxy,

$$f_{\text{collapse}}\ (90\%\ \text{confidence}) < 0.13\ \text{yr}^{-1}.$$

Hence, the mean time interval $\Delta T_{\text{collapse}}$ between the expected Galactic events exceeds

$$\Delta T_{\text{collapse}}(90\%\ \text{confidence}) > 7.7\ \text{years}.$$

Thus, the first twenty years on the path to detecting the first neutrino burst in our Galaxy have been traversed, although this may prove to be only a small part of the required path.

ACKNOWLEDGMENTS

We wish to thank the staff of the Baksan neutrino observatory, those who are still working and those who have left it, for a long and fruitful collaboration in this experiment. This study was supported by the Russian Foundation for Basic Research (project no. 00-02-17778).

REFERENCES

1. W. Pauli, in *Structures et proprietes des noyaux atomiques: Instituts Solvay, VI Conseil de Physique, 1933* (Gauthier-Villars, Paris, 1934), p. 324.
2. F. Reines and C. L. Cowan, Phys. Rev. **90**, 492 (1953); **113**, 273 (1959).
3. K. Lande, B. Cleveland, T. Daily, *et al.*, AIP Conf. Proc. **243**, 1122 (1992).
4. V. N. Gavrin, Nucl. Phys. (Proc. Suppl.) **91**, 36 (2001).
5. W. Hampel, Phys. Lett. B **447**, 127 (1999).
6. V. Baade and F. Zwicky, Phys. Rev. **45**, 138 (1934).
7. Ya. B. Zel'dovich and O. Kh. Gusseĭnov, Dokl. Akad. Nauk SSSR **162**, 791 (1965) [Sov. Phys. Dokl. **10**, 524 (1965)].
8. G. V. Domogatsky and G. T. Zatsepin, in *Proceedings of the 9th International Cosmic Ray Conference, ICRC, London, 1965*, Vol. 2, p. 1030.
9. S. A. Colgate and R. H. White, Astrophys. J. **143**, 626 (1966).
10. W. J. Cocke, H. J. Disney, and D. J. Taylor, Nature **221**, 525 (1969).
11. M. I. Large, A. E. Vaugham, and B. Y. Mills, Nature **220**, 340 (1968).
12. G. G. Pavlov, D. Sanwal, G. P. Garmire, and V. E. Zavlin, Astron. Soc. Pac. Conf. Ser. (2002); astro-ph/0112322.
13. V. M. Kaspi and D. J. Helfand, Astron. Soc. Pac. Conf. Ser. (2002); astro-ph/0201183.
14. K. Lande, G. Bozoki, W. Frati, *et al.*, Nature **251**, 485 (1974).
15. M. Deabyne, W. Frati, K. Lande, *et al.*, in *Proceedings of the 16th International Cosmic Ray Conference, ICRC, Kyoto, Japan, 1979*, Vol. 13, p. 325.
16. G. Bozoki, in *Proceedings of the 13th International Cosmic Ray Conference, ICRC, Denver, USA, 1973*, Vol. 3, p. 1994.
17. E. N. Alexeyev, L. N. Alexeyeva, and A. E. Chudakov, in *Proceedings of the 16th International Cosmic Ray Conference, ICRC, Kyoto, Japan, 1979*, Vol. 10, p. 282.
18. A. E. Chudakov and O. G. Ryazhskaya, in *Proceedings of the International Conference "Neutrino 77," Baksan Valley, USSR, 1077* (Nauka, Moscow, 1978), Vol. 1, p. 155.
19. K. Arisaka, K. Hirata, T. Kajita, *et al.*, J. Phys. Soc. Jpn. **54**, 3213 (1985).
20. R. M. Bionta, G. Blewitt, G. B. Bratton, *et al.*, Phys. Rev. Lett. **51**, 27 (1983).
21. V. S. Imshennik and D. K. Nadezhin, Itogi Nauki Tekh., Ser.: Astron. **21**, 63 (1982).
22. B. Bowers and J. R. Wilson, Astrophys. J. **263**, 366 (1982).
23. E. E. Salpeter and S. L. Shapiro, Astrophys. J. **251**, 311 (1981).
24. A. Burrows and T. L. Mazurek, Nature **301**, 315 (1983).
25. K. S. Hirata, T. Kajita, M. Koshiba, *et al.*, Phys. Rev. Lett. **58**, 1490 (1987).
26. R. M. Bonita, G. Blewitt, C. B. Bratton, *et al.*, Phys. Rev. Lett. **58**, 1494 (1987).
27. E. N. Alekseev, L. N. Alekseeva, V. I. Volchenko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 461 (1987) [JETP Lett. **45**, 589 (1987)]; E. N. Alexeyev, L. N. Alexeyeva, I. V. Krivosheina, *et al.*, Phys. Lett. B **205**, 209 (1988).
28. N. Panagia, in *Invited Lecture at the International Summer School "Experimental Physics of Gravitational Waves," Urbino (Italy)*, Ed. by G. Calamai, M. Mazzoni, R. Stanga, and F. Vetrano (World Scientific, Singapore, 1999); astro-ph/0003083.
29. A. Fassia, W. P. S. Meikle, and J. Spyromilio, Mon. Not. R. Astron. Soc. **332**, 296 (2002); astro-ph/0112516.
30. S. Park, D. N. Burrows, G. P. Garmire, *et al.*, Astrophys. J. **567**, 314 (2002); astro-ph/0111116.
31. J. Middleditch, J. A. Kristian, W. E. Kunkel, *et al.*, New Astron. **5**, 243 (2000).
32. S. Nagataki and K. Sato, Prog. Theor. Phys. **105**, 429 (2001).
33. H.-T. Janka, Astron. Astrophys. **368**, 527 (2001).
34. H.-T. Janka, K. Kifonidis, and M. Rampp, Lect. Notes Phys. **578**, 333 (2001).
35. V. M. Suslin, S. D. Ustyugov, V. M. Chechetkin, and G. P. Churkina, Astron. Zh. **78**, 281 (2001) [Astron. Rep. **45**, 241 (2001)].
36. J. C. Wheeler, D. L. Meier, and J. R. Wlson, astro-ph/0112020.
37. S. W. Bruenn, K. R. DeNisco, and A. Mezzacappa, astro-ph/0101400.
38. C. J. Horowitz, astro-ph/0109209; Phys. Rev. D **65**, 043001 (2002).
39. K. Takahashi, M. Watanabe, K. Sato, and T. Totani, Phys. Rev. D **64**, 093004 (2001).
40. Th. J. Loredo and D. Q. Lamb, astro-ph/0107260.
41. A. Habig, in *Proceedings of the 10th Annual October Astrophysics Conference "College Park 1999, Cosmic Explosions," Maryland, 1999*, p. 169.
42. J. F. Beacom and P. Vogel, Phys. Rev. D **58**, 053010 (1998).
43. C. J. Virtue, Nucl. Phys. (Proc. Suppl.) **100**, 326 (2001).

44. M. Aglietta, Nuovo Cimento A **105**, 1793 (1992).
45. S. Ahlen, Astropart. Phys. **1**, 11 (1992).
46. J. Ahrens, X. Bai, G. Barouch, *et al.*, Astropart. Phys. **16**, 345 (2002).
47. K. Scholberg, Nucl. Phys. (Proc. Suppl.) **91**, 331 (2000).
48. A. V. Filippenko, in *Cosmic Explosions*, Ed. by S. Holt and W. W. Zhang (American Institute of Physics, New York, 2000); astro-ph/0002264.
49. E. N. Alekseev, L. N. Alekseeva, V. I. Volchenko, *et al.*, Zh. Éksp. Teor. Fiz. **104**, 289 (1993) [JETP **77**, 339 (1993)].
50. J. H. Taylor and R. N. Manchester, Astrophys. J. **215**, 885 (1977).
51. V. Trimble, Rev. Mod. Phys. **54**, 1183 (1982).
52. J. N. Bahcall and T. Piran, Astrophys. J. Lett. **267**, L77 (1983).
53. S. van den Bergh, R. D. McClure, and R. Evans, Astrophys. J. **323**, 44 (1987).
54. E. Cappellaro and M. Turatto, Astron. Astrophys. **190**, 10 (1988).
55. O. G. Richter and M. Rosa, Astron. Astrophys. **206**, 219 (1988).
56. R. Narayan and K. J. Schaudt, Astrophys. J. Lett. **325**, L43 (1988).
57. S. van den Bergh, Astron. Astrophys. **231**, L2 (1990).
58. D. Battacharya, J. Astrophys. Astron. **11**, 125 (1990).
59. B. N. G. Guthrie, Astron. Astrophys. **234**, 84 (1990).
60. G. A. Tammann and A. Schroeder, Astron. Astrophys. **236**, 149 (1990).
61. P. Madau, in *The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift*, Ed. by S. D'Odorico, A. Fontana, and E. Giallongo, Astron. Soc. Pac. Conf. Ser. **146**, 289 (1998); astro-ph/9801005.
62. P. Madau, M. D. Valle, and N. Panagia, submitted to Mon. Not. Roy. Astron. Soc. **297**, L17 (1998); astro-ph/9803284.
63. E. Cappellaro, M. Turatto, D. Yu. Tsvetkov, *et al.*, Astron. Astrophys. **322**, 431 (1997).
64. E. Cappellaro, R. Evans, and M. Turatto, Astron. Astrophys. **351**, 459 (1999).
65. E. Cappellaro, in *Invited Review at the Meeting "The Influence of Binaries on Stellar Population Studies," Brussels, 2000,* Ed. by D. Vanbeveren; astro-ph/0012455.
66. S. van den Bergh and G. A. Tammann, Annu. Rev. Astron. Astrophys. **29**, 363 (1991).
67. G. A. Tammann, W. Loeffler, and A. Schroeder, Astrophys. J., Suppl. Ser. **92**, 487 (1994).
68. A. G. Lyne, Mon. Not. R. Astron. Soc. **295**, 743 (1998).
69. S. F. Portgies Zwart and E. P. J. van den Heuvel, New Astron. **4**, 355 (1999).
70. D. R. Lorimer, in *Proceedings of NATO ASI "The Neutron Star-Black Hole Connection,"* astro-ph/9911519.

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 57–63.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 68–75.
Original Russian Text Copyright © 2002 by Kaznacheev, Bogdanov, Taraskin.

FLUIDS

The Nature of Prolate Shape of Tactoids in Lyotropic Inorganic Liquid Crystals

A. V. Kaznacheev*, M. M. Bogdanov, and S. A. Taraskin
Nesmeyanov Institute of Organoelemental Compounds, Russian Academy of Sciences, Moscow, 117813 Russia
**e-mail: son@pmc.ineos.ac.ru*

Received January 9, 2002

Abstract—A tactoid phase is obtained, which is an anisotropic mesophase dispersed in an anisotropic solution. A model is constructed that describes the shape of anisotropic regions, i.e., tactoids. It is demonstrated that the prolate shape of tactoids is caused by the competition between the elastic energy of the nematic phase and the surface energy. This model is used to find the ratio of elastic constants K_3/K_1 from the experimental data. It is experimentally found that a magnetic field orients and stretches tactoids. An explanation of this phenomenon is suggested. *© 2002 MAIK “Nauka/Interperiodica”.*

Thermotropic and lyotropic liquid crystals are well known at present [1–3]. The former of these crystals are chemically pure substances consisting of rod-shaped or disk-shaped molecules. The interaction between such molecules results in the orientational order of their preferred axes in the absence of long-range three-dimensional translation order. The lyotropic liquid crystals are formed when some substances are dissolved in certain solvents [2–4]. In this case, anisometric structure elements are provided by aggregates of molecules. Their size and shape depend on the concentration and temperature. This brings about a wide variety of liquid-crystal phases and phase transitions between them. By the chemical classification, the lyotropic liquid crystals include a class of inorganic liquid crystals. The latter crystals arise in dispersions of inorganic substances [5, 6].

In recent years, investigations of the molecular structure of such liquid crystals have been developing [5–8]. However, the macroscopic properties characteristic of liquid crystals, such as the elastic constants and coefficients of viscosity [1], have hardly been investigated. Note two papers [9, 10] which deal with the problem of elastic constants in the nematic phase of the lyotropic inorganic liquid crystal of vanadium pentoxide (V_2O_5)–water. Generalova *et al.* [10] revealed for the first time the effect of a magnetic field on the nematic phase of this system and the emergence of magnetohydrodynamic domains which are typical of other types of liquid crystals as well [11, 12]. The data on the Freedericksz transition were used to obtain the value of splay elastic constant $K_1 = 4 \times 10^{-7}$ dyn [10]. In order of magnitude, it agrees with the analogous data for thermotropic and other types of lyotropic liquid crystals. No fundamental difference between them is observed.

However, a phenomenon absent from liquid crystals of other types is observed in lyotropic inorganic liquid crystals. This phenomenon was revealed in the 1920s by Zocher and consists in that, under conditions of low concentration (~1% by mass) of inorganic substances, an isotropic phase is observed containing anisotropic droplets of prolate shape [5, 6, 13]. Zocher referred to such a two-phase system preceding the nematic phase as a tactoid phase, and to the anisotropic droplets as tactoids.

The fact of existence of two-phase systems in the form of dispersion is well known, for example, liquid crystal droplets in an isotropic matrix [14] and droplets of liquid-crystal phase coexisting with the isotropic phase in the process of melting. In these cases, droplets of different sizes have the shape of a sphere, which is associated with the minimum of surface energy. On the other hand, in the tactoid phase, the anisotropic droplets are of prolate shape.

Until recently, the question of such a shape of tactoids remained open. We suggest an explanation of this phenomenon. Comparison of the theoretical results with the data of experiments involving the measurement of tactoids of different volumes made it possible to estimate the ratio of the splay and bend elastic constants of the nematic phase of droplets. In addition, it was experimentally observed that a magnetic field orients and stretches tactoids in its direction. An explanation of this phenomenon is suggested.

A typical representative of lyotropic inorganic liquid crystals is provided by sols of vanadium pentoxide in water. They were prepared by the procedure described by Biltz [15]. The texture of the tactoid phase of such sols is given in Fig. 1. Against the dark background corresponding to the isotropic phase, one can see anisotropic regions, i.e., tactoids. The basic facts following from the observation of tactoids may be summarized as follows.

1. We observed the coalescence of tactoids. The stages of this process are given in Fig. 2. After coales-

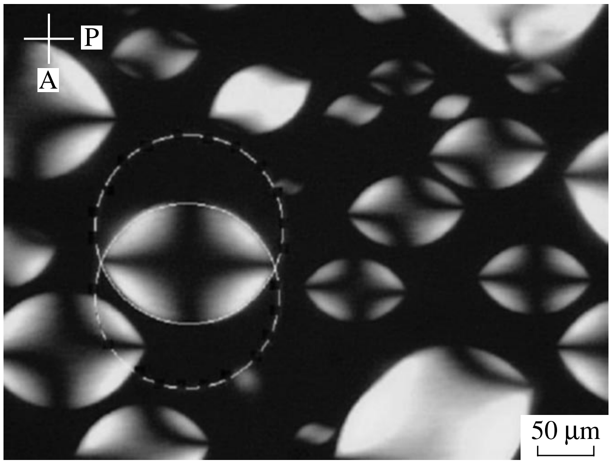


Fig. 1. The texture of tactoid phase in a vanadium pentoxide–water system with the concentration of V_2O_5 of 1.1% by mass. The cell thickness is 200 μm.

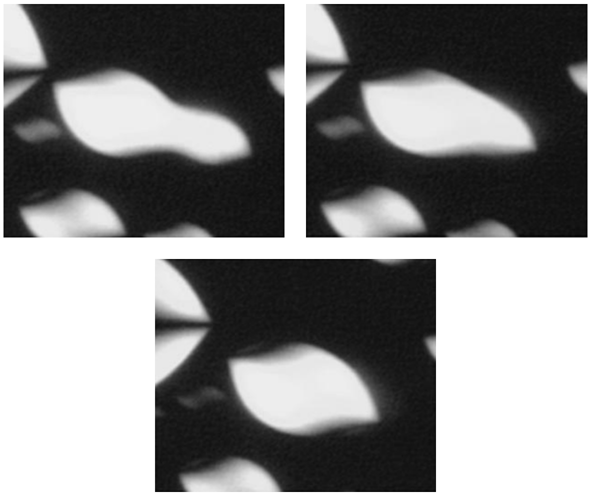


Fig. 2. The process of coalescence of tactoids.

cence, the shape of the resultant tactoid remains spindle-shaped. This is indicative of the fact that tactoids are equilibrium droplets.

2. The tactoid boundaries are circular arcs (Fig. 1).

3. The texture of a tactoid varies during its rotation between crossed polaroids (Fig. 3). This variation of the texture indicates that the field of director on the tactoid boundaries has a tangential orientation, and defects such as boojums are located at the tactoid vertexes [14].

Based on these observations, it is assumed that a tactoid is a droplet of the nematic phase coexisting with the isotropic phase. A model of its shape and director field is given in Fig. 4. The tactoid boundary is the surface of revolution of a circular arc of radius R about the z axis. This choice of boundary is based on the experimental data (Fig. 1). Observations indicate that, for tactoids of different sizes less than the measuring cell thickness, the boundaries remain circular arcs.

We assume that the field of director n at each point of a tactoid coincides with the unit vector $\mathbf{e}_\eta$ of bispherical coordinates (see Fig. 4) [16]. This choice of the field of director satisfies the tangential boundary conditions. A strong anchoring between the director and tactoid surface is presumed in this case. This assumption is based on the fact that, in all problems in which the cohesive energy W is included, the latter energy appears in the dimensionless form of $WV^{1/3}/K$, where V is the tactoid volume and K is one of the elastic constants of liquid crystal. Therefore, at least for large tactoids, the condition of strong anchoring must be valid. With the given choice of the field of director, the twist associated with the elastic constant K_2 is absent. This follows from the form of tactoid textures (Fig. 3). In the presence of twist of the field of director in droplets of thermotropic nematic liquid crystals, the textures have a different form [14].

So, the shape of a tactoid is defined by two parameters, namely, R and α (Fig. 4). At $\alpha = \pi/2$, a tactoid transforms into a sphere. The tactoid free energy Φ is made up of the elastic energy Φ_{el} of the nematic phase and the surface energy $\Phi_s = \sigma S$, where σ is the surface tension and S is the tactoid surface area. The expression for free energy Φ may be written directly, proceeding from the dimensional representation,

$$\Phi = K_1 R\Phi_{el}^{(1)}(\alpha) + K_3 R\Phi_{el}^{(3)}(\alpha) + \sigma R^2\Phi_s(\alpha), \quad (1)$$

where K_1 and K_3 are the splay and bend elastic constants, respectively [1, 17], and $\Phi_{el}^{(1)}(\alpha)$ and $\Phi_{el}^{(3)}(\alpha)$ are dimensionless functions of the angle α associated with these strains. The calculation of these functions is given in Appendix I. Note that, on approaching the point defects on the tactoid surface, i.e., boojums, the elastic energy density diverges. However, its integration over the tactoid volume brings about a finite value. The expression for the dimensionless function $\Phi_s(\alpha)$ is

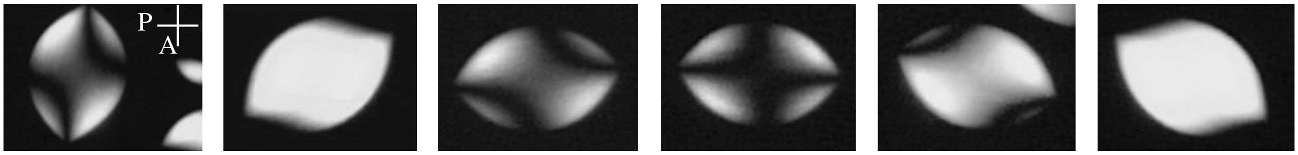


Fig. 3. The variation of the texture of a tactoid during its rotation between crossed polaroids.

derived in calculating the tactoid surface area and has the form

$$\Phi_s(\alpha) = 4\pi(\sin\alpha - \alpha\cos\alpha). \qquad (2)$$

A tactoid of constant volume V takes such values of R and α at which its total energy Φ is minimal. The function given by Eq. (1) contains two variables R and α related by

$$V = R^3\Psi(\alpha) = \text{const}, \qquad (3)$$

where

$$\Psi(\alpha) = 2\pi\left(\sin\alpha - \alpha\cos\alpha - \frac{\sin^3\alpha}{3}\right)$$

is introduced.

We express R from Eq. (3) and substitute into Eq. (1) to derive

$$\tilde{\Phi} = \frac{K_i}{\sigma V^{1/3}}\Psi_i(\alpha) + \Psi_s(\alpha), \qquad (4)$$

where $\tilde{\Phi} = \Phi/\sigma V^{2/3}$ is the dimensionless energy, $\Psi_i(\alpha) = \Phi_{el}^{(i)}/\Psi^{1/3}$ ($i = 1, 3$), the recurring index i here and in what follows indicates summation, and $\Psi_s(\alpha) = \Phi_s(\alpha)/\Psi^{2/3}$. The first two terms of formula (4) are related to the elasticity of the nematic phase. They increase with increasing angle α. The surface energy is described by the function $\Psi_s(\alpha)$. It is universal and decreases with increasing α. As a result of competition between the elastic and surface terms in formula (4), the minimum of free energy is attained at $\alpha < \pi/2$; i.e., a nematic droplet will have a prolate shape. The position of the minimum of free energy depends on the dimensionless parameters $K_i/\sigma V^{1/3}$ ($i = 1, 3$). It follows from formula (4) that an increase in the tactoid volume causes a decrease in the elastic energy contribution. Consequently, the shape of large tactoids must approach spherical. By studying for extremum the free energy given by Eq. (4), one can derive the formula for the dependence of the tactoid curvature radius R_0 on the angle α,

$$R_0 = C_i f_i(\alpha), \qquad (5)$$

where $C_i = K_i/\sigma$ denotes the characteristic dimensions of the problem and $f_i(\alpha) = -\Psi_i'/(\Psi^{1/3}\Psi_s')$ denotes increasing dimensionless functions of the angle α. They have the form

$$f_1(\alpha) = \frac{\alpha\sin\alpha - 2\alpha^2\cos\alpha + \sin^2\alpha\cos\alpha}{\cos\alpha[\alpha(\alpha + \sin\alpha\cos\alpha) - 2\sin^2\alpha]},$$

$$f_3(\alpha) = \frac{(\alpha^2 - \sin^2\alpha)[\alpha(1 + 2\cos^2\alpha) - 3\sin\alpha\cos\alpha]}{4\sin\alpha\cos\alpha[\alpha(\alpha + \sin\alpha\cos\alpha) - 2\sin^2\alpha]}.$$

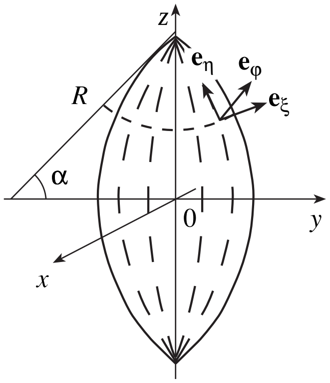


Fig. 4. A model of tactoid.

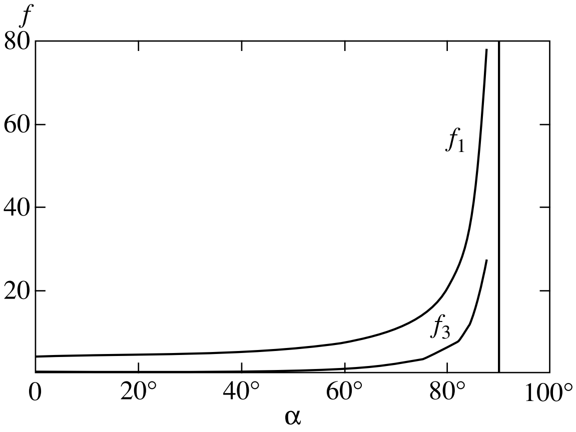


Fig. 5. Graphs of the functions $f_1(\alpha)$ and $f_3(\alpha)$.

For $\alpha \longrightarrow 0$, we have $f_1 \longrightarrow 4$ and $f_3 \longrightarrow 10\alpha^2$. As $\alpha \longrightarrow \pi/2$, $f_{1,3} \longrightarrow \infty$. The graphs of these functions are given in Fig. 5.

By measuring experimentally the correlation $R_0(\alpha)$ for tactoids of different volume and comparing it with formula (5), one can obtain data on C_1 and C_3. For this purpose, we selected a lyotropic inorganic liquid crystal of V_2O_5–water. Its sols were prepared by the procedure described by Biltz [15]. We observed the tactoid phase in the range of concentration of V_2O_5 from 0.3 to 1.2% by mass. The system being investigated was fed into plane-parallel glass capillaries with a thickness of the order of 200 μm, which were hermetically sealed with picein. The thickness of the capillaries was defined by Teflon spacers and, prior to charging, was measured by the interference method. Immediately after charging, the texture of substance in the capillaries was anisotropic. We observed the separation of the texture into the isotropic and anisotropic regions. Later on, the anisotropic regions assumed the form of tactoids. In doing so, their size distribution was observed.

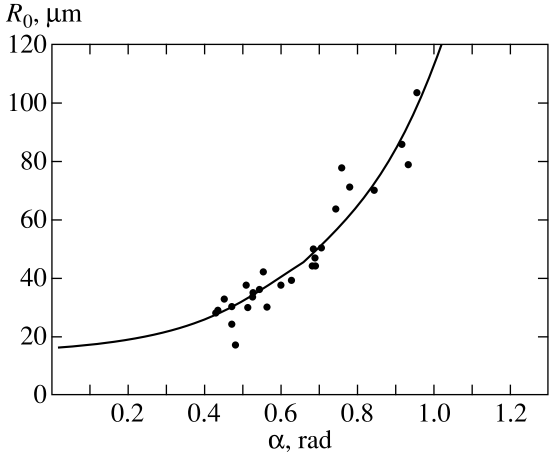


Fig. 6. The correlation $R_0(\alpha)$ at the concentration of V_2O_5 of 0.33% by mass. The values of $C_1 = 4.1 \pm 0.6$ μm and $C_3 = 92.1 \pm 8.6$ μm.

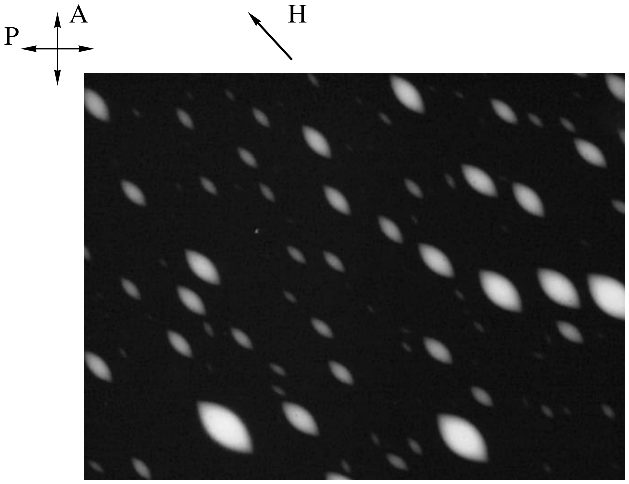


Fig. 7. The orientation of tactoids in a magnetic field.

In order to support the fact that the prolate shape of tactoids is equilibrium, flows were mechanically excited in the capillaries with the substance being investigated. The tactoid boundaries were smeared, and birefringence was induced in the isotropic phase due to the flows. The texture as a whole became anisotropic. In the course of time, the same process occurred as that observed immediately after charging the substance into the capillary. The anisotropic regions again assumed the equilibrium shape of tactoids.

In order to let the system to come to equilibrium, all measurements were performed a day after the substance was charged into the capillaries. The measurements involved the use of an Axiolab Pol polarizing microscope by Zeiss. The image was displayed on a monitor via a videocamera. The tactoid dimensions were determined with the aid of a Linkam VTO 232 adapter. Only those tactoids were measured whose transverse dimensions were less than the cell thickness.

The measurements gave the values of the parameters R_0 and α for tactoids of different volumes. Figure 6 shows a typical experimental dependence $R_0(\alpha)$. The method of least squares and the function given by Eq. (5) were used to calculate the coefficients C_1 and C_3. The solid line in Fig. 6 gives the approximating curve. The C_3/C_1 ratio is equal to the K_3/K_1 ratio. For all of the investigated samples, $C_1 \sim 1$ μm, $C_3 \sim 10$ to 100 μm, and $K_3/K_1 \sim 10$ to 100. The obtained values of K_3/K_1 are very high compared with those in the case of thermotropic and organic lyotropic liquid crystals. Usually, K_3/K_1 does not exceed three [1, 18]. If we use the values of $K_1 = 4 \times 10^{-7}$ dyn (obtained from the data on the Freedericksz transition in the nematic phase of the system being investigated [10]) and $C_1 \sim 1$ μm, we can estimate the surface tension $\sigma \sim 10^{-3}$ erg/cm^2 at the tactoid boundary. As a result of such low values of surface tension, macroscopic tactoids are prolate.

It is known that a magnetic field orients tactoids in its direction [13] (Fig. 7). We have further found that a magnetic field stretches tactoids. This is most evident in the case of large tactoids (Fig. 8). In order to explain this phenomenon, one must include in expression (1) for the free energy Φ the magnetic term arising as a result of the effect of magnetic field **H** on the field of director **n**. This term is associated with the anisotropy of magnetic susceptibility χ_a. The magnetic energy density F_m has the form [1, 17]

$$F_m = -\frac{\chi_a}{2}(\mathbf{n}\cdot\mathbf{H})^2. \qquad (6)$$

At $\chi_a > 0$, the magnetic field tends to orient the director in its direction. The integration of expression (6) over the tactoid volume enables one to calculate the magnetic energy (see Appendix II). The result may be written directly, proceeding from the dimensional representation. It has the form

$$-\chi_a H^2 R^3 \Phi_m(\alpha), \qquad (7)$$

where $\Phi_m(\alpha)$ is a dimensionless function of the angle α (see Appendix II). We add formula (7) to the right-hand part of expression (1) and, as previously, eliminate R to derive

$$\tilde{\Phi} = \frac{K_i}{\sigma V^{1/3}}\Psi_i(\alpha) + \Psi_s(\alpha) - \frac{\chi_a H^2 V^{1/3}}{\sigma}\Psi_m(\alpha), \qquad (8)$$

where $\Psi_m(\alpha) = \Phi_m(\alpha)/\Psi$ is a decreasing dimensionless function of the angle α.

It follows from formula (8) that, at low values of V, the magnetic term may be ignored. In this case, however, elastic terms are important. Therefore, small tactoids are stretched owing to the competition between the elastic and surface forces. For large tactoids (high

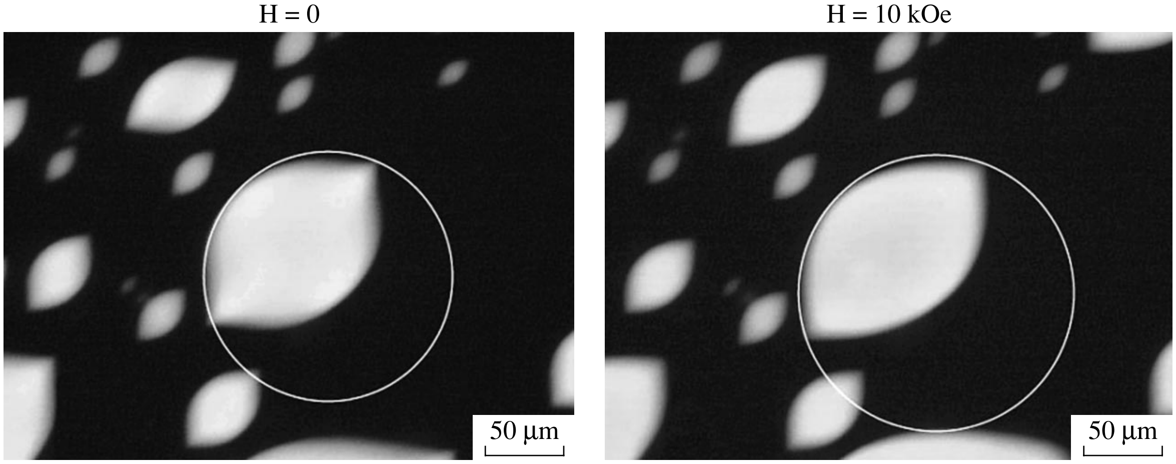


Fig. 8. A magnetic field stretches large tactoids.

values of V), on the contrary, the elastic terms are small, and the competition arises between the magnetic and surface terms in expression (8). This brings about the prolate shape of large tactoids.

For quantitative solution of the problem, one must study expression (8) of the energy $\tilde{\Phi}$ for an extremum at constant volume V. This leads to a quadratic equation relative to the tactoid curvature radius R,

$$R^2 - \left[\frac{\sigma}{\chi_a H^2}\frac{\Psi'_s}{\Psi'_m\Psi^{1/3}}\right]R - \frac{K_i}{\chi_a H^2}\frac{\Psi'_i}{\Psi'_m\Psi^{2/3}} = 0. \qquad (9)$$

The solution to this equation has two roots $R_\pm$ which are

$$R_\pm = \frac{\sigma}{2\chi_a H^2}\frac{\Psi'_s}{\Psi'_m\Psi^{1/3}}\left[1 \pm \sqrt{1 + \frac{4K_i\chi_a H^2}{\sigma^2}\frac{\Psi'_i\Psi'_m}{(\Psi'_s)^2}}\right]. \qquad (10)$$

If we use formula (5) for $R_0(\alpha)$, i.e., the curvature radius of equilibrium tactoid in the absence of a magnetic field, Eq. (10) may be conveniently written as

$$R_\pm = \frac{\sigma}{2\chi_a H^2}\frac{1}{f(\alpha)}\left[1 \pm \sqrt{1 - \frac{4\chi_a H^2}{\sigma}R_0(\alpha)f(\alpha)}\right], \qquad (11)$$

where

$$f(\alpha) = \frac{\Psi'_m\Psi^{1/3}}{\Psi'_s} = \frac{\sin^2\alpha\cos\alpha(6+\sin^2\alpha) + 6\alpha^2\cos^3\alpha - \alpha\sin\alpha(12 - 11\sin^2\alpha)}{24\cos\alpha[2\sin^2\alpha - \alpha(\alpha + \sin\alpha\cos\alpha)]}$$

is an increasing function of the angle α. At $\alpha \longrightarrow 0$, $f \longrightarrow 0$. At $\alpha \ll 1$, the radical in formula (11) may be expanded into a series. Then, we have

$$R_- \approx R_0(\alpha),$$

$$R_+ \approx \frac{\sigma}{\chi_a H^2}\frac{1}{f(\alpha)}.$$

If $\alpha > \alpha_c$ found from the solution to the equation

$$\frac{4\chi_a H^2}{\sigma}R_0(\alpha_c)f(\alpha_c) = 1,$$

real solutions of $R_\pm$ disappear. This corresponds to the situation in which even large droplets are prolate due to the magnetic field and, for these droplets, $\alpha < \alpha_c$. Both values of $R_\pm$ in formula (11) correspond to the minimum of free energy given by Eq. (8). For one and the same value of α, a small tactoid corresponds to R_-, and a larger tactoid corresponds to R_+. Graphs of function (11) for $\sigma = 10^{-3}$ erg/cm^2, $\chi_a = 10^{-9}$ [10], $C_1 = 4$ μm, $C_3 = 40$ μm, and for different values of the magnetic field H are given in Fig. 9. The two-valuedness of this function is associated with the fact that small tactoids are stretched due to elastic forces, and large tactoids, due to magnetic forces.

The ratios $K_3/K_1 \sim 10$ obtained by us agree with analogous data for another lyotropic nematic system of tobacco mosaic virus (TMV)–water. For this system, a study into magnetohydrodynamic domains produced a value of $K_3/K_1 = 8.8$ [19]. Note that, although the TMV–water system does not belong to lyotropic inorganic liquid crystals, it exhibits the tactoid phase as well [20]. In addition, the ratio $K_3/K_1 \sim 10$ was measured in the nematic phase N_1 of the lyotropic inorganic liquid crystal of a tetrapalladium organyl–pentadecane system [21]. These measurements were based on the

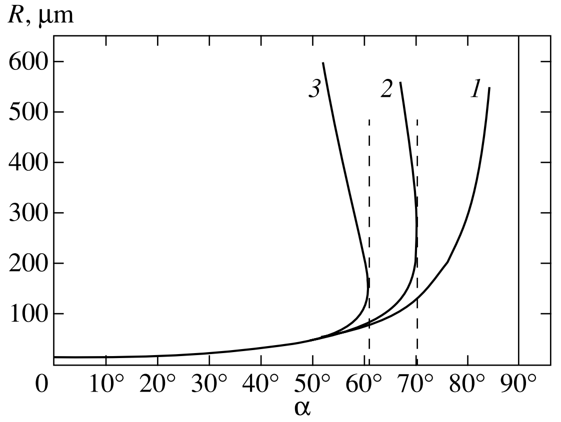


Fig. 9. The correlation $R_0(\alpha)$ for different values of the magnetic field intensity: (*1*) 0, (*2*) 10, (*3*) 20 kOe; C_1 = 4 μm, C_3 = 40 μm, $\sigma = 10^{-3}$ erg, $\chi_a = 10^{-9}$.

investigation of the Freedericksz transition. The high values of K_3/K_1 are apparently associated with the high values of the ratio of length L to diameter D of structure elements. For example, for TMV, $L/D \approx 17$ [19]; at the same time, for a typical representative of thermotropic liquid crystals such as paraazoxy anisole (PAA), $L/D \approx 4$ [1].

The high values of ratio $K_3/K_1 > 10$ obtained by us may be attributed to the fact that a strong anchoring between the director and boundary is assumed in the tactoid model. However, as is demonstrated above, this condition may be violated in the case of small tactoids. Therefore, for processing such experimental results, the model must be refined with due regard for the cohesive energy.

The suggested explanation of the prolate shape of tactoids is based on the competition between the elastic energy of the nematic phase and the surface energy. This is possible on condition of strong anchoring between the director and boundary. In this case, the surface tension $\sigma \sim 10^{-3}$ erg/cm^2 is low. Usually, in the case of thermotropic liquid crystals, the situation is opposite: $\sigma \sim 10$ erg/cm^2 and $W \sim 10^{-2}$ erg/cm^2 [22]. Therefore, no prolate shape of droplets is observed in these crystals. In order to compare W and σ, one needs additional experimental data on the value of cohesive energy in the system being investigated. No such data are available at present; however, even now one can expect that $W \geq \sigma$. Apparently, this points to one of the specific features of lyotropic inorganic liquid crystals.

APPENDIX I

The expression for the elastic energy density $F_{\rm el}$ of the nematic phase has the form [1, 17]

$$F_{\rm el} = \frac{K_1}{2}(\operatorname{div}\mathbf{n})^2 + \frac{K_2}{2}(\mathbf{n}\cdot\operatorname{curl}\mathbf{n})^2 + \frac{K_3}{2}(\mathbf{n}\times\operatorname{curl}\mathbf{n})^2, \qquad \text{(I.1)}$$

where K_1, K_2, and K_3 are the splay, twist, and bend elastic constants, respectively, and $\mathbf{n}$ is the director. The unit vector $\mathbf{n}$ coincides with the unit vector $\mathbf{e}_\eta$ in the bispherical coordinates ξ, η, φ. In these coordinates, $\mathbf{n} = (0, 1, 0)$. The bispherical coordinates are related to the Cartesian coordinates x, y, z by [16]

$$x = \frac{a\sin\xi\cos\varphi}{\cosh\eta - \cos\xi},$$

$$y = \frac{a\sin\xi\sin\varphi}{\cosh\eta - \cos\xi},$$

$$z = \frac{a\sinh\eta}{\cosh\eta - \cos\xi},$$

where $a = R\sin\alpha$ is the transformation parameter (Fig. 4). The elastic energy $\Phi_{\rm el}$ of the nematic phase is obtained by integration of expression (I.1) over the tactoid volume. In the bispherical coordinates, it has the form

$$\Phi_{\rm el} = \frac{a}{2}\int_0^{2\pi} d\varphi \int_{-\infty}^{\infty} d\eta \times \int_{\pi-\alpha}^{\pi} \frac{4K_1\sinh^2\eta\sin\xi + K_3\sin^3\xi}{(\cosh\eta - \cos\xi)^3}d\xi. \qquad \text{(I.2)}$$

We integrate Eq. (I.2) to derive

$$\Phi_{\rm el} = K_1R\times 4\pi(\sin\alpha - \alpha\cos\alpha) + K_3R\pi(3\sin\alpha - 3\alpha\cos\alpha - \alpha^2\sin\alpha). \qquad \text{(I.3)}$$

Comparison of Eq. (I.3) with the first two terms of formula (1) enables one to write

$$\Phi_{el}^{(1)}(\alpha) = 4\pi(\sin\alpha - \alpha\cos\alpha),$$

$$\Phi_{el}^{(3)}(\alpha) = \pi(3\sin\alpha - 3\alpha\cos\alpha - \alpha^2\sin\alpha).$$

APPENDIX II

Expression (6) for the magnetic energy density F_m depends on $(\mathbf{n}\cdot\mathbf{H})^2$. If $\mathbf{H}$ coincides with the tactoid axis Z (Fig. 4), then $(\mathbf{n}\cdot\mathbf{H})^2 = n_z^2H^2$. For the field of director $\mathbf{n}$, we can write $\mathbf{n} = \mathbf{e}_\eta/e_\eta$, where

$$\mathbf{e}_\eta = \frac{\partial x}{\partial\eta}\mathbf{i} + \frac{\partial y}{\partial\eta}\mathbf{j} + \frac{\partial z}{\partial\eta}\mathbf{k};$$

$\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors of Cartesian coordinates; and e_η is the magnitude of $\mathbf{e}_\eta$. Then, for n_z^2, we have

$$n_z^2 = \left(\frac{1 - \cosh\eta\cos\xi}{\cosh\eta - \cos\xi}\right)^2. \qquad \text{(II.1)}$$

In order to calculate the magnetic energy Φ_m, we integrate expression (6) in view of (II.1) over the tactoid volume in bispherical coordinates. This integral has the form

$$\Phi_m = -\frac{\chi_a H^2}{2} a^3 \int_0^{2\pi} d\varphi \int_{-\infty}^{\infty} d\eta \times \int_{\pi-\alpha}^{\pi} \frac{\sin\xi(1-\cosh\eta\cos\xi)^2}{(\cosh\eta-\cos\xi)^5} d\xi. \qquad \text{(II.2)}$$

The calculation of (II.2) results in

$$\Phi_m = -\chi_a H^2 R^3 \frac{\pi}{36} \times [\sin\alpha(20+\cos^2\alpha) - 3\alpha\cos\alpha(7+2\sin^2\alpha)]. \qquad \text{(II.3)}$$

Comparison of Eq. (II.3) with formula (7) enables one to write

$$\Phi_m(\alpha) = \frac{\pi}{36} \times [\sin\alpha(20+\cos^2\alpha) - 3\alpha\cos\alpha(7+2\sin^2\alpha)].$$

REFERENCES

1. P. de Gennes, *The Physics of Liquid Crystals* (Clarendon, Oxford, 1974; Mir, Moscow, 1977).
2. A. A. Vedenov and E. B. Levchenko, Usp. Fiz. Nauk **141**, 3 (1983) [Sov. Phys. Usp. **26**, 747 (1983)].
3. A. S. Sonin, Usp. Fiz. Nauk **153**, 273 (1987) [Sov. Phys. Usp. **30**, 875 (1987)].
4. A. S. Vasilevskaya, É. V. Generalova, and A. S. Sonin, Usp. Khim. **58**, 1575 (1989).
5. A. S. Sonin, Kolloidn. Zh. **60**, 149 (1998).
6. A. S. Sonin, J. Mater. Chem. **8**, 2557 (1998).
7. P. Davidson, C. Bourgaux, L. Schoutteten, *et al.*, J. Phys. II **5**, 1577 (1995).
8. J.-C. P. Gabriel and P. Davidson, Adv. Mater. **12**, 9 (2000).
9. A. V. Kaznacheev, A. Yu. Kovalevskiĭ, I. A. Ronova, *et al.*, Kolloidn. Zh. **62**, 606 (2000).
10. É. V. Generalova, A. V. Kaznacheev, and A. S. Sonin, Kristallografiya **46**, 121 (2001) [Crystallogr. Rep. **46**, 111 (2001)].
11. E. Guyon, R. B. Meyer, and J. Salan, Mol. Cryst. Liq. Cryst. **54**, 261 (1979).
12. S. Fraden, A. J. Hurd, R. B. Meyer, *et al.*, J. Phys. Colloq. **46** (3), C3-85 (1985).
13. H. Zocher and K. Jacobsohn, Kolloidchem. Beih. **28**, 167 (1929).
14. M. V. Kurik and O. D. Lavrentovich, Usp. Fiz. Nauk **154**, 381 (1988) [Sov. Phys. Usp. **31**, 196 (1988)].
15. W. Biltz, Ber. Dtsch. Chem. Ges. **37**, 109 (1904).
16. G. B. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1966; Atomizdat, Moscow, 1970).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
18. W. H. de Jen, *Physical Properties of Liquid Crystalline Materials* (Gordon and Breach, New York, 1980; Mir, Moscow, 1982).
19. A. J. Hurd, S. Fraden, F. Lonberg, *et al.*, J. Phys. (Paris) **46**, 905 (1985).
20. J. Bernal and J. Fankuchen, Nature **139**, 923 (1937).
21. A. V. Kaznacheev, K. Praefcke, A. S. Sonin, and N. V. Usol'tseva, Kolloidn. Zh. **64**, 1 (2002).
22. L. M. Blinov, E. I. Kats, and A. A. Sonin, Usp. Fiz. Nauk **152**, 449 (1987) [Sov. Phys. Usp. **30**, 604 (1987)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 64–66.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 76–78.
Original Russian Text Copyright © 2002 by Stishov.

FLUIDS

On Phase Diagram of the System of “Collapsing” Hard Spheres

S. M. Stishov
Institute of High-Pressure Physics, Russian Academy of Sciences,
Troitsk, Moscow oblast, 142190 Russia
Los Alamos National Laboratory, Los Alamos, New Mexico, 87545 USA
e-mail: sergei@hppi.troitsk.ru

Received February 26, 2002

Abstract—The basic phase diagram of a system with a repulsive step interaction potential is constructed on the basis of simple physical arguments by using the properties of a system of hard spheres as the starting point. Various versions of the behavior of the melting curve are discussed. The possibility of a phase transition in the liquid phase is indicated. *© 2002 MAIK “Nauka/Interperiodica”.*

We will refer to a system of particles interacting through a step potential of the form (Fig. 1b)

$$\begin{aligned}\Phi(r) &= \infty, \quad 0 \le r < \sigma,\\ \Phi(r) &= \varepsilon, \quad \sigma \le r < \sigma',\\ \Phi(r) &= 0, \quad \sigma' \le r,\end{aligned} \tag{1}$$

as the system of “collapsing” hard spheres. Systems of this type are studied in connection with anomalous melting curves, isomorphic phase transitions, transformations in colloidal systems, etc. (see, for example, [1–3]).

Let us consider basic properties of a standard system of hard spheres, i.e., a system with an interaction of the form (Fig. 1a)

$$\begin{aligned}\Phi(r) &= \infty, \quad 0 \le r < \sigma,\\ \Phi(r) &= 0, \quad \sigma \le r.\end{aligned} \tag{2}$$

The system of hard spheres is described by three parameters: temperature T and two characteristic lengths σ and l, viz., the hard sphere diameter and the mean distance between particles, which is equal to $(V/N)^{1/3}$ (V is the volume of the system and N is the number of particles). These quantities make it possible to obtain two dimensionless combinations: PV_0/NkT and V/V_0 (P is the pressure and $V_0 = N\sigma^3/\sqrt{2}$), which are obviously connected through the dependence

$$PV_0/NkT = f(V/V_0). \tag{3}$$

Melting takes place in the system for $l/\sigma = \text{const}$. Accordingly, relation (3) leads to the equation for the melting curve

$$P = \frac{c}{\sigma^3}T, \tag{4}$$

where $c \approx 12$ [4]. It should be emphasized that the phase-equilibrium curves for systems of hard particles are always straight lines emerging from the origin.

Let us now analyze a collapsing system. It can be seen from Fig. 1b that a system of collapsing hard spheres is characterized by three lengths, $l = (V/N)^{1/3}$, σ, and σ' and two energy parameters, viz., ε and T (T is the temperature).

In the two limiting cases, $T \gtrsim \varepsilon$, $l > \sigma'$, and $T \lesssim \varepsilon$, the system apparently behaves as a simple system of hard spheres of diameter σ' or σ. The low-temperature and high-temperature branches of the melting curve are described by expressions of type (4). A transition between the two branches of the melting curve occurs at $T \approx \varepsilon$; in this case, we can expect the emergence of a more or less strong anomaly on the melting curve up to a temperature peak (or a triple point associated with a phase transition in the liquid) depending on the ratio σ/σ'. The equation for the melting curve of the collapsing system can be written in the form of the superposition

$$P = \alpha\frac{cT}{\sigma'^3} + \beta\frac{cT}{\sigma^3}, \tag{5}$$

or

$$P = \frac{cT}{\sigma'^3}\left(\alpha + \beta\frac{\sigma'^3}{\sigma^3}\right), \tag{6}$$

where $\alpha = f(\varepsilon/T)$, $\beta = \varphi(\varepsilon/T)$, and $\alpha + \beta = 1$; by definition, $\alpha \longrightarrow 1$ and $\beta \longrightarrow 0$ for $\varepsilon/T \longrightarrow \infty$, while $\alpha \longrightarrow 0$ and $\beta \longrightarrow 1$ for $\varepsilon/T \longrightarrow 0$. However, the peculiarities of the transition region are mainly determined by the behavior of temperature-dependent coefficients α and β for $\varepsilon/T \approx 1$ and by the magnitude of the ratio σ'/σ. Figure 2 shows different scenarios of the behavior of the melting curve for a collapsing system. It should be

1063-7761/02/9501-0064$22.00 © 2002 MAIK “Nauka/Interperiodica”

emphasized that the features of the melting curve as well as a number of other specific properties of a collapsing system are determined by exciting the particles to a state with energy $\geq\varepsilon$ and with diameter σ upon an increase in temperature and pressure in accordance with the interaction potential (see Fig. 1b). This process occurring in a condensed system is correlated to a certain extent, which, in turn, may lead to its localization in a certain region in the P–V–T space.[1]

The versions of the behavior of the melting curve presented in Fig. 2 correspond to various extents of correlation upon a transition of particle to a state with energy $\geq\varepsilon$. The phase transition in the liquid (line *9* in Fig. 2) and melting lines *8* correspond to the extreme correlation case, while a weaker correlation leads to curves of the type *5*, *6*, and *7* (see Fig. 2).

It should be noted that the phase state of a substance is significant for a transition of particles to an excited state. In particular, the inversion of the density jump, which necessarily takes place in the case of a melting curve *5* with a peak (Fig. 2), indicates that a transition of particles to a state with energy $\geq\varepsilon$ in the liquid must occur earlier (at a lower pressure) than in the crystal phase.

The situation with relatively weak anomalies on the melting curve of type *6* or *7* (see Fig. 2) takes place for $\Delta\sigma/\sigma < 0.1$ (here, $\Delta\sigma = \sigma' - \sigma$) [3]. The peak on the melting curve *5* (Fig. 2) apparently occurs for $\Delta\sigma/\sigma \geq 0.1$ [1, 2]. The scenario with the first-order phase transition in the liquid (line *8* in Fig. 2) is hardly probable for this type of system, although, according to Franzese *et al.* [5], this scenario is realized in a collapsing system with attraction. However, generally speaking, the correlation effects must increase with the ratio $\Delta\sigma/\sigma$ upon a transition of the system to the excited state, and it cannot be ruled out that a detailed analysis of the liquid phase of the collapsing system for $\Delta\sigma/\sigma \geq 1.5$ might lead to new results.[2]

It should be remarked further that the form of the interaction potential in a collapsing system suggests that an isostructural phase transition may exists in the crystal phase, which is indeed observed in computer experiments[3] [2, 3]. Let us derive the expression for the pressure corresponding to the expected transition at $T = 0$.

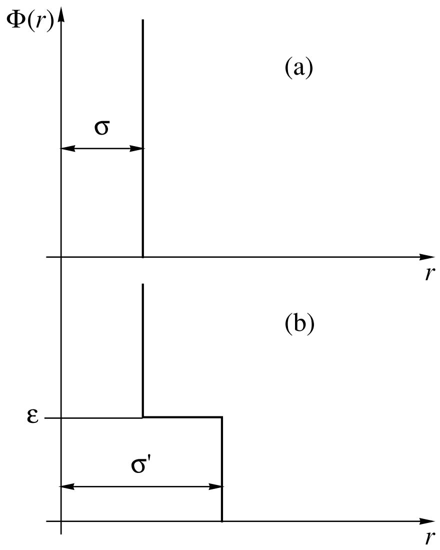


Fig. 1. (a) Potential of hard spheres and (b) step potential.

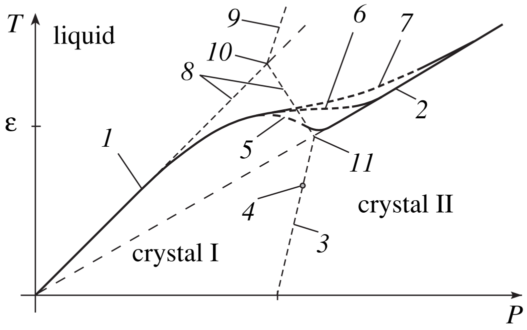


Fig. 2. Phase diagram of a system of collapsing hard spheres: curves *1* and *2* are the melting curves corresponding to the equations $P = cT/\sigma'^3$ and $P = cT/\sigma^3$; line *3* corresponds to an isostructural phase transition and may terminate at the critical point *4*; curves *5*–*8* describe the versions of the behavior of the melting curve in the transition region; curve *9* describes a possible first-order phase transition in the liquid phase (it can be seen that line *9* and two segments of the melting line *8* form the triple point *10*); *11* is the triple point emerging as a result of intersection of the line of isostructural phase transition with the melting curve; if the isostructural transition terminates at the critical point, no triple point is formed.

From the equality of the chemical potentials at a phase transition, we have

$$E_1 + PV_1 = E_2 + PV_2, \qquad (7)$$

where $V_1 = N(\sigma')^3/\sqrt{2}$, $V_2 = N\sigma^3/\sqrt{2}$, $E_1 = 0$, and $E_2 = 6N\varepsilon$. Equation (7) leads to the following expression for the transition pressure:

$$P = \frac{8.48\varepsilon}{(\Delta V/V_1)(\sigma')^3}. \qquad (8)$$

---

[1] In a rarefied system, i.e., for $l \gtrsim \sigma'$ (in the case of infrequent pair collisions), this process does not cause strong changes and leads to an evolutionary variation of the effective diameter of particles. In a condensed system ($l \approx \sigma'$), multiple collisions make possible the formation of clusters consisting of particles of diameter σ. The volume effect emerging in this case facilitates the cluster growth, which may ultimately induce a phase transition in the system.

[2] It should be noted in this connection that all (although scarce) results of investigations [6, 7] indicate that phase transitions in liquids are accompanied by considerable changes in the volume (~50%).

[3] Here and in the literature, the term “isostructural phase transition” is applied to a first-order phase transition which is not accompanied by a change in the crystal lattice symmetry.

It is interesting to compare the value of this quantity with a certain average pressure corresponding to the transition between the two branches of the melting curve. It was assumed above that this transition occurs at temperature $T \approx \varepsilon$. Substituting this estimate into Eq. (4) and replacing σ by σ', we obtain the following estimate for the transition pressure:

$$P \approx \frac{12}{(\sigma')^3}\varepsilon. \tag{9}$$

For the sake of comparison, we substitute into Eq. (8) the value of the quantity $\Delta V/V \approx 0.20$ corresponding to the isostructural transition in cerium [8]. This gives

$$P \approx \frac{42}{(\sigma')^3}\varepsilon. \tag{10}$$

It should be emphasized that the pressure corresponding to the phase transition in the solid phase for reasonable values of the volume jump considerably exceeds estimate (9) for the pressure corresponding to the transition between the two branches of the melting curve. Although this is in qualitative agreement with our predictions, such a strong discrepancy between the numerical factors in formulas (9) and (10) probably indicates that the estimate $T \approx \varepsilon$ is not quite accurate.

Thus, using the properties of the system of hard spheres, we have proved that the phase diagram of a collapsing system or a system of particles interacting through a repulsive step potential possesses or may possess specific features the most remarkable of which is the anomalous behavior of the melting curve. The melting point of such a system may pass through a peak upon a change in pressure. First-order phase transitions in the liquid and solid phases are also expected features of the system, although the phase transition in the liquid may turn out to be unobservable (i.e., may correspond to a metastable region).

REFERENCES

1. E. Brindeau, R. Levant, and J.-P. Hansen, Phys. Lett. A **60A**, 424 (1977).
2. D. A. Young and B. J. Alder, Phys. Rev. Lett. **38**, 1213 (1977).
3. P. Bolhuis and D. Frenkel, J. Phys.: Condens. Matter **9**, 381 (1997).
4. W. G. Hoover, S. G. Gray, and K. W. Johnson, J. Chem. Phys. **55**, 1128 (1971).
5. G. Franzese, G. Malesclo, A. Skibinsky, *et al.*, Nature **409**, 693 (2001).
6. J. N. Glosli and F. N. Ree, Phys. Rev. Lett. **82**, 4659 (1999).
7. Y. Katayama, T. Mizutani, W. Utsumi, *et al.*, Nature **403**, 170 (2000).
8. D. A. Young, *Phase Diagrams of the Elements* (Univ. of California Press, Berkeley, 1991).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 67–76.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 79–89.
Original Russian Text Copyright © 2002 by Teodorovich.

Renormalization Group Method in the Problem of the Effective Conductivity of a Randomly Heterogeneous Porous Medium

E. V. Teodorovich

Institute for Problems in Mechanics, Russian Academy of Sciences, pr. Vernadskogo 101, Moscow, 119526 Russia
e-mail: teodor@ipmnet.ru

Received December 21, 2001

Abstract—Within the field formulation of the renormalization group (RG) method, an expression for the Fourier transform of the integral kernel of the effective conductivity of a randomly heterogeneous porous medium is obtained. In the large-scale limit, the result obtained corresponds to the well-known Landau–Lifshits–Matheron formula. The validity of the technical tricks applied to the calculations performed within the RG approach is discussed. *© 2002 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

To describe a fluid flow through a porous medium, one usually applies Darcy’s law, which is obtained by averaging over microscales corresponding to the characteristic sizes of the pores. Darcy’s law establishes a relation between the seepage velocity $v_i(\mathbf{r})$ and the pressure gradient $\partial_i p(\mathbf{r})$ and for an isotropic medium is expressed as

$$v_i(\mathbf{r}) = -\kappa\partial_i p(\mathbf{r}), \qquad (1.1)$$

where κ is the conductivity of the porous medium.

In a randomly heterogeneous medium, the conductivity is a certain random function of coordinates, $\kappa = \kappa(\mathbf{r})$; henceforth, we assume that the statistics of realizations of the field $\kappa(\mathbf{r})$ are specified. In practice, the problem of interest is the relation between the seepage velocity and the pressure gradient, both averaged over the ensemble of realizations. In the general case, this relation is nonlocal and is expressed as

$$\langle v_i(\mathbf{r})\rangle = -\int d\mathbf{r}' K(\mathbf{r}, \mathbf{r}')\langle\partial_i p(\mathbf{r}')\rangle, \qquad (1.2)$$

and a task of the theory is to calculate the integral kernel $K(\mathbf{r}, \mathbf{r}')$ on the basis of the known statistics of realizations of the random field $\kappa(\mathbf{r})$.

In a statistically homogeneous system, the kernel $K(\mathbf{r}, \mathbf{r}')$ is a function of the difference of coordinates $\mathbf{r} - \mathbf{r}'$ and is appreciably different from zero only in a certain region $|\mathbf{r} - \mathbf{r}'| \le l$, where l is the correlation length. If the characteristic scale in which the mean values of the seepage velocity and pressure gradient vary significantly is small as compared with the correlation length (the large-scale limit), then one can assume that

$$K(\mathbf{r} - \mathbf{r}') \approx K_{\text{eff}}\delta(\mathbf{r} - \mathbf{r}'), \quad K_{\text{eff}} = \int d\mathbf{r} K(\mathbf{r}). \qquad (1.3)$$

In this case, the problem reduces to the calculation of K_{eff}, the so-called effective conductivity of a randomly heterogeneous medium.

Consider an infinite medium with a given deterministic (regular) density of fluid sources $\rho(\mathbf{r})$. From the incompressibility condition, we obtain the equation

$$\partial_i v_i(\mathbf{r}) = \rho(\mathbf{r}), \qquad (1.4)$$

which, with regard to Darcy’s law (1.1), leads to the following differential equation for pressure:

$$\partial_i[\kappa(\mathbf{r})\partial_i p(\mathbf{r})] = -\rho(\mathbf{r}). \qquad (1.5)$$

Thus, to calculate the effective conductivity, one has to solve first the stochastic differential equation for pressure (1.5); then, using Darcy’s law, determine the seepage velocity; and finally, applying the averaging procedure over the ensemble of conductivity field realizations, find the mean values of the seepage velocity and the pressure gradient. This enables one to calculate the effective conductivity.

To solve the stochastic differential equation (1.5) approximately, one usually applies the method of perturbation theory. In this case, a solution is constructed as a series in powers of conductivity fluctuations $\delta\kappa(\mathbf{r}) = \kappa(\mathbf{r}) - \langle\kappa(\mathbf{r})\rangle$. The subsequent term-by-term averaging of the series for the seepage velocity and the pressure gradient enables one to get the expression for the effective conductivity in the form of a series in powers of the conductivity variance. However, actually, one has to restrict oneself to low-order approximations of perturbation theory due to the increasing difficulties associated with the passage to higher order approximations. Regarding the estimation of the role of higher order approximations, in [1], an assumption was made that the low-order approximations of perturbation theory represent the first terms in the Taylor series expansion of the exponential function and that the dependence of

K_{eff} on the log-conductivity variance σ^2 in the space of dimension d has the form

$$K_{\text{eff}} = K_G \exp\left(\frac{d-2}{2d}\sigma^2\right), \tag{1.6}$$

where K_G is the geometric mean of the conductivity. This formula is exact in the one-dimensional case when $K_{\text{eff}} = K_H$ (K_H is the harmonic mean of the conductivity, $K_H = K_G\exp(-\sigma^2/2)$ for a lognormal distribution). In the two-dimensional case, Matheron obtained the formula $K_{\text{eff}} = K_G$ for the lognormal distribution. In the three-dimensional case, Landau and Lifshits [2], on the basis of general phenomenological considerations, proposed the formula $K_{\text{eff}} = \langle\kappa^{1/3}\rangle^3$. All these cases for the lognormal distribution agree with hypothesis (1.6), which is called the Landau–Lifshits–Matheron formula. Calculations within perturbation theory carried out taking into account the terms of order σ^4 turn out to be in agreement with the Landau–Lifshits–Matheron formula; however, it was pointed out that, in a number of cases, this formula proves not to be valid for terms of order σ^6 (see a more detailed survey [3]). At present, the substantiation (or refutation) of the Landau–Lifshits–Matheron formula is the subject of numerous investigations.

2. PERTURBATION THEORY AND DIAGRAMMATIC TECHNIQUE

In the conventional method, one first constructs a perturbation theory series for pressure by solving Eq. (1.5) and then finds a series for the seepage velocity with the use of Darcy's law. In contrast to this approach, in the present study, we first construct a series for the velocity of the filtration flow and then apply Darcy's law to construct a series for pressure. In this approach, an expression for the velocity is obtained naturally in the form of expansion in powers of $u(\mathbf{r}) = \ln[\kappa(\mathbf{r})/K_G]$ and $\langle u(\mathbf{r})\rangle = 0$.

Let us represent the velocity field as a superposition of the potential and solenoidal parts:

$$\begin{gathered} v_i(\mathbf{r}) = v_i^p(\mathbf{r}) + v_i^s(\mathbf{r}), \\ \partial_i v_j^p(\mathbf{r}) - \partial_j v_i^p(\mathbf{r}) = 0, \quad \partial_i v_i^s(\mathbf{r}) = 0. \end{gathered} \tag{2.1}$$

In this case, Eq. (1.4) only determines the potential part

$$v_i^p(\mathbf{r}) = (\partial)_i^{-1}\rho(\mathbf{r}) = \partial_i\Delta^{-1}\rho(\mathbf{r}), \tag{2.2}$$

where $\Delta^{-1} = G^{(0)}(\mathbf{r})$ is the Green's function for the Laplace operator, which is a solution to the equation

$$\Delta G^{(0)}(\mathbf{r}) = \delta(\mathbf{r}). \tag{2.3}$$

To determine the equation for the solenoidal part of the seepage flow velocity v_i^s, we calculate $\partial_j[\partial_i v_j(\mathbf{r}) - \partial_j v_i(\mathbf{r})]$. Taking into account relations (2.1) and Darcy's law (1.1), we find

$$\begin{aligned} \partial_j[\partial_i v_j(\mathbf{r}) - \partial_j v_i(\mathbf{r})] &= -\partial_j\partial_j v_i^s(\mathbf{r}) \\ &= -\partial_j[\partial_i\kappa(\mathbf{r})\partial_j p(\mathbf{r}) - \partial_j\kappa(\mathbf{r})\partial_i p(\mathbf{r})]. \end{aligned}$$

Applying again Darcy's law to eliminate pressure, we obtain

$$\begin{aligned} \partial_j\partial_j v_i^s(\mathbf{r}) &= -[\partial_k u_k(\mathbf{r})\delta_{ij} - \partial_j u_i(\mathbf{r})]v_j(\mathbf{r}), \\ u_i(\mathbf{r}) &= \partial_i u(\mathbf{r}). \end{aligned} \tag{2.4}$$

Passing from differential to integral equations, we rewrite the equation for the velocity components of the filtration flow as

$$\begin{aligned} v_i(\mathbf{r}) = v_i^p(\mathbf{r}) - \int d\mathbf{r}' G^{(0)}(\mathbf{r} - \mathbf{r}') \\ \times [\partial'_k u_k(\mathbf{r}')\delta_{ij} - \partial'_j u_i(\mathbf{r}')]v_j(\mathbf{r}'). \end{aligned} \tag{2.5}$$

This equation can also be represented in the form corresponding to the Martin–Siggia–Rose formalism [4]:

$$\begin{aligned} v_i(\mathbf{r}) = v_i^p(\mathbf{r}) - \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 G^{(0)}(\mathbf{r} - \mathbf{r}_1) \\ \times \Gamma_{ij}^{(0)}(\mathbf{r}_1|\mathbf{r}_2, \mathbf{r}_3)u(\mathbf{r}_2)v_j(\mathbf{r}_3), \end{aligned} \tag{2.6}$$

where the tensor $\Gamma_{ij}^{(0)}$, called a vertex in what follows, is defined by the formula

$$\begin{aligned} \Gamma_{ij}^{(0)}(\mathbf{r}_1|\mathbf{r}_2, \mathbf{r}_3) = \lambda_0[(\partial^{(2)}\cdot\partial^{(1)})\delta_{ij} - \partial_i^{(2)}\partial_j^{(1)}] \\ \times \delta(\mathbf{r}_1 - \mathbf{r}_2)\delta(\mathbf{r}_1 - \mathbf{r}_3), \end{aligned} \tag{2.7}$$

where the parenthesized numbers at the differentiation operators indicate on what variable the operator acts and λ_0 is a formal parameter of expansion in a perturbation theory series (equal to one).

In the space of Fourier transforms, Eqs. (2.6) and (2.7) are rewritten as

$$\begin{aligned} \tilde{v}_i(\mathbf{q}) = \tilde{v}_i^p(\mathbf{q}) - \tilde{G}^{(0)}(\mathbf{q})\int\frac{d\mathbf{p}}{(2\pi)^d} \\ \times \tilde{\Gamma}_{ij}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p})\tilde{u}(\mathbf{p})\tilde{v}_j(\mathbf{q}-\mathbf{p}), \end{aligned} \tag{2.8}$$

$$\tilde{\Gamma}_{ij}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = \lambda_0[(\mathbf{p}\cdot\mathbf{q})\delta_{ij} - p_i q_j]. \tag{2.9}$$

Note that the vertex $\Gamma_{ij}^{(0)}$ possesses the properties of the transverse projection operator:

$$\begin{gathered} q_i\tilde{\Gamma}_{ij}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = 0, \\ P_{ik}(\mathbf{q})\tilde{\Gamma}_{kj}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = \tilde{\Gamma}_{ij}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}), \\ P_{ij}(\mathbf{q}) = \delta_{ij} - \frac{q_i q_j}{q^2}, \quad \tilde{\Gamma}_{ij}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p})p_j = 0. \end{gathered} \tag{2.10}$$

The application of the iteration method enables us to write the solution to the integral equation in the form of

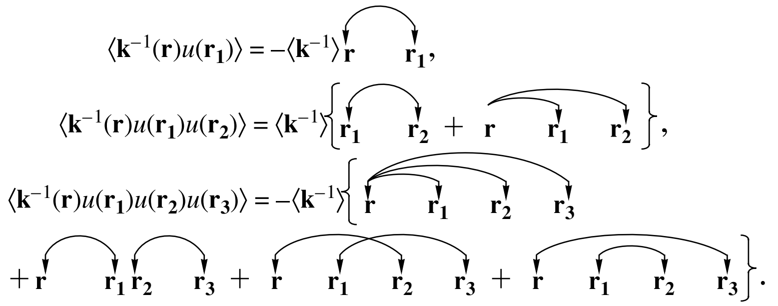


Fig. 1. Diagrammatic representations of mixed means of different orders.

a perturbation theory series in powers of the formal expansion parameter λ_0; this series represents a functional of a random field $u(\mathbf{r})$. The averaging of the functional series leads to the problem of calculating the statistical moments of the form $\langle u(\mathbf{r}_1)\ldots u(\mathbf{r}_n)\rangle$. In the case of a centered normal distribution, these means are represented as sums of all possible covariance functions for the even-order moments and vanish for odd-order moments (the Wick theorem).

In a statistically homogeneous system, we have the following relation for the covariance functions:

$$\langle u(\mathbf{r}_1)u(\mathbf{r}_2)\rangle = B(\mathbf{r}_1 - \mathbf{r}_2). \qquad (2.11)$$

To abridge the expressions and facilitate the analysis of high-order approximations, we introduce rules for assigning certain graphical symbols, called the Feynman diagrams, to analytic expressions. These rules are formulated as follows.

1. The Green's function $G^{(0)}(\mathbf{r}_1 - \mathbf{r}_2)$ of an unperturbed system (propagator) is assigned a thin horizontal line outgoing from the point $\mathbf{r}_2$ and incoming to the point $\mathbf{r}_1$.

2. The function $u(\mathbf{r})$ is assigned an arrow directed vertically to the point $\mathbf{r}$.

3. The function $\Gamma_{ij}^{(0)}(\mathbf{r}_1|\mathbf{r}_2, \mathbf{r}_3)$ (a bare vertex) is assigned a triangle such that the horizontal arrows corresponding to the Green's functions are incoming to its right vertex $\mathbf{r}_3$ and outgoing from the left vertex $\mathbf{r}_1$, and the vertical arrow corresponding to $u(\mathbf{r}_2)$ is incoming to the middle vertex $\mathbf{r}_2$.

4. The pair correlation function $B(\mathbf{r}_1 - \mathbf{r}_2)$ (correlator) is assigned a line with two arrows directed vertically to the points $\mathbf{r}_1$ and $\mathbf{r}_2$.

5. The full Green's function $G(\mathbf{r}_1 - \mathbf{r}_2)$ averaged over the ensemble of realizations (a full propagator) is assigned a heavy horizontal line directed from the point $\mathbf{r}_2$ to the point $\mathbf{r}_1$.

6. The full vertex $\Gamma_{ij}(\mathbf{r}_1|\mathbf{r}_2, \mathbf{r}_3)$ averaged over the ensemble of realizations is assigned a dark triangle with two incoming and one outgoing arrows.

7. Integration is performed over the coordinates of all points corresponding to the vertices of the triangles.

When passing to the wavenumber space (the space of Fourier transforms), the structure of the diagrams remains unchanged; however, the lines of correlators and propagators in this case are assigned a wavenumber, and these quantities correspond to the Fourier transforms $\tilde{B}(\mathbf{q})$ and $\tilde{G}_{ij}^{(0)}(\mathbf{q})$, while a triangular vertex is assigned $\tilde{\Gamma}^{(0)}(\mathbf{q}|\mathbf{p}, \mathbf{q} - \mathbf{p})$; in this case, the algebraic sum of incoming and outgoing wavenumbers is equal to zero.

The calculation of the mean gradient of pressure by the formula

$$\begin{aligned}\langle \partial_i p(\mathbf{r})\rangle &= -\langle \kappa^{-1}(\mathbf{r}) v_i(\mathbf{r})\rangle \\ &= -\langle \kappa^{-1}\rangle v_i^p(\mathbf{r}) - \langle \kappa^{-1}(\mathbf{r}) v_i^s(\mathbf{r})\rangle\end{aligned} \qquad (2.12)$$

gives rise to mixed means of the form

$$\begin{aligned}&\langle \kappa^{-1}(\mathbf{r})u(\mathbf{r}_1)\ldots u(\mathbf{r}_n)\rangle \\ &= K_G^{-1}\langle e^{-u(\mathbf{r})}u(\mathbf{r}_1)\ldots u(\mathbf{r}_n)\rangle.\end{aligned} \qquad (2.13)$$

The calculation of the mixed means is performed in Appendix. According to (A.7), these means correspond to the sum of products of all possible covariance functions of the form $\langle u(\mathbf{r})u(\mathbf{r}_k)\rangle$ and $\langle u(\mathbf{r}_k)u(\mathbf{r}_l)\rangle$, multiplied by $(-1)^n\langle\kappa^{-1}\rangle$. To the corresponding quantities, we assign diagrams containing a large light circle (the so-called scalar vertices [5]) at which an arbitrary number of lines of correlators can arrive whose other ends are connected to the triangular vertices. Note that, in contrast to [5], there are no lines of correlators connected only to a scalar vertex in the formalism used. This is associated with the fact that we did not use the Taylor series expansion of the exponential function when calculating the statistical moments of the form (2.13). The diagrams for a few low-order moments of the form (2.13) calculated according to (A.7) in the Appendix are shown in Fig. 1.

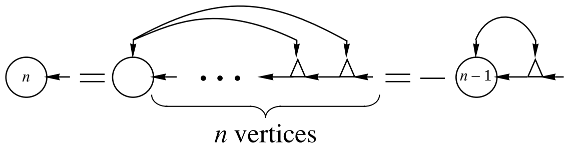


Fig. 2. Diagrammatic representations of $H_{ij}^{(n)}$.

3. EQUATION FOR THE EFFECTIVE CONDUCTIVITY

The application of the iteration method to solving Eq. (2.5) yields the following representation for the solution:

$$v_i(\mathbf{r}) = \int d\mathbf{r}' R_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n)) v_j^p(\mathbf{r}'). \tag{3.1}$$

here, the resolvent kernel R_{ij} is a functional of the random field $u(\mathbf{r}_n)$ and is given by

$$\begin{aligned} R_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n)) &= \delta_{ij}\delta(\mathbf{r}-\mathbf{r}') + T_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n)), \\ T_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n)) &= \int d\mathbf{r}'' P_{ij}(\mathbf{r}-\mathbf{r}'') T(\mathbf{r}'', \mathbf{r}'|u(\mathbf{r}_n)). \end{aligned} \tag{3.2}$$

The averaging of Eq. (3.2) over the realizations of the random field $u(\mathbf{r})$ under the assumption that the system is statistically homogeneous yields

$$\langle T(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n))\rangle = T(\mathbf{r}-\mathbf{r}'); \tag{3.3}$$

i.e., the solution can be represented as a convolution of the potential vector $v_i^p(\mathbf{r})$ and the expression containing the transverse projection operator; such forms identically vanish. Hence, we obtain

$$\langle v_i(\mathbf{r})\rangle = v_i^p(\mathbf{r}) = \partial_i \Delta^{-1}\rho(\mathbf{r}). \tag{3.4}$$

Thus, to determine the integral kernel $K_{ij}(\mathbf{r}-\mathbf{r}')$, one should calculate, using Darcy's law, the pressure gradient by the formula

$$\langle \partial_i p(\mathbf{r})\rangle = -\int d\mathbf{r}' \langle \kappa^{-1}(\mathbf{r}) R_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n))\rangle v_j^p(\mathbf{r}'); \tag{3.5}$$

this leads to the relation

$$\begin{aligned} K_{ij}^{-1}(\mathbf{r}-\mathbf{r}') &= H_{ij}(\mathbf{r}-\mathbf{r}') = \langle \kappa^{-1}(\mathbf{r}) R_{ij}(\mathbf{r}, \mathbf{r}')\rangle \\ &= \langle \kappa^{-1}(\mathbf{r})\rangle \delta(\mathbf{r}-\mathbf{r}') + K_G^{-1}\langle e^{-u(\mathbf{r})} T_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n))\rangle. \end{aligned} \tag{3.6}$$

In turn, the kernel R_{ij} is a solution to the integral equation

$$\begin{aligned} R_{ij}(\mathbf{r}, \mathbf{r}'|u(\mathbf{r}_n)) &= \delta_{ij}\delta(\mathbf{r}-\mathbf{r}') - \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 \\ &\times G^{(0)}(\mathbf{r}-\mathbf{r}_1)\Gamma_{il}^{(0)}(\mathbf{r}_1|\mathbf{r}_2, \mathbf{r}_3) u(\mathbf{r}_2) R_{lj}(\mathbf{r}_3, \mathbf{r}'|u(\mathbf{r}_n)), \end{aligned} \tag{3.7}$$

which follows from (2.6) and (3.1). The iterative solution of Eq. (3.7) yields a representation for R_{ij} in terms of power series in λ. The substitution of this series into (3.6), followed by the term-by-term averaging of the series obtained, yields an expression for $H_{ij}(\mathbf{r}-\mathbf{r}')$ in the form of an infinite series each of whose terms can be assigned an appropriate Feynman diagram.

We will carry out the further analysis in the space of Fourier transforms. Denote by $H_{ij}^{(n)}(\mathbf{q})$ the contribution, taken with the sign $(-1)^n$, of the diagram containing n triangular vertices and a scalar vertex with n incoming lines of correlators whose other ends meet triangular vertices (when determining $H_{ij}^{(n)}$, we took into account that, according to (A.7), the odd-order statistical moments are expressed in terms of pair correlators taken with the minus sign). The corresponding diagram is shown in Fig. 2. The form of this diagram clearly shows that there exists the following recurrence relation between the functions $H_{ij}^{(n)}$:

$$\begin{aligned} H_{ij}^{(n)}(\mathbf{q}) &= -\int \frac{d\mathbf{p}}{(2\pi)^d} H_{il}^{(n-1)}(\mathbf{q}-\mathbf{p}) \\ &\times G^{(0)}(\mathbf{q}-\mathbf{p})\Gamma_{lj}^{(0)}(\mathbf{q}-\mathbf{p}|-\mathbf{p}, \mathbf{q}) B(\mathbf{p}). \end{aligned} \tag{3.8}$$

Let us introduce the function

$$H'_{ij}(\mathbf{q}) = \sum_{n=0}^{\infty} H_{ij}^{(n)}(\mathbf{q}). \tag{3.9}$$

The recurrence relations (3.8) imply the equation

$$\begin{aligned} H'_{ij}(\mathbf{q}) &= H_{ij}^{(0)}(\mathbf{q}) - \int \frac{d\mathbf{p}}{(2\pi)^d} H'_{il}(\mathbf{q}-\mathbf{p}) \\ &\times \tilde{G}^{(0)}(\mathbf{q}-\mathbf{p})\tilde{\Gamma}_{lj}^{(0)}(\mathbf{q}-\mathbf{p}|-\mathbf{p}, \mathbf{q})\tilde{B}(\mathbf{p}), \\ H_{ij}^{(0)} &= \delta_{ij}\langle \kappa^{-1}\rangle, \end{aligned} \tag{3.10}$$

which can be referred to as the Dyson equation for the tensor function H'_{ij}. If we represent H'_{ij} by a large dark circle, then the Dyson equation is represented by the diagrammatic relation depicted in Fig. 3a.

The contributions of all the remaining diagrams are obtained by inserting all possible subdiagrams containing the lines of correlators that connect only triangular vertices. A large part of these diagrams determine the corrections to the Green's function and to the vertex, and the consideration of these diagrams is reduced to the replacement of the bare Green's functions $G^{(0)}$ and vertex $\Gamma^{(0)}$ in (3.10) by their exact (renormalized) values G and Γ. As a result, the Dyson equation for the tensor H_{ij} can be represented as

$$\begin{aligned} H_{ij}(\mathbf{q}) &= H_{ij}^{(0)} - \int \frac{d\mathbf{p}}{(2\pi)^d} H_{il}(\mathbf{q}-\mathbf{p}) \\ &\times \tilde{G}(\mathbf{q}-\mathbf{p})\tilde{\Gamma}_{lj}(\mathbf{q}-\mathbf{p}|-\mathbf{p}, \mathbf{q})\tilde{B}(\mathbf{p}). \end{aligned} \tag{3.11}$$

The corresponding diagrammatic equation is shown in Fig. 3b.

Note that, in an isotropic system, the second-rank tensor is given by

$$H_{ij}(\mathbf{q}) = \delta_{ij}H(q) + \frac{q_j q_j}{q^2}H_1(q). \qquad (3.12)$$

In view of property (2.10), only the term proportional to the Kronecker δ is preserved in the integrand of (3.11).

To determine the full Green's function and a full vertex, one has to solve a system of equations for these functions constructed in a certain perturbation theory approximation. These equations are constructed in a usual manner, and the renormalization group (RG) method proves to be very efficient for solving these equations simultaneously. In the next section, we present the solution to these equations and Eq. (3.11) by the RG method.

4. APPLICATION OF THE RENORMALIZATION GROUP METHOD

The RG method initially appeared in quantum field theory as a method for improving a perturbation theory series by summing a certain infinite subsequence of the full series of perturbation theory under the requirement that the series should be invariant under the RG transform [6]. Somewhat later, a new direction and a new formulation of the RG method appeared that were initiated by Wilson's works [7, 8], according to which the RG method represents a method for the asymptotic analysis of an essentially multimode nonlinear system by successively reducing the number of modes considered by averaging over a part of fast modes in the equation for slow modes (Kadanoff's procedure [9]). An analogue of a similar approach can be given by a phenomenological consideration of the effect of motions of the molecular scale by introducing molecular transfer terms when passing from the kinetic to a hydrodynamic description. Another analogue is given by the Krylov–Bogolyubov method in the theory of nonlinear oscillations, in which the equation for slowly varying quantities (amplitudes and phases) is obtained by averaging over basic-frequency oscillations.

To investigate processes in random media, the RG method was first applied in [10], where the authors used the Wilson formulation of the RG method for the further improvement of perturbation theory; according to this formulation, the self-energy operator in the Dyson equation was not calculated in the low order of perturbation theory; instead, it was a solution to a certain equation corresponding to the theory of self-consistent field. This approach was further developed in a series of works of other authors [11–14]. The variant of the RG method applied by these authors involves many unjustified assumptions based on the formal borrowing of certain ideas and technical tricks of the RG method from the theory of critical phenomena. In particular, the authors of [12–14] applied the RG method to a system that did not possess the property of RG invariance and, when carrying out scale transformations, did not take into account that there is an additional parameter in the theory that violates the scale invariance; this parameter has a dimension of length and is related to the correlation length of the conductivity fluctuations. Note that this situation, associated with the unjustified use of the RG method, proves to be very widespread. In particular, a similar situation is encountered in the Yakhot–Orszag RG theory of turbulence [15]. To justify the illegitimate application of the RG method in the theory of turbulence, the authors of [15] (see also [16]) formulated the so-called ideology of ε expansion (which has nothing to do with the method of ε expansion in the theory of critical phenomena [17, 18]). According to this ideology, there is a certain magical compensation of the errors arising when one uses the approximation $\varepsilon \longrightarrow 0$ to calculate the universal constants of turbulence and the errors associated with the neglect of local interactions responsible for a cascade mechanism of energy transport along the wavenumber spectrum. A detailed critical analysis of the ideology of ε expansion in the theory of turbulence is presented in [19, 20]. The main objections are associated with the incorrect interpretation of the role of local interactions, which, according to Wilson, dominate within the RG description, and with the invalid identification of the diffusion coefficient $\tilde{K}(q, \Lambda)|_{q=0}$ $(q/\Lambda \longrightarrow 0)$, taken in the large-scale limit $q \longrightarrow 0$ and depending on the cutoff parameter in the wavenumber space Λ, with the diffusion coefficient $\tilde{K}(\mathbf{q})$ taken for $q = \Lambda$.

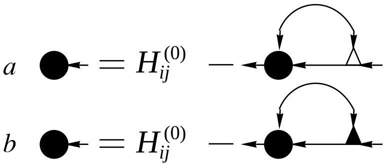


Fig. 3. Diagrammatic equations for the function $H_{ij}^{(0)}$ (*a*) corresponding to Eq. (3.10) and (*b*) corresponding to Eq. (3.11).

In [19, 20], the author of the present paper has demonstrated that the application of the field formulation of the RG method to the theory of turbulence yields correct and more general results without using unjustified hypotheses. We will follow this approach when carrying out the further analysis of our problem.

To determine $\tilde{K}(\mathbf{q})$, we have to solve a system of integral equations for the full Green's function expressed in the form

$$\tilde{G}(\mathbf{q}) = -1/\tilde{D}(q)q^2$$

and the full vertex function

$$\tilde{\Gamma}_{ij}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = \tilde{\lambda}(q)[(\mathbf{p}\cdot\mathbf{q})\delta_{ij} - p_i q_j]$$

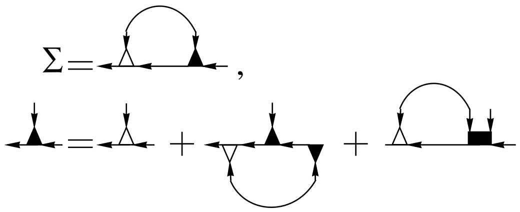


Fig. 4. Diagrammatic equations for the self-energy operator and a vertex.

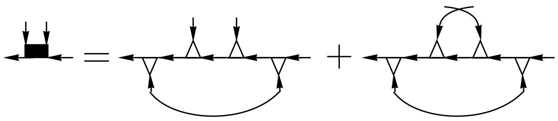


Fig. 5. Diagrams of the low-order approximation of perturbation theory with two incoming lines of correlators.

and solve Eq. (3.11) for the function $H_{ij}(\mathbf{q})$; i.e., we have to determine $\tilde{D}(q)$, $\tilde{\lambda}(q)$, $H(q)$, and $H_1(q)$. We will not write expressions for the full Green's function and the full vertex; we only present the diagrammatic representations of these equations (see Fig. 4). Notice that the left vertex in these equations is a bare vertex, although, usually, when trying to formulate a system of equations of improved perturbation theory (the skeleton diagrammatic technique), one always inserts full vertices into the diagrams of low-order approximation instead of bare vertices. However, an accurate analysis of the diagrammatic series shows that such an approach is incorrect because, in this case, part of the diagrams prove to be taken into account twice. A similar situation occurs in the theory of turbulence when constructing the Wyld diagrammatic technique [21]. This fact was first pointed out by Lee [22], who suggested that one should replace only one bare vertex (the right one) in the self-energy operator by a full vertex in order to avoid a double consideration of the same diagrams. This rule was confirmed in the theory of turbulence in a number of works by the present author [23], where the Dyson diagrammatic equations were obtained beyond the framework of perturbation theory, using the formulations of the statistical theory of turbulence on the basis of the characteristic functional. It was also demonstrated that one of the vertices (the left one) in the equation for the vertex function should not be renormalized. As applied to the problem considered, exact diagrammatic equations for the vertex have the form shown in Fig. 4, where the dark square with two incoming vertical lines contains only strongly coupled (one-particle irreducible) diagrams. In the low-order approximation of perturbation theory, these equations correspond to the diagrams shown in Fig. 5. If we neglect these diagrams, we obtain a closed system of equations for the self-energy operator and a vertex, in which the left vertex is not renormalized. Note that, when writing Eq. (3.11), we did not take into account the diagrams shown in Fig. 5.

The iterative solution of this system with the use of the values

$$\tilde{D}(q) = 1, \quad \tilde{\lambda}(q) = \lambda_0 = 1, \quad H(q) = \langle \kappa^{-1} \rangle, \quad H_1 = 0$$

as the zero-order approximation reproduces a nonrenormalized perturbation theory series. To construct a renormalized perturbation theory, we perform a renormalization using the replacements

$$\tilde{G}^{(0)}(\mathbf{q}) = -1/q^2 \longrightarrow \tilde{G}^R(\mathbf{q}) = -1/Dq^2,$$

$$\tilde{\Gamma}^{(0)}_{ij}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = \lambda_0[(\mathbf{p}\cdot\mathbf{q})\delta_{ij} - p_i q_j]$$
$$\longrightarrow \tilde{\Gamma}^R_{ij}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p}) = \lambda[(\mathbf{p}\cdot\mathbf{q})\delta_{ij} - p_i q_j]$$

in the zero-approximation terms of the equations for the Green's function and a vertex, replacing $\langle \kappa^{-1} \rangle \delta_{ij}$ on the right-hand side of Eq. (3.11) by a certain arbitrary quantity $H\delta_{ij} + H_1 q_i q_j / q^2$, and adding counterterms compensating for these replacements in the relevant equations. In addition, we require that the exact functions $\tilde{G}(\mathbf{q})$, $\tilde{\Gamma}_{ij}(\mathbf{q}|\mathbf{p}, \mathbf{q}-\mathbf{p})$, and $H_{ij}(\mathbf{q})$ coincide with the renormalized functions at the normalization point ($q = \mu$, $p = 0$), i.e.,

$$\begin{aligned} \tilde{D}(\mu) &= D, \quad \tilde{\lambda}(\mu) = \lambda, \\ H(\mu) &= H, \quad H_1(\mu) = H_1. \end{aligned} \tag{4.1}$$

Then, the iterative solution of the equations yields a renormalized perturbation theory series for the functions $\tilde{D}(q)$, $\tilde{\lambda}(q)$, $H(q)$, and $H_1(q)$. However, it turns out that one can actually calculate only the first few terms of this series. To obtain information on the properties of the full series, one should sum this series (or its infinite subsequence). The RG method is one of the methods for solving such a problem.

Assume that a pair correlation function of the conductivity logarithm has the form

$$B(\mathbf{r}) = B_0 F(mr), \quad \tilde{B}(q) = \frac{B_0}{m}\tilde{F}\left(\frac{q}{m}\right),$$

where F and $\tilde{F}$ are dimensionless functions of a dimensionless variable and m^{-1} has a dimension of length and is related to the correlation length of fluctuations of the conductivity logarithm; here, if $F(0) = 1$, then the constant B_0 is equal to the variance σ^2 of the conductivity logarithm. Then, using dimensional arguments and estimating the structure of the terms of perturbation theory series, we can represent the required functions $\tilde{D}(q)$, $\tilde{\lambda}(q)$, $H(q)$, and $H_1(q)$ as

$$\begin{aligned}
\tilde{D}(q) &= D f_1\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D^2}\right),\\
\tilde{\lambda}(q) &= \lambda f_2\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D^2}\right),\\
H(q) &= H f_3\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D}\right),\\
H_1(q) &= H_1 f_4\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D}\right).
\end{aligned} \tag{4.2}$$

The RG invariance of the renormalized perturbation theory series consists in the requirement that this series should not depend on the choice of the renormalization point μ and the set of values of the renormalized parameters D, λ, H, and $H^{(1)}$; i.e., the functions $\tilde{D}(q)$, $\tilde{\lambda}(q)$, $H(q)$, and $H_1(q)$ should not vary under the RG transformation

$$\mu \longrightarrow \mu',\ D \longrightarrow D',\ \lambda \longrightarrow \lambda',\ H \longrightarrow H',\ H_1 \longrightarrow H_1'.$$

As a rule, real systems are not RG invariant. Wilson [11] drew attention to the fact that a specific feature of an RG-invariant system is the presence of a large number of modes of various scales whose contributions to the physical characteristics of the system are equally essential. A similar situation is realized in a system where the interaction occurs between modes of close scales, while the interaction between essentially different scales occurs through a cascade sequence of interactions between modes of intermediate scales, i.e., “interactions are local in the space of scales.” Because the contributions of the modes of different scales are approximately equal (there is no distinguished characteristic scale), the sums over modes and the corresponding integrals should contain logarithmic divergences or logarithmically depend on the cutoff scale in a system with a finite number of modes. According to Wilson [7], the presence of logarithmic divergences in the theory is the characteristic feature of an RG-invariant system.

Since not all theories are “logarithmic,” Wilson and Fisher [17] proposed the so-called method of ε expansion, according to which the analysis is carried out in a “logarithmic” theory, and a transition to the “real” theory is performed by analytic continuation, with respect to a certain parameter, of the results of the logarithmic theory to the results of the real theory. Most often, the dimension d of the space serves as a parameter with respect to which the analytic continuation is performed. For example, in the theory of critical phenomena, the logarithmic theory corresponds to $d_c = 4$. In the theory of turbulence, such a parameter is given by the exponent $n = n_c + \varepsilon$ of the powerlike correlation function of effective random forces [24]; here, the logarithmic theory corresponds to $n_c = 4 - d$, while a real theory, when a cascade mechanism of energy transfer along the wavenumber spectrum (according to the Kolmogorov theory of turbulence) is realized, corresponds to $n = -d = n_c + \varepsilon$ ($\varepsilon = -4$). The dimensional regularization in the field theory can be considered as an analogue of the ε-expansion method. Within the logarithmic theory, the actual expansion parameter proves to be proportional to ε; as a result, it becomes possible to apply perturbation theory even in the case of strong intermode interactions.

The system considered is not RG invariant; i.e., the theory does not contain logarithmic divergences. If we wish to separate the interactions that are local in the wavenumber space and use the RG-invariance properties, then, acting within the ε-expansion method, we should choose a certain type of Fourier transform of the correlation function $\tilde{F}(q/m)$. Let us choose this function so that the integrals over modes contain logarithmic divergences and that the Fourier transforms of the real and chosen correlation functions coincide at the normalization point. In turbulence theory, such an approach was used in [25], where the results obtained by the author proved to be equivalent to the procedure of continuation in ε.

Simple analysis shows that, to satisfy the above requirements, one should make the change

$$\tilde{F}\left(\frac{q}{m}\right) \longrightarrow \tilde{F}\left(\frac{\mu}{m}\right)\left(\frac{\mu^2}{q^2}\right)^{d/2+\varepsilon}, \tag{4.3}$$

where the parameter $\varepsilon \longrightarrow +0$ is introduced for convenience when separating logarithmically diverging terms.

The requirement of the RG invariance implies that the following relations should hold under the change of the normalization point $\mu \longrightarrow \mu'$:

$$\begin{aligned}
D f_1\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D^2}\right) &= D' f_1\left(\frac{q^2}{\mu'^2}, \frac{m^2}{\mu'^2}, \frac{\sigma^2\lambda'}{D'^2}\right),\\
\lambda f_2\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D^2}\right) &= \lambda' f_2\left(\frac{q^2}{\mu'^2}, \frac{m^2}{\mu'^2}, \frac{\sigma^2\lambda'}{D'^2}\right),\\
H f_3\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D}\right) &= H' f_3\left(\frac{q^2}{\mu'^2}, \frac{m^2}{\mu'^2}, \frac{\sigma^2\lambda'}{D'^2}\right),\\
H_1 f_4\left(\frac{q^2}{\mu^2}, \frac{m^2}{\mu^2}, \frac{\sigma^2\lambda}{D}\right) &= H_1' f_4\left(\frac{q^2}{\mu'^2}, \frac{m^2}{\mu'^2}, \frac{\sigma^2\lambda'}{D'^2}\right).
\end{aligned} \tag{4.4}$$

According to the normalization conditions (4.1),

$$\begin{gathered}
f_1(1, y, g) = f_2(1, y, g) = f_3(1, y, h)\\
= f_4(1, y, h) = 1,\\
g = \frac{\sigma^2\lambda}{D^2}, \quad h = \frac{\sigma^2\lambda}{D}.
\end{gathered} \tag{4.5}$$

It follows from (4.2) and (4.4) that the functions $f_{1,2}(x, y, g)$ satisfy the RG functional equations

$$\begin{aligned} f_{1,2}(x, y, g) &= f_{1,2}(\alpha, y, g) f_{1,2}\left(\frac{x}{\alpha}, \frac{y}{\alpha}, \tilde{g}(\alpha, y, g)\right), \\ \tilde{g}(x, y, g) &= g f_2(x, y, g)/f_1^2(x, y, g), \end{aligned} \tag{4.6}$$

while the functions $f_{3,4}(x, y, h)$ satisfy the equations

$$\begin{aligned} f_{3,4}(x, y, h) &= f_{3,4}(\alpha, y, h) f_{3,4}\left(\frac{x}{\alpha}, \frac{y}{\alpha}, \tilde{h}(\alpha, y, g)\right), \\ \tilde{h}(x, y, g) &= h f_2(x, y, g)/f_1(x, y, g). \end{aligned} \tag{4.7}$$

Differentiating Eqs. (4.6) and (4.7) with respect to the parameter α and then setting $\alpha = 1$, we obtain the following differential equations for the functions $f_{1,2}(x, y, g)$ and $f_{3,4}(x, y, h)$:

$$\begin{gathered} \left\{-x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} + \beta_g(y, g)\frac{\partial}{\partial g} + \gamma_{1,2}(y, g)\right\} \\ \times f_{1,2}(x, y, g) = 0, \\ \beta_g(y, g) = \left.\frac{\partial \tilde{g}(x, y, g)}{\partial x}\right|_{x=1}, \\ \gamma_{1,2}(y, g) = \left.\frac{\partial f_{1,2}(x, y, g)}{\partial x}\right|_{x=1}, \quad \beta_g = g(\gamma_2 - 2\gamma_1) \end{gathered} \tag{4.8}$$

and

$$\begin{gathered} \left\{-x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} + \beta_h(y, h)\frac{\partial}{\partial g} + \gamma_{3,4}(y, h)\right\} \\ \times f_{3,4}(x, y, h) = 0, \\ \beta_h(y, h) = \left.\frac{\partial \tilde{h}(x, y, h)}{\partial x}\right|_{x=1}, \\ \gamma_{3,4}(y, h) = \left.\frac{\partial f_{3,4}(x, y, h)}{\partial x}\right|_{x=1}, \quad \beta_h = h(\gamma_2 - \gamma_1). \end{gathered} \tag{4.9}$$

To solve the RG differential equations, one must know the RG functions γ_i that are determined, according to (4.7) and (4.9), by the behavior of the functions f_i near the normalization point. According to the RG method [6], we determine these functions in the low-order approximation of the renormalized perturbation theory. Appropriate calculations yield the following results for the RG functions γ_i:

$$\begin{gathered} \gamma_1(y, g) = \gamma_2(y, g) = A_d g \tilde{F}(y^{-1/2}) y^{-d/2}, \\ \gamma_3(y, h) = (d-1) A_d h \tilde{F}(y^{-1/2}) y^{-d/2}, \\ \gamma_4(y, h) = 0, \end{gathered} \tag{4.10}$$

where

$$A_d = \frac{1}{2d}\frac{S_d}{(2\pi)^d},$$

and S_d is the surface area of a d-dimensional sphere of unit radius.

The solution of the RG equations for the functions $f_1(x, y, g)$ and $f_2(x, y, g)$ yields

$$\tilde{h}(q) = \frac{\sigma^2 \tilde{\lambda}(q)}{\tilde{D}(q)} = h = \text{const} \quad (\beta_h = 0). \tag{4.11}$$

A solution to Eq. (4.9) for the function $f_3(x, y, h)$ is given by

$$\begin{aligned} f_3(x, y, h) &= \exp\left\{-\int_y^{y/x} \frac{\gamma_3(t, h)\,dt}{t}\right\} \\ &= \exp\left\{-(d-1) A_d h \int_y^{y/x} \tilde{F}(t^{-1/2}) t^{-d/2} dt\right\}, \end{aligned} \tag{4.12}$$

which yields

$$\begin{aligned} &H(q) \\ &= H \exp\left\{\frac{2(d-1)A_d h}{m^d}\int_\mu^q \tilde{F}\left(\frac{p}{m}\right) p^{d-1} dp\right\}. \end{aligned} \tag{4.13}$$

The "asymptotic freedom" requirements

$$H(\infty) = \langle \kappa_1 \rangle, \quad \tilde{D}(\infty) = 1, \quad \tilde{\lambda}(\infty) = 1$$

lead to $h = \sigma^2$, which allows one to determine the following relation between H and μ:

$$\begin{aligned} H &= \langle \kappa^{-1} \rangle \\ &\times \exp\left\{-\frac{d-1}{d}\frac{S_d \sigma^2}{(2\pi m)^d}\int_\mu^\infty \tilde{F}\left(\frac{p}{m}\right) p^{d-1} dp\right\}. \end{aligned} \tag{4.14}$$

Eliminating these parameters, we obtain

$$\begin{aligned} H(q) &= \langle \kappa^{-1} \rangle \\ &\times \exp\left\{-\frac{d-1}{d}\frac{S_d \sigma^2}{(2\pi m)^d}\int_q^\infty \tilde{F}\left(\frac{p}{m}\right) p^{d-1} dp\right\} \\ &= \langle \kappa^{-1} \rangle \exp\left\{-\frac{d-1}{d}\int \frac{d\mathbf{p}}{(2\pi)^d} \tilde{B}(\mathbf{p}) \theta(p - q)\right\} \\ &= \langle \kappa^{-1} \rangle \exp\left\{-\frac{d-1}{d}\sigma^2 + \frac{d-1}{d}\int \frac{d\mathbf{p}}{(2\pi)^d} \tilde{B}(\mathbf{p}) \theta(q - p)\right\}, \end{aligned} \tag{4.15}$$

where $\theta(p)$ is the Heaviside function.

The solution of the equation for $H_1(q)$ yields $H_1(q) = \text{const} = 0$.

Taking into account the lognormal distribution formula

$$\langle \kappa^{-1} \rangle = K_G^{-1} \exp(\sigma^2/2),$$

from (4.15) we obtain

$$(\tilde{K}_{ij}(\mathbf{q}) = K_G \delta_{ij})$$

$$\times \exp\left\{\frac{d-2}{2d}\sigma^2 - \frac{d-1}{d}\int \frac{d\mathbf{p}}{(2\pi)^d}\tilde{B}(\mathbf{p})\theta(q-p)\right\}, \quad (4.16)$$

which in the large-scale approximation $q \longrightarrow 0$ reproduces the Landau–Lifshits–Matheron formula (1.6).

5. DISCUSSION

The application of the RG method in its quantum-field formulation allows one not only to obtain the Landau–Lifshits–Matheron formula, which is valid in the large-scale limit, but also to determine $\tilde{K}_{ij}(\mathbf{q})$ in the entire range of wavenumbers; in this case, the shape of the correlation function and the scale parameter with the dimension of length prove to be taken into account. Note that the Wilson RG approach allows one to determine only an averaged response of the medium to large-scale perturbations.

Our result shows that, within the RG approach, the contributions of the effects of the Green's function renormalization and a vertex are not negligible; however, they compensate for each other, and therefore we can ignore them; i.e., we can seek a solution to Eq. (3.10) from the very beginning. We have to emphasize that the predictions based on the application of the RG approach should not be considered as exact results since formula (4.3) was used when deriving them; this formula is an analogue of the ε-expansion procedure, which has not been rigorously substantiated. This procedure is a payment for the right to use the RG-invariance considerations when summing infinite series of perturbation theory.

To find out to what extent the technical tricks used within the RG approach are justified, we calculated the effective permittivity in the second-order approximation of perturbation theory for the model case

$$B(r) = B_0 \exp(-r^2/4m).$$

The appropriate calculation yields

$$K_{\text{eff}}^{-1} = K_G^{-1}\left\{1 + \frac{d-1}{d}\sigma^2 + \frac{d-1}{2d}\sigma^4[a_1 + a_2 + a_3]\right\},$$

$$a_1 = 1 - \beta\left(\frac{d}{2}\right), \quad a_2 = -1 + (d-1)\beta\left(\frac{d}{2}\right),$$

$$a_3 = \frac{d-1}{d} - (d-2)\beta\left(\frac{d}{2}\right), \quad (5.1)$$

$$\beta\left(\frac{d}{2}\right) = \int_0^1 \frac{t^{d/2-1}}{1+t}dt,$$

where a_1 is the contribution of the second iteration when solving Eq. (3.10) and a_2 and a_3 take into account the contributions due to the renormalization of the Green's function and a vertex, respectively. It follows from (5.1) that the above-listed assumptions prove to be incorrect at least when $d = 2$, although the final result obtained corresponds to the Landau–Lifshits–Matheron formula. However, taking into account that $\beta(d/2) \longrightarrow 1/d$ as d increases, we obtain

$$a_1 \longrightarrow (d-1)/d, \quad a_2 \longrightarrow -1/d, \quad a_3 \longrightarrow 1/d,$$

and the assumptions used prove to be valid in the limit of large d. Numerical estimates show that, even for $d = 3$, the error due to the above approximations does not exceed 15% and, for a hypothetical case of $d = 4$, it does not exceed 8%. These estimates give hope that the application of formula (4.16) to determining the form of the integral kernel $K_{ij}^{-1}(r)$ will not lead to large errors when constructing a statistical solution to a transfer problem in a randomly heterogeneous medium.

APPENDIX

CALCULATION OF STATISTICAL MEANS

To determine quantities of the form

$$\langle \exp\{-u(\mathbf{r})\}u(\mathbf{r}_1)\ldots u(\mathbf{r}_n)\rangle, \quad (A.1)$$

we apply the method of a characteristic functional, which represents a functional Fourier transform of the distribution for $u(\mathbf{r})$. By definition,

$$\Phi[\eta(\mathbf{x})] = \left\langle \exp\left\{i\int d\mathbf{x}u(\mathbf{x})\eta(\mathbf{x})\right\}\right\rangle. \quad (A.2)$$

According to (A.2), we have

$$\langle F\{u(\mathbf{r}_1), u(\mathbf{r}_2), \ldots, u(\mathbf{r}_n)\}\rangle$$

$$= F\left\{\frac{\delta}{i\delta\eta(\mathbf{r}_1)}, \frac{\delta}{i\delta\eta(\mathbf{r}_2)}, \ldots, \frac{\delta}{i\delta\eta(\mathbf{r}_n)}\right\}\Phi[\eta(\mathbf{x})]\Bigg|_{\eta(\mathbf{x})=0} \quad (A.3)$$

for the mean values of the random-field function, and, for the means of the form (A.1), we obtain

$$\langle \exp\{-u(\mathbf{r})\}u(\mathbf{r}_1)\ldots u(\mathbf{r}_n)\rangle$$

$$= \exp\left\{i\frac{\delta}{\delta\eta(\mathbf{r})}\right\}\frac{\delta}{i\delta\eta(\mathbf{r}_1)}\ldots\frac{\delta}{i\delta\eta(\mathbf{r}_n)}\Phi[\eta(\mathbf{x})]\Bigg|_{\eta(\mathbf{x})=0}. \quad (A.4)$$

The operator exponential function in (A.4) represents an operator of functional translation, whose action is defined by the relation

$$\exp\left\{i\frac{\delta}{\delta(\mathbf{r})}\right\}\Phi[\eta(\mathbf{x})] = \Phi[\eta(\mathbf{x}) + i\delta(\mathbf{x}-\mathbf{r})]. \tag{A.5}$$

The application of formulas (A.3)–(A.5) allows us to calculate the means on the basis of the explicit form of the characteristic functional. In the case of a lognormal distribution for $\langle u(\mathbf{r})\rangle = 0$, we can determine

$$\begin{aligned}&\Phi[\eta(\mathbf{x})]\\ &= \exp\left\{-\frac{1}{2}\int d\mathbf{x}\, d\mathbf{x}'\eta(\mathbf{x})B(\mathbf{x}-\mathbf{x}')\eta(\mathbf{x}')\right\}.\end{aligned} \tag{A.6}$$

Below, we present the results of appropriate calculations for several low-order moments:

$$\begin{gathered}\langle \kappa^{-1}(\mathbf{r})u(\mathbf{r}_1)\rangle = -\langle \kappa^{-1}\rangle B(\mathbf{r}-\mathbf{r}_1),\\ \langle \kappa^{-1}(\mathbf{r})u(\mathbf{r}_1)u(\mathbf{r}_2)\rangle = \langle \kappa^{-1}\rangle\\ \times\{B(\mathbf{r}_1-\mathbf{r}_2) + B(\mathbf{r}-\mathbf{r}_1)B(\mathbf{r}-\mathbf{r}_2)\},\\ \langle \kappa^{-1}(\mathbf{r})u(\mathbf{r}_1)u(\mathbf{r}_2)u(\mathbf{r}_3)\rangle = -\langle \kappa^{-1}\rangle\{B(\mathbf{r}-\mathbf{r}_1)B(\mathbf{r}_2-\mathbf{r}_3)\\ + B(\mathbf{r}-\mathbf{r}_2)B(\mathbf{r}_1-\mathbf{r}_3) + B(\mathbf{r}-\mathbf{r}_3)B(\mathbf{r}_1-\mathbf{r}_2)\\ + B(\mathbf{r}-\mathbf{r}_1)B(\mathbf{r}-\mathbf{r}_2)B(\mathbf{r}-\mathbf{r}_3)\}.\end{gathered} \tag{A.7}$$

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project nos. 99-01-00957 and 99-01-01083.

REFERENCES

1. L. W. Gelhar and C. L. Axness, Water Resour. Res. **19**, 161 (1983).
2. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Gostekhizdat, Moscow, 1957).
3. É. V. Teodorovich, Prikl. Mat. Mekh. **66**, 448 (2002).
4. P. C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).
5. D. T. Hristopulos and G. Christakos, J. Stoch. Hydr. Hydraul. **11**, 369 (1997).
6. N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Nauka, Moscow, 1984; Wiley, New York, 1980).
7. K. Wilson, Adv. Math. **16**, 170 (1975).
8. K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
9. L. P. Kadanoff, Physics (N. Y.) **2**, 263 (1966).
10. D. S. Dean, I. T. Drummond, and R. R. Horgan, J. Phys. A **27**, 5135 (1994).
11. Q. Zhang, Water Resour. Res. **31**, 577 (1995).
12. U. Jaekel and H. Vereecken, Water Resour. Res. **33**, 2287 (1997).
13. D. T. Hristopulos and G. Christakos, Stochastic Environmental Research and Risk Assessment: Research Joirnal **13**, 131 (1999).
14. B. Noetinger, C. R. Paris **331** (5), 353 (2000).
15. V. Yakhot and S. A. Orszag, J. Sci. Comput. **1**, 3 (1986).
16. W. P. Dannevik, V. Yakhot, and S. A. Orszag, Phys. Fluids **30**, 2021 (1987).
17. K. Wilson and M. Fisher, Phys. Rev. Lett. **28**, 240 (1972).
18. K. Wilson and G. Kogut, Phys. Rep. C **12**, 75 (1974).
19. É. V. Teodorovich, Izv. Akad. Nauk SSSR, Ser. Mekh. Zhidk. Gaza, No. 4, 36 (1990).
20. É. V. Teodorovich, Izv. Ross. Akad. Nauk, Ser. Mekh. Zhidk. Gaza, No. 1, 40 (1994).
21. H. W. Wyld, Ann. Phys. **14**, 143 (1961).
22. L. L. Lee, Ann. Phys. **32**, 292 (1965).
23. É. V. Teodorovich, Teor. Mat. Fiz. **101**, 28 (1994).
24. C. De Dominicis and P. C. Martin, Phys. Rev. A **19**, 419 (1979).
25. É. V. Teodorovich, Dokl. Akad. Nauk SSSR **299**, 836 (1988) [Sov. Phys. Dokl. **33**, 247 (1988)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 77–82.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 90–96.
Original Russian Text Copyright © 2002 by Voronov, Stal'gorova, Gromnitskaya.

SOLIDS
Structure

Anomalies in the Variation of Elastic Properties of Cesium during Phase Transformations under a Pressure up to 5 GPa

F. F. Voronov†, O. V. Stal'gorova, and E. L. Gromnitskaya*
Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: grom@ns.hppi.troitsk.ru*

Received December 26, 2001

Abstract—The pulsed ultrasonic method is used for studying polycrystalline cesium under a pressure up to 5 GPa. Elastic parameters and compression ratio of cesium are determined and anomalies in the elastic properties of cesium during phase transitions CsI–CsII–CsIII–CsIV are revealed. It is found that the CsII–CsIII electron-structure transformation is preceded by anomalous compressibility of the fcc phase of cesium and by softening of longitudinal and transverse acoustic modes of the cesium phonon spectrum. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The heavy alkali metal cesium is distinguished from other metals in that it experiences a series of phase transformations associated with changes in its crystalline and electronic structure in the range of pressures p and temperatures T accessible for investigations. The behavior of macroscopic parameters of cesium exhibits a number of anomalies. Under a pressure $p \sim 2.3$ GPa, a structural transformation from the bcc (CsI) to the fcc (CsII) phase takes place; during this transition, the volume decreases jumpwise ($\Delta V/V_0 = 0.7\%$). However, of special interest are two close phase transformations in cesium compressed by 50%: CsII–CsIII ($p = 4.21$ GPa) and CsIII–CsIV ($p = 4.33$ GPa) [1–3]. Until recently, the CsII–CsIII transformation during which the volume collapses approximately by 9.1% was interpreted by many authors as a 6*s*–5*d* electron transition inducing an isomorphic fcc–fcc phase transformation [4, 5]. However, the possibility of an isostructural phase transformation in cesium under the above conditions was questioned in [6, 7]. In addition, it was concluded on the basis of estimates of the decrease in the atomic radius in cesium (2% and 10% during transformations CsII–CsIII and CsIII–CsIV, respectively) [8] that the 6*s*–5*d* avalanche transition mainly occurs during the CsIII–CsIV transformation. For $p \sim 4.33$ GPa, the CsIII structure is transformed into the tetragonal CsIV structure [8] with a decrease in volume by 4.3%.

Anomalies in the variation of physical properties of cesium are manifested even in the bcc phase. For example, according to the results of optical studies, cesium with the bcc structure exhibited strong absorption in the pressure range of 1.8 GPa, which is attributed to a change in the *d* nature of conduction electrons [9]. In our studies using ultrasonic technique, anomalies in the behavior of elastic parameters of bcc cesium were detected for pressures $p > 1.4$ GPa. Starting from this pressure, the dependence of compressibility of bcc cesium on the reduced pressure p/K_t deviates towards higher pressures as compared to other alkali metals [10].

The fcc phase of cesium is of special interest to researchers. It was found using neutron diffraction analysis that CsII possesses an anomalously soft compression isotherm [11], while our early studies using the ultrasonic method [12] revealed a very small variation of the velocity $v_l(p)$ of longitudinal ultrasonic waves in fcc cesium (i.e., upon an increase in pressure from 2.28 to 4.21 GPa). Moreover, after passing the peak (at $p \sim 3.8$ GPa), the value of $v_l(p)$ decreased as the CsII–CsIII transition was approached. The melting curve is known to have a negative slope after passing through a peak at 3.5 GPa. These anomalies can easily be explained as a consequence of a continuous variation of the electron structure of cesium under pressure [8, 13, 14].

It was undoubtedly interesting to continue our ultrasonic experiments with cesium [10, 12] under higher pressures, to determine anomalies in the variation of the low-frequency part of the phonon spectrum of cesium, to obtain the dependence of elastic parameters of cesium on pressure up to 5 GPa, and to find the variation of these parameters during the phase transformations CsII–CsIII and CsIII–CsIV.

2. EXPERIMENTAL TECHNIQUE

We studied the pressure dependences of the velocities of longitudinal, $v_l(p)$, and transverse, $v_t(p)$, ultrasonic waves propagating in cesium under a pressure up to 5.0 GPa by using an "ultrasonic piezometer" (cylinder–piston-type chamber) [10] and a high-pressure

---

†Deceased.

1063-7761/02/9501-0077$22.00 © 2002 MAIK "Nauka/Interperiodica"

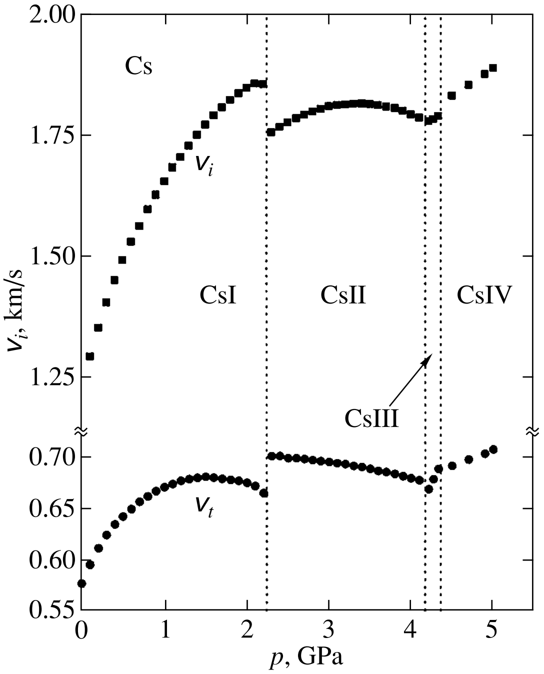


Fig. 1. Pressure dependences of the velocities of longitudinal, v_l, and transverse, v_t, ultrasonic waves in cesium.

setup of the Toroid type [15]. In our experiments, we used cesium with 99.9% content of the matrix material. The samples were prepared under a layer of thoroughly dehydrated oil as described in [10]. During the measurements of $v_l(p)$ and $v_t(p)$ in the cylinder–piston chamber in the initial interval of pressures 0–1.1 GPa, we succeeded in confining the substance in the chamber under pressure without encapsulating the samples by using only thin protecting Teflon films (with a thickness of the order of 0.02 mm), which was attained by a thorough adjustment of the sizes of chamber parts. The measurements during the elevation and removal of pressure in the interval from 0 to 2.5 GPa in the piezometer were made on encapsulated samples. In this case, the samples cut from an ingot were placed in aluminum capsules in the form of cylinders with a wall thickness of approximately 1 mm and with an end face thickness of approximately 0.6 mm.

In the main series of experiments (i.e., under pressures from 0 to 5.0 GPa), we used a version of the Toroid high-pressure solid-state chamber in the form of a "lentil with a flat bottom" [16]. Cesium contained in a cylindrical capsule of diameter 14 mm with a 7-mm height was placed in a catlinite container. The measurements of $v_l(p)$ and $v_t(p)$ were made only for increasing pressure since a decrease in pressure led to leakage of the capsule due to irreversible deformation of the container upon a decrease in the cesium volume to half the initial value after the attainment of pressure $p \sim 5$ GPa. The dependence of the pressure in the chamber in the load, $p_c = f(p_n)$, was determined from the phase-transition points in cesium, which were detected according to jumps on the time dependence of an ultrasonic signal passing through the sound-duct pistons and through the sample. The first point corresponding to the emergence of an acoustic contact, $p = (0.005 \pm 0.002)$ GPa, was determined from the graph quite reliably. The second point corresponded to the CsI–CsII transition at $p = (2.28 \pm 0.03)$ GPa. The value of pressure corresponding to this transition was thoroughly measured in experiments made on an ultrasonic piezometer and described in detail in our earlier publication [10]. For the third ($p = 4.21$ GPa) and fourth ($p = 4.33$ GPa) points on the gauge curve, which correspond to the CsII–CsIII and CsIII–CsIV transformations, we used the values of pressure obtained from an analysis of the results presented in [2, 3, 17] on the basis of the available pressure scale [18]. The graph of the gauge curve obtained by us is given in [12]. For pressures higher than 4.21 GPa, we assumed that the gauge dependence $p_c = f(p_n)$ is linear on account of a decrease in the efficiency of the chamber operation after the CsII–CsIII and CsIII–CsIV phase transformations; in this case, we assumed that the decrease in dp_c/dp_n is proportional to the decrease in the cesium volume as a result of these transformations (9.1% and 4.3%) [3].

3. RESULTS AND DISCUSSION

The experimental dependences $L(p)$ of the sample height on pressure obtained by us in the Toroid chamber, as well as the pressure dependences $T_t(p)$ and $T_l(p)$ of the time of travel of transverse and longitudinal ultrasonic waves, which were determined by the modified method of visual superposition of pulses [19], were reduced to the actual pressure according to the gauge dependence $p_c = f(p_n)$ and are matched to the results of our previous studies [10] at the point $p = 2.28$ GPa of the CsI–CsII phase transformation.

The velocities $v_i(p)$ of propagation of ultrasonic waves in cesium were determined from the relations

$$\begin{aligned} & v_i(p) \\ & = [L(p) + \Delta L(p)]/\{[L(2.28\text{ GPa}) + \Delta L(2.28\text{ GPa})] \\ & \times v_i^{-1}(2.28\text{ GPa}) - T_i(2.28\text{ GPa}) - T_i(p) - \Delta T_i(p)\}, \end{aligned}$$

where $i = l$ or t, and $\Delta L(p)$ and $\Delta T_i(p)$ are the corrections to the deformations of the plungers and to the change in the transit time of an ultrasonic signal through the sound-duct plungers upon loading, which were determined in separate experiments.

The velocities $v_l(p)$ and $v_t(p)$ of propagation of longitudinal and transverse ultrasonic waves in cesium, which were obtained by processing the results of 24 experiments in the pressure range 2.28–5.0 GPa, are presented in Fig. 1. The same figure shows our results obtained under a pressure up to 2.28 GPa [10]. The error in determining the velocities of propagation of ultrasound did not exceed 1%. It should be noted that the velocity variations at each of the four phases of cesium are different. The pressure dependence of velocities exhibits a behavior typical for simple metals only under a low pressure ($p < 0.4$ GPa) and only for the CsIV phase. For $p \sim 0.5$–1.0 GPa, the dependences $v_l(p)$ and $v_t(p)$ become nonlinear, indicating pretransition phenomena in bcc cesium. In addition, in the vicinity of the bcc–fcc transition, a strong decrease in the amplitude of ultrasonic signals is observed, which is apparently caused by changes in the phonon–phonon interaction due to anharmonic effects. The bcc–fcc phase transformation in cesium occurs almost instantaneously with a small change in volume (about 0.7%), which is typical of martensite transformations. When the fcc phase of cesium is formed, the amplitude of the ultrasonic signal increases considerably. The bcc–fcc transition is accompanied by a 5.6% increase in the velocity $v_t(p)$ of transverse waves and a 5.3% decrease in the velocity $v_l(p)$ of longitudinal waves.

The most clearly manifested anomalies in the behavior of velocities are observed for the fcc phase of cesium. In the entire range of existence of this phase, the velocity of shear ultrasonic waves decreases, while the velocity of longitudinal waves decreases only after passing through a flat peak at approximately 3.5 GPa. The variation of $v_l(p)$ in this pressure range is very small and does not exceed 3%. The CsIV phase is characterized by an increase in both velocities, but it is difficult to obtain an exact quantitative estimate since the gauge dependence $p_c = f(p_n)$ has an indeterminacy for $p > 4.33$ GPa.

Calculating elastic parameters of cesium from ultrasonic data, we obtained the dependence of the volume of the substance on pressure for all the four phases of cesium up to 5 GPa. The compression curve for cesium obtained by us (Fig. 2) is in excellent agreement with the results obtained by the X-ray, piezometric, neutron diffraction, and other methods [1, 3, 8, 13, 20–25].

Figures 3 and 4 show the bulk elastic modulus, shear modulus, and their derivatives as functions of the variation of the atomic volume of cesium in the pressure range from 0 to 5.0 GPa, which were calculated from the velocities of propagation of ultrasound. Since the bulk modulus is a measure of rigidity of a crystal and the shear modulus is a measure of shape variation, the form of variation in these parameters can be used to draw conclusions on changes occurring upon the deformation of the Cs lattice under pressure.

Prior to the bcc–fcc transition, the derivative $\partial K_t/\partial V$ of the bulk modulus with respect to volume exhibits a sharp kink and starts decreasing; i.e., as noted above, a dependence of the properties of cesium on the population of the *s* band starts apparently being manifested. The shear characteristics also indicate a softening of the lattice in the vicinity of the bcc–fcc structural phase transformation; in this case, the value of the derivative $\partial G/\partial p$ decreases from 0.76 under atmospheric pressure to zero prior to the transition. The negative value of the Grüneisen parameter calculated from the shear modu-

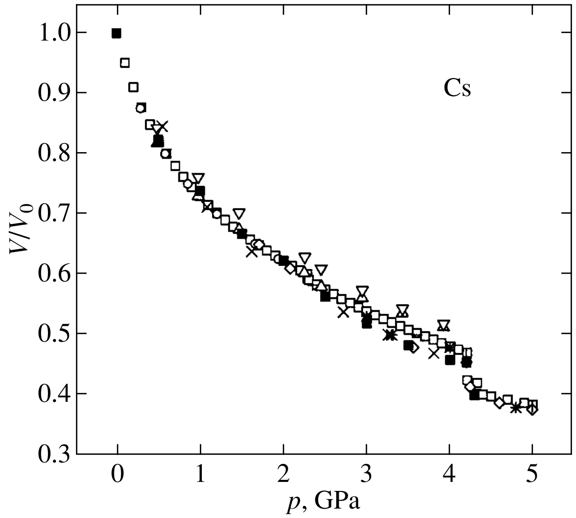


Fig. 2. Dependence of compression of cesium on pressure up to 5.0 GPa: our results (□), results obtained in [20] (○), 21 (△), [1] (▽), [7] (◇), [22] (+), [13] (×), [8] (*), [23] (–), [24] (|), and [25] (■).

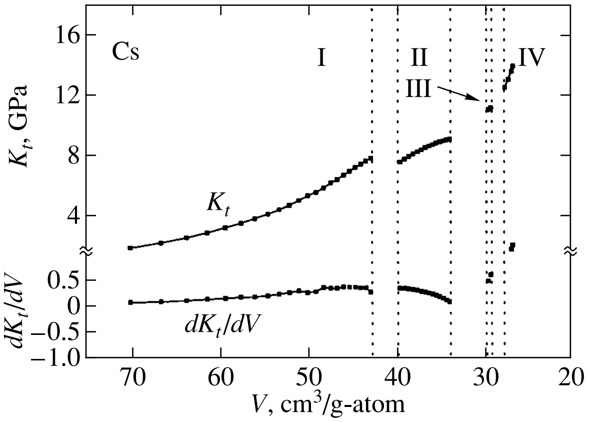


Fig. 3. Dependence of the bulk modulus K_t and its derivative dK_t/dV on the molar volume.

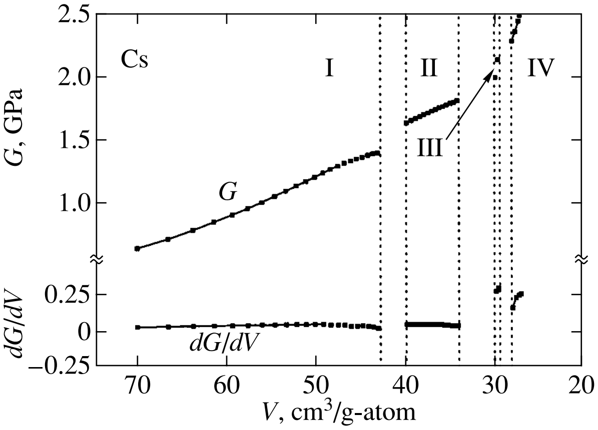


Fig. 4. Dependence of the shear modulus G and its derivative dG/dV on the molar volume.

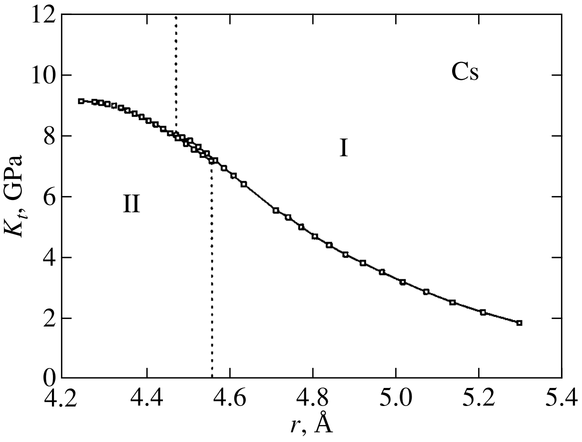


Fig. 5. Dependence of the bulk modulus K_t on the distance r between the nearest neighbors in the lattice.

lus [10] in the vicinity of the phase transition also indicates the emergence of a soft shear (in all probability, acoustic TA_1) mode in the phonon spectrum of the bcc phase of cesium.

A transition to a more densely packed fcc structure upon an increase in the coordination number is accompanied by an increase in the shear modulus (about 13%). At the same time, the bulk modulus decreases (by about 14%) during this transition, which is apparently associated with an increase in the distance between the nearest neighbors. This feature is visually demonstrated by the dependence of the isothermal bulk modulus K_t on the distance between the nearest neighbors in the lattice (Fig. 5), which shows no jump in the bulk modulus during this transition.

It should be noted that no experiments were made to our knowledge on the direct measurement of the bulk modulus of cesium and its derivative with respect to pressure. Ultrasonic studies make it possible to measure K_t quite accurately as a function of pressure in the continuous mode. The experiment shows (see Fig. 3) that the fcc phase of cesium is compressible to a greater extent than the bcc phase, and the derivative dK_t/dV of the bulk modulus decreases upon an increase in pressure in the entire range of the existence of the fcc phase. Since the value of dK_t/dV determines the degree of softness in the equation of state, we can state that, as we approach the CsII–CsIII phase transformation, the resistance of cesium to volume deformation decreases, and an intense transition of electrons from the repulsive *s* bands to the attracting *d* band takes place in the range of the fcc phase, its intensity increasing as the transformation is approached. The strongest anomalies in $K_t(V)$ are observed for compressions $V/V_0 \sim 0.5$, which is in accord with the results of calculations made in [23, 26], where it was shown that the bottom of the 5*d* band passes through the Fermi level exactly for compressions $V/V_0 = 0.47–0.48$.

Figure 4 shows the experimental dependences of the shear modulus and its derivative on the atomic volume of cesium. It should be noted that the function $G(p)$ for the fcc phase exhibits a weak (virtually, linear) increase upon a decrease in the volume derivative of the shear modulus from 0.03 to 0.02 by the instant of the CsII–CsIII transformation. The calculations of the stability of the fcc lattice of cesium show that the softening of the acoustic shear mode TA_1 leading to a decrease in the shear modulus $c' = (1/2)(c_{11} - c_{12})$ plays the leading role in this structural transformation [7, 14, 27]. In addition, a softening of the TA_2 mode and, hence, a decrease in the modulus c_{44} cannot be ruled out either [14]. The shear modulus of a polycrystal ($G = \rho v_t^2$) can be presented in the Voigt approximation [28] for crystal with a cubic system in terms of the elastic constants of a single crystal, $G = (2/5)[(1/2)c_{11} - c_{12}] + (3/5)c_{44}$. An analysis of our experimental dependences of the velocities of elastic shear waves and of the shear modulus taking into account a slight increase in the shear modulus with pressure indicates a softening of the acoustic mode TA_1 in the phonon spectrum of fcc cesium upon an increase in pressure in the region as the CsII–CsIII transition is approached; the second transverse mode TA_2 associated with the elastic constant c_{44} is not softened in this case. The form of variation of the velocity of longitudinal waves in cesium and of the bulk modulus also indicates a softening of the longitudinal acoustic mode prior to this phase transition.

The Grüneisen parameter $\gamma_l(p)$ calculated from our experimental dependences becomes negative, while $\gamma_t(p)$ assumes a close-to-zero value for $V(p)/V(0) \sim 0.48$; i.e., for pressures $p > 4.0$ GPa, the longitudinal *LA* and transverse TA_1 modes become soft, and the crystal lattice loses its stability. A decrease in phonon frequencies in the vicinity of the volume collapse leads to an increase in the amplitude of thermal vibrations, which correlates with a decrease in the melting temperature and a sharp increase in the resistivity, which were observed under such pressures in [1, 3].

An analysis of the phase of CsIII is a complicated problem since the region of existence of this phase is very narrow; it amounts to 0.14 GPa at $T = 461$ K, i.e., in the vicinity of the triple points CsII–*L*–CsIII and CsIII–*L*–CsIV, and 0.12 GPa at $T = 298$ K [6]. McWhan and Stevens [13] assumed that the boundaries of the CsIII phase contract rapidly from the triple points to the point $T = 288$ K, $p = 4.3$ GPa, and the CsIII phase does not exist below this point. Indeed, it is shown in [2, 13] that the determination of the region of existence of the CsIII phase at $T = 273$ K is complicated, but this may be due also to a slow kinetics of the transformation at a low temperature [2]. In our experiments at room temperature, the region of existence of CsIII was clearly defined in a number of experiments involving the measurements of the velocities of transverse waves, in which the parameters being measured could be

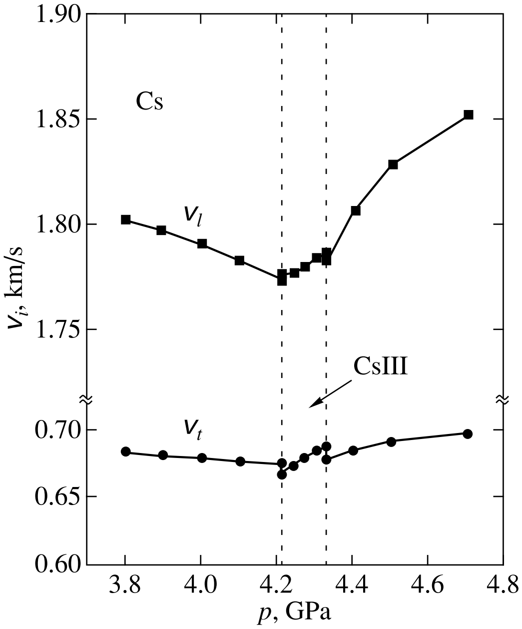


Fig. 6. Experimental pressure dependences of the velocities of longitudinal, v_l, and transverse, v_t, ultrasonic waves in the CsIII phase.

detected at 5–8 points for this phase. Individual measurements at temperatures of 291 and 300 K proved that the region of existence of the CsIII phase, determined as a function of the load exerted by a press, changes insignificantly in this temperature range and may disappear at 288 K. Thus, the region of existence of the CsIII phase in our experiments amounts to 0.12 GPa, which confirms the results obtained in [2]. The CsII–CsIII transformation was detected from slight jumps in the sample height $L(p)$ and the time of travel of transverse, $T_t(p)$, and longitudinal, $T_l(p)$, ultrasonic waves. In addition, the range corresponding to the CsIII phase differs from the phases of adjacent phases CsII and CsIV in the variation of the derivatives dT_i/dp. The CsIII–CsIV transformation is manifested more clearly in the jump and in a change in the slope of the pressure dependence $T_t(p)$ of the travel time of a transverse ultrasonic wave. The weak manifestation of this transformation in the measurements of the velocity of longitudinal waves led us to the erroneous conclusion that only one transformation CsII–CsIV takes place at room temperature [12].

The experimental dependences $v_l(p)$ and $v_t(p)$ obtained by us for the region of phase transformations CsII–CsIII–CsIV are presented separately on a magnified scale in Fig. 6. The relative increase in the bulk modulus and the shear modulus upon the phase transformation CsII–CsIII was 9.4 and 24.5%, respectively, indicating a higher rigidity of the CsIII structure as compared to the CsII.

Figure 7 shows the pressure dependences of the Debye temperature $\Theta(p)$ and the Poisson ratio $\sigma(p)$ of cesium in the pressure range under investigation. It should be noted that the value of $\Theta(3.8\ \text{GPa})/\Theta(0) = 1.5$ is in good agreement with neutron diffraction measurements [13]: $\Theta(3.8\ \text{GPa})/\Theta(0) < 1.3 \pm 0.2$.

The experimental dependences of compression and Debye temperature for cesium were used for estimating the melting point (T_m) from the Lindemann criterion. A satisfactory agreement between the results of experiments and computations is observed for bcc phase of cesium, especially for low pressures for which the condition $T_{\text{exp}}/T_m \approx 1$ is satisfied, where T_{exp} is the experimental temperature. The conditions of applicability of

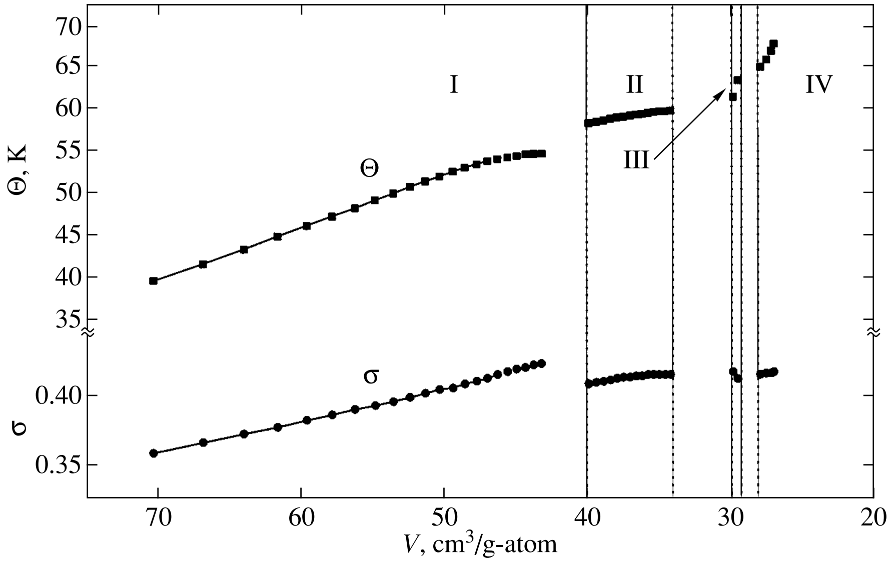


Fig. 7. Experimental dependences of the Debye temperature Θ and the Poisson ratio σ on the molar volume of cesium.

the Lindemann relation for the fcc phase are violated except for pressures close to $p \approx 3.8$ GPa.

As regards the CsIV phase, we can give only a qualitative characteristic of the behavior of its elastic parameters in view of an indeterminacy in the gauge dependence $p_c = f(p_n)$ for pressures $p > 4.33$ GPa. The increase in both velocities and, accordingly, in the elastic moduli $K_t(V)$ and $G(V)$ indicates the rigidity of the formed phase. It was noted in [11] that the elastic modulus $K_t(p)$ for CsIV approaches the values typical of hard metals; i.e., the estimate $dK_t/dp = 4.2$ obtained by us for $p = 4.33$ GPa is quite reliable.

4. CONCLUSIONS

Thus, the experimental pulsed ultrasonic studies of polycrystalline cesium under a pressure up to 5.0 GPa and during the phase transitions CsI–CsII, CsII–CsIII, and CsIII–CsIV enabled us to determine the behavior of elastic parameters of cesium under these conditions and revealed anomalies associated with phase transformations. Theoretical computations based on different models indicate a strong pressure dependence of the electron structure of cesium and the existence of an avalanche *s*–*d* electron transition for compressions $V/V_0 \sim 0.48$. Assumptions were made [23, 27] concerning a decrease in the vibrational frequencies of the lattice upon a compression of fcc cesium in the vicinity of this transition. In our experiments, we observed anomalies in the elastic properties of fcc cesium in the entire range of the existence of this phase, which reflects an intense variation of the population of the *s* band upon a change in pressure. This process began even in bcc cesium for $V(p)/V_0 \sim 0.68$. It was found the electron-structure transformation CsII–CsIII takes place after softening of both longitudinal and transverse modes in the acoustic part of the phonon spectrum. The anomalies in the variation of the bulk modulus of cesium with pressure indicate that the principal avalanche *s*–*d* transition apparently takes place at the boundary of the CsII–CsIII rather than CsIII–CsIV phase transformation, as assumed in [8]. It would be interesting to study in the future cesium under pressure at various temperatures to refine its structure and the regions of existence of the CsIII phase, the boundaries of the CsII–CsIII–CsIV phase transitions, and also the main tendencies in the variation of elastic parameters of cesium upon cooling and heating.

ACKNOWLEDGMENTS

This work was financed by the Russian Foundation for Basic Research (project nos. 01-02-16557 and 02-02-16298), the program “Leading Scientific Schools” (project no. 00-15-96593), a grant from the president of the Russian Federation “Young Doctors of Sciences” (grant no. 00-15-99308), and INTAS (grant no. 00-807).

REFERENCES

1. P. W. Bridgman, Proc. Amer. Acad. Arts Sci. **76**, 55 (1948).
2. A. Jayaraman, R. C. Newton, and J. M. McDonough, Phys. Rev. **159**, 527 (1967).
3. H. T. Hall, L. Merill, and J. D. Barnett, Science **146**, 1297 (1964).
4. R. L. Sterngeimer, Phys. Rev. **78**, 235 (1950).
5. E. S. Alekseev and R. G. Arkhipov, Fiz. Tverd. Tela (Leningrad) **4**, 1077 (1962) [Sov. Phys. Solid State **4**, 795 (1962)].
6. U. Schwarz, K. Takemura, M. Hanfland, and K. Syassen, Phys. Rev. Lett. **81**, 2711 (1998).
7. N. E. Christensen, D. J. Boers, J. L. van Velsen, and D. L. Novikov, Phys. Rev. B **61**, 3764 (2000).
8. K. Takemura, S. Minomura, and O. Shimomura, Phys. Rev. Lett. **49**, 1772 (1982).
9. H. Tups, K. Takemura, and K. Syassen, Phys. Rev. Lett. **49**, 1776 (1982).
10. F. F. Voronov, O. V. Stal’gorova, and E. L. Gromnitskaya, Zh. Éksp. Teor. Fiz. **105**, 1456 (1994).
11. D. B. McWhan, G. Parisot, and D. Bloch, J. Phys. F **4**, L69 (1994).
12. F. F. Voronov, O. V. Stalgorova, and E. L. Gromnitskaya, High Press. Res. **6**, 213 (1991).
13. D. B. McWhan and A. L. Stevens, Solid State Commun. **7**, 301 (1969).
14. V. G. Vaks, M. T. Katsnelson, A. I. Likhtenstein, *et al.*, J. Phys.: Condens. Matter **3**, 1409 (1991).
15. L. G. Khvostantsev, L. F. Vereshchagin, and A. P. Novikov, High Temp.–High Press. **9**, 637 (1977).
16. F. F. Voronov and S. B. Grigor’ev, Izmer. Tekh. **7**, 47 (1977).
17. G. C. Kennedy and P. N. LaMori, J. Geophys. Res. **67**, 851 (1962).
18. D. L. Decker, W. A. Basset, L. Merill, *et al.*, J. Phys. Chem. Ref. Data **1**, 773 (1972).
19. E. P. Papadakis, Rev. Sci. Instrum. **47**, 806 (1976).
20. M. S. Anderson and C. A. Swenson, Phys. Rev. B **31**, 668 (1985).
21. P. W. Bridgman, Proc. Amer. Acad. Arts Sci. **72**, 207 (1938).
22. C. E. Weir, G. J. Piermarini, and S. Bloch, J. Chem. Phys. **54**, 2768 (1971).
23. D. Glötzel and A. K. McMahan, Phys. Rev. B **20**, 3210 (1979).
24. K. Takemura and K. Syassen, Phys. Rev. B **32**, 2213 (1985).
25. R. Böehler, Mater. Res. Soc. Symp. Proc. **22**, 21 (1984).
26. I. Yamashita and S. Asano, J. Phys. Soc. Jpn. **29**, 264 (1970).
27. J. Xie, S. P. Chen, J. Tse, *et al.*, Phys. Rev. B **62**, 3624 (2000).
28. W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (Van Nostrand, New York, 1950; Inostrannaya Literatura, Moscow, 1952).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 83–86.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 97–100.
Original Russian Text Copyright © 2002 by Pokrovskiĭ, Smirnova, Khval'kovskiĭ.

SOLIDS
Structure

Infrared Absorption by Pairs of Group III and V Impurities in Silicon

Ya. E. Pokrovskiĭ*, O. I. Smirnova, and N. A. Khval'kovskiĭ

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 101999 Russia

**e-mail: yaep@mail.cplire.ru*

Received January 11, 2002

Abstract—Infrared absorption bands, shifted to regions of lower energies relative to narrow lines of transitions of impurities to excited states, are investigated in silicon doped with group III and V impurities in concentrations above 10^{16} cm^{-3}. It is found that the band structure is peculiar to each of the investigated impurities but independent of their concentrations, and the absorption coefficients in the bands increase approximately quadratically with concentration. This leads one to infer that the bands are caused by the excitation of charge carriers bound on impurity pairs localized within several Bohr radii. *© 2002 MAIK "Nauka/Interperiodica".*

Series of narrow lines in absorption spectra of group III and V impurities in silicon have long been studied and identified as transitions from the ground state of single impurities to excited states [1]. However, as the impurity concentration N increases, it is natural to expect a manifestation of the contribution to absorption due to electronic transitions in complexes where the impurity atoms are spaced at relatively close distances from one another. Indeed, a broad absorption band with a maximum at the energy $h\nu = 23$ meV was observed in silicon doped with boron [2]. This band and the long-wave absorption band in silicon doped with arsenic were investigated in [3] for various concentrations of impurities. A similar absorption band was observed in silicon doped with gallium [4]. An absorption spectrum was recorded in silicon doped with indium [5], which reminded one of a regular absorption spectrum for indium, but was shifted towards the region of lower energies. This absorption was associated in [5] with the shallow state of indium impurity, revealed previously in [6] by the temperature dependences of the hole concentration and by the photoconductivity spectrum. Also observed in [6] was a superlinear increase in the concentration N_2 of shallow states with increasing N. In the opinion of Baron *et al.* [6], this did not rule out the possibility of the shallow states being impurity pairs. We have confirmed this assumption (previously made in [7]) for the majority of group III and V impurities in silicon.

Figure 1 gives absorption spectra for doped silicon in the region of energies less than the energies of transition of group III and V impurities to excited states at 4.2 K. We have previously believed that these broad absorption bands are caused by transitions to allowed bands from 1S impurity states populated by background radiation at room temperature [3]. Indeed, the band edges in Si[B] and Si[As] correspond well to such energies [1]. However, the provision of cooled filters arranged before the samples, which reduce the effect of background radiation by more than three orders of magnitude, had almost no effect on the absorption in long-wave bands. From this, it follows that the bands cannot be associated with transitions of charge carriers from the excited states of impurities.

The structure of long-wave bands is peculiar to each impurity. In the Si[B] spectrum, two flat maxima are observed, the distance between which is close to the difference between the energies of the first and second excited states of the acceptor series $P_{3/2}$ [1], and in Si[Al], flat maxima are observed, as in the case of boron impurity; however, a peak is observed against their background, which is caused by the aluminum–oxygen complex [5], which, in our crystals grown by the method of crucibleless zone crystallization, was present in low concentration. In the low-energy region, weak peaks of absorption by residual boron impurity are observed: in the case of Si[Ga], it is a flat step with a weakly defined fine structure. In the case of a deep indium impurity, the long-wave absorption bands are very weak and may be recorded only at concentrations $N > 10^{17}$ cm^{-3}. No samples doped as strongly as that were available to us; therefore, Fig. 1d shows the spectrum recorded in [5] for Si[In] at 8 K. The spectrum is very similar to regular absorption spectra at low concentrations of indium but is shifted to the region of lower energies by 43 meV, and the absorption lines are broadened significantly. A wide step is observed in the Si[As] spectrum, while the Si[Sb] and Si[P] spectra are poorly defined with a gently sloping long-wave tail and a maximum or a step between the $2P_0$ and $2P_\pm$ absorption lines.

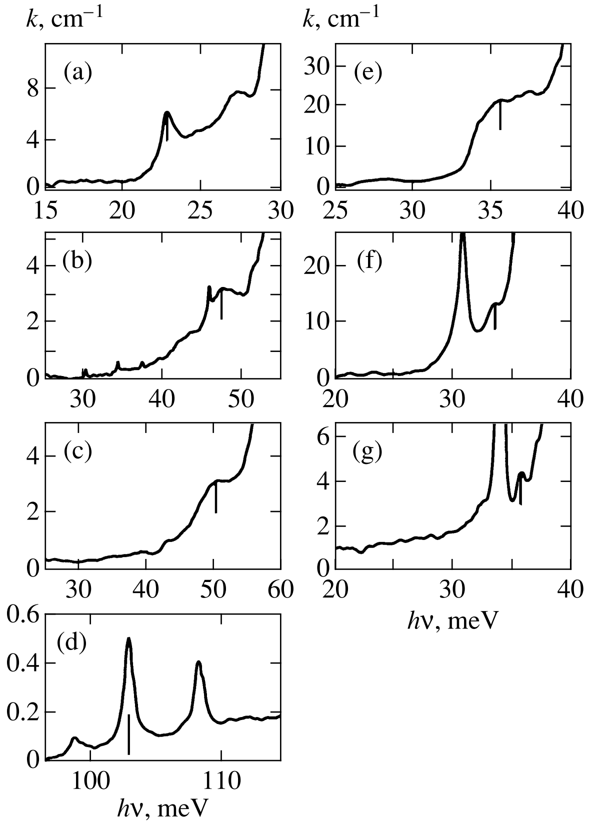


Fig. 1. Long-wave absorption spectra for silicon doped with impurities in concentrations N (cm^{-3}) of (a) B (7.6×10^{16}), (b) Al (1.5×10^{17}), (c) Ga (1.6×10^{17}), (d) In ($4\ 10^{17}$) [5], (e) As (1.8×10^{17}), (f) Sb (2.5×10^{16}), (g) P (1.5×10^{16}). The vertical lines indicate the photon energies $h\nu$ for which the absorption coefficients k in Fig. 2 were determined.

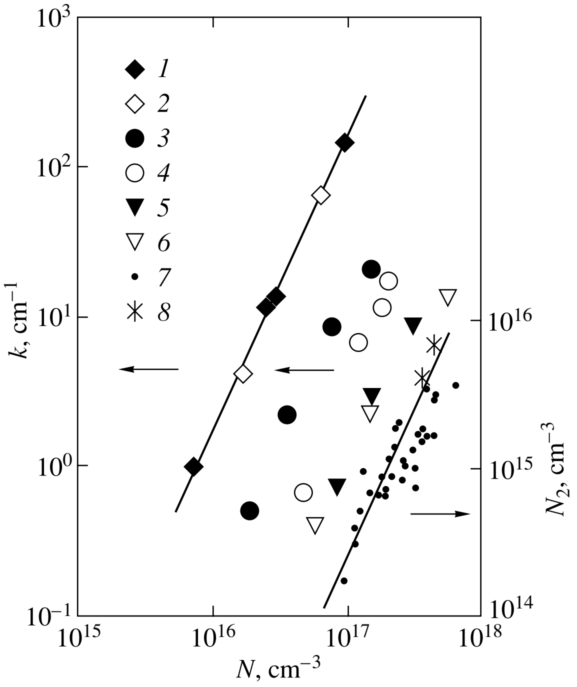


Fig. 2. The absorption coefficients k in silicon as functions of the impurity concentration N (*1*, Sb; *2*, P; *3*, B; *4*, As; *5*, Al; *6*, Ga), as well as (*7*) the N dependences of the concentration N_2 of pairs for Si[In] borrowed from [6] and (*8*) those determined by spectra from [5, 6]. The solid lines correspond to quadratic dependences.

In spite of this difference between the long-wave absorption spectra in doped silicon, they exhibit a number of features in common with one another. First, the structure of the spectra does not depend on the impurity concentration N. Second, the higher the ionization energy and the stronger the localization of the charge carriers in the ground state of impurities, the higher the values of N at which the long-wave absorption may be recorded. Third, the dependence of the absorption coefficient k in the bands on N has the same form for all impurities.

Figure 2 gives the dependences of k on concentration N, which were determined at energies fixed for each one of the impurities (Sb, P, B, As, Al, Ga). One can see that the absorption depends approximately quadratically on the impurity concentration. Note that the errors in determining this dependence were maximal in the regions with $k < 1$ cm^{-1} and at concentrations $N > 10^{17}$ cm^{-3}. The concentration N was determined by the specific resistance of the samples at 300 K and by its dependences on N according to the graphs given in [8] for each of the impurities investigated by us. These dependences become weaker in the range of high values of N, which results in a lower accuracy in determining the concentration. Nevertheless, even in these regions the error did not exceed 20%. Also given in Fig. 2 is the dependence of the concentration N_2 of shallow states of indium impurity on its total concentration N from [6]. One can see that, in this case as well, in view of the low accuracy of the method employed in [6], the “cloud” of points in the figure follows adequately the quadratic dependence on N.

The quadratic dependence of the absorption coefficient k on N indicates that long-wave bands arise during the excitation of charge carriers bound on pairs of impurities. We will treat the situation proceeding from the simplest analog, i.e., a system of two protons and two electrons. This well-known pattern of electron terms of a hydrogen molecule is given in Fig. 3 (see, for example, [9]). Unlike the molecules, the impurities at crystal lattice sites are fixed at random at distances r. The energy differences between the dashed and solid curves in Fig. 3 correspond to a decrease in the energies of ionization and excitation of pairs to the 2S state compared with the energies for remote hydrogen atoms. One can see that these regions lie at distances $r = (2–5)a$, where a is the Bohr radius of the hydrogen atom. One can further see that the differences between the energies of ionization and transitions to the first excited state differ significantly. Note that, for other systems similar to diatomic molecules, intervals of distances r must also exist in which the energies of ionization and excitation of pairs are less than those in the case of isolated partners. This is associated with the fact that the interaction energy in the case of a neutral pair decreases with increasing distance r faster than in the case of a singly ionized pair.

A similar situation must be observed in the case of impurity pairs in crystals, but on different scales of distances and energies depending on the Bohr radii a of impurities. For example, the results of calculation for an acceptor molecule in a cubic crystal have demonstrated [10] that the interaction energy in such a system at $r \gg a$ decreases as r^{-5}. However, in contrast to regular diatomic molecules, the absorption spectrum must be a superposition of the pair contribution under conditions of random distribution of impurities, and, in this case, broad bands rather than narrow lines must show up in the absorption spectra. The structure of these bands may be different for transitions to different excited states of each of the impurities. In this case, only those electronic transitions whose energies are less than those for analogous transitions in solitary impurity atoms may be observed in pairs during absorption. Transitions with high energies will find themselves in regions of strong absorption by single impurities and may hardly be observed against this background. So, the existence of impurity pairs brings about the emergence in doped silicon of a whole spectrum of local states, both deeper and shallower than the ground states of impurities. In compensated silicon, the photoexcitation must lead to optical recharging of impurity states. The relaxation of optical recharging at low temperatures is a slow process whose rate depends on the experimental conditions. This fact may prove to be important in studying the kinetics of electronic transitions in doped silicon. As to the validity of the term “molecules” for impurity pairs, this term, though incorrect, points to the existence of an analogy between them. Indeed, pairs represent a certain group of interacting impurity atoms with a characteristic electronic energy spectrum independent of the impurity concentration.

The concentration N_2 of impurity molecules may be estimated assuming that the oscillator strengths in pairs and in solitary impurity atoms for similar transitions differ insignificantly. First of all, we will use the results of [5, 6] for an indium impurity in silicon. According to [6], in the photoconductivity spectrum of Si[In] in the spectral range in which the impurity response is determined by pairs, the signal is approximately two orders of magnitude lower than in the region of photoionization of single atoms with the concentration $N = 3.1 \times 10^{17}$ cm^{-3}. From this, it follows that the pair concentration $N_2 = (3–5) \times 10^{15}$ cm^{-3}. One can use the absorption spectrum of Si[In], given in [5], to estimate N_2 by the area of one of the bands by comparing it with the area of the respective band for solitary impurity atoms of indium with the known concentration N [1]. For the sample with $N = 4 \times 10^{17}$ cm^{-3} in Fig. 1, the thus estimated concentration N_2 is 6×10^{15} cm^{-3}. These concentrations of impurity pairs of indium are also given in Fig. 2. One can make a similar estimation for a boron impurity, for example, for a sample with $N = 7.6 \times 10^{16}$ cm^{-3} (Fig. 1). We assumed that a band with a maximum at $h\nu = 23$ meV was due to the transition of a hole in a pair to the first excited state of the series $P_{3/2}$ and derived $N_2 = 6 \times 10^{15}$ cm^{-3}. These estimates appear to be reasonable if we assume that the main contribution to long-wave absorption is made by impurity pairs which are spaced two–five Bohr radii from one another and have no closer neighbors.

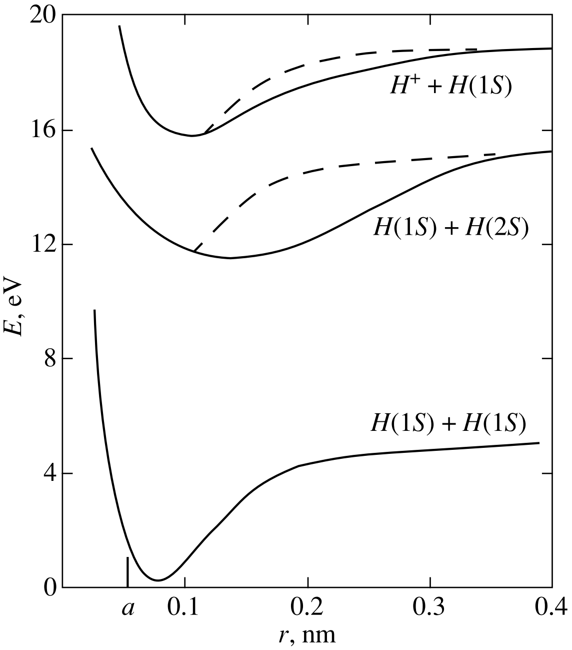


Fig. 3. The electron terms of a hydrogen molecule; a is the Bohr radius. The differences in the energy E between the dashed and solid lines correspond to a decrease in the energies of ionization and excitation at distances r between protons and at $r = \infty$.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-16078) and by the Program for Support of Leading Scientific Schools (project no. 00-15-96663).

REFERENCES

1. A. K. Ramdas and S. Rodrigues, Rep. Prog. Phys. **44**, 1278 (1981).
2. H. Chandrasekhar, A. K. Ramdas, and S. Rodrigues, Phys. Rev. B **12**, 5780 (1975).

3. Ya. E. Pokrovskii, O. I. Smirnova, and N. A. Khvalkovskii, Solid State Commun. **93**, 458 (1995).
4. Ya. E. Pokrovskiĭ, O. I. Smirnova, and N. A. Khval'kovskiĭ, Zh. Éksp. Teor. Fiz. **117**, 457 (2000) [JETP **90**, 404 (2000)].
5. W. Scott, Appl. Phys. Lett. **32**, 540 (1978).
6. R. Baron, M. H. Young, J. K. Neeland, *et al.*, Appl. Phys. Lett. **30**, 594 (1977).
7. Ya. E. Pokrovskiĭ, O. I. Smirnova, and N. A. Khval'kovskiĭ, in *Proceedings of the V Russia Conference on Physics of Semiconductors, Nizhni Novgorod, 2001*, p. 189.
8. N. Sclar, Progr. Quant. Electr. **9**, 149 (1984).
9. *Tables of Physics Quantities: Reference Book*, Ed. by K. I. Kikoin (Atomizdat, Moscow, 1976).
10. Sh. M. Kogan and A. F. Polupanov, in *Proceedings of the 3rd International Conference on Shallow Impurities in Semiconductors*, Ed. by B. Monemar, Inst. Phys. Conf. Ser., No. 95, 527 (1988).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 95, No. 1, 2002, pp. 87–100.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 122, No. 1, 2002, pp. 101–115.
Original Russian Text Copyright © 2002 by Ovchinnikov, Ovchinnikova.

SOLIDS
Structure

Models of Spin Structures in Sr_2RuO_4

A. A. Ovchinnikov[a, b, *] **and M. Ya. Ovchinnikova**[a]

[a]*Institute of Chemical Physics, Russian Academy of Sciences, Moscow, 117977 Russia*
[b]*Max-Planck Institute for Physics of Complex Systems, Dresden, 01187 Germany*
**e-mail: aovchin@mpipks-dresden.mpg.de*

Received February 6, 2002

Abstract—The mean-field method is used to calculate the bands, Fermi surfaces, and spin susceptibilities of a three-band model of the RuO_4 plane of Sr_2RuO_4 rutinate for states with different spin structures. In particular, the spiral state is studied with the "incommensurate" vector $Q = 2\pi(1/3, 1/3)$ corresponding to the nesting of bands with the population $n = 4$. This state proves to be the lowest with respect to energy among other (paramagnetic, ferromagnetic, antiferromagnetic, and periodic) solutions. In the spiral state, in addition to the main α, β, and γ sheets of the Fermi surface, shadow Fermi boundaries along the $\Gamma(0, 0)$–$M(\pi, 0)$ line (previously observed in the ARPES experiments) are revealed and explained. This may change the interpretation of the data on dispersionless peaks in photoemission, previously ascribed to surface states. The spin susceptibilities of the spiral state exhibit peaks in the dependence $\mathrm{Im}\chi(q, \omega)$ at $q = Q$ in accordance with the observed magnetic peak in neutron scattering. The hypothesis of the presence of spin structures with $q = Q$ in the normal state of Sr_2RuO_4 and the methods of checking this hypothesis are discussed. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Single-layer quasi-two-dimensional rutinate attracts attention as a superconductor ($T_c \sim 1$ K) with a possible triplet type of pairing [1–3]. Arguments in favor of triplet superconductivity are provided by the behavior of the Knight shift [4, 5] and by the possible part played in a superconducting transition by ferromagnetic fluctuations. The latter are assumed by analogy with the ferromagnetism of the parent cubic perovskite $SrRuO_3$. The determination of the type and mechanism of pairing depends directly on the understanding of the electronic and magnetic structures of the normal state. The band structure and the properties of magnetic fluctuations of Sr_2RuO_4 were investigated in [6–8]. One of the singularities of rutinates is the dependence of their properties on the composition of the compound. So, substituted analogs of $Ca_{2-x}Sr_xRuO_4$ are characterized by a very complex magnetic phase diagram which is controlled by structural distortions of the lattice and includes phases of paramagnetic (PM) or ferromagnetic (FM) metal or antiferromagnetic (AFM) dielectric [9–11].

In contrast to substituted compounds, the lattice of original Sr_2RuO_4 remains undistorted. Measurements of inelastic neutron scattering revealed incommensurate magnetic fluctuations with the quasi-momentum $Q \sim (0.3, 0.3, 1)$ [in $(2\pi/a, 2\pi/a, 2\pi/c)$ units] [12]. In the weakly substituted compound $Sr_2Ru_{1-x}Ti_xO_4$, dynamic fluctuations with the same values of Q are replaced by a static incommensurate order with the same values of Q [13]. The position of peaks of magnetic susceptibility $\mathrm{Im}\chi(q, \omega)$ in Sr_2RuO_4 is related to the structure of valence bands [6, 7]. Namely, for quasi-one-dimensional bands d_{xz} and d_{yz}, the Fermi surface is characterized by nesting with the quasi-momentum $Q = (0.3, 0.3, q_z)$, i.e., by the presence of parallel segments of the Fermi surface which are matched during the shift $k \longrightarrow k + Q$. At the same time, high values of homogeneous static susceptibility and heat capacity [14, 15], as well as the high ratio U/t of on-site repulsion to the widths of bands of suitable models [7, 8], point to strong electric correlations in these compounds.

Different approaches exist to the description of the magnetic and electronic properties of correlated systems under conditions of nesting. In one of these approaches, the starting point is the Lindhard zero magnetic susceptibility $\chi_0(q, \omega)$. Nesting at some $q = Q$, leads to a maximum in the dependence $\chi''(q, \omega)$ at $\omega \longrightarrow 0$ and $q \longrightarrow Q$. The nesting vectors of zero electron bands $Q = (\pi, \pi)$ in cuprates or $Q = (2\pi/3, 2\pi/3)$ in Sr_2RuO_4 indeed correspond to the observed peaks of low-frequency absorption $\chi''(q, \omega)$ in these compounds. However, the absolute values of integral intensity, calculated using χ_0, differ by an order of magnitude from the values observed in strongly correlated compounds such as cuprates and rutinates. The contradiction is removed by renormalization of susceptibility of the type $\chi(q, \omega) = \chi_0/[1 - J(q)\chi_0]$ or $\chi(q, \omega) = \chi_0/[1 - U\chi_0]$ by way of exchange or on-site interaction in the t–J model or Hubbard model [16, 17]. The renormalization enables one to describe a high density of low-frequency excitations and a high susceptibility only in the case of smallness of denominators in χ, i.e., in a situation close to that of instability of a correlated system relative to the emergence of a spin structure with a given value of q. In that case, however, the procedure of renormalization on the basis of zero spectrum may itself become inconsistent.

Another way of describing the electronic structure under conditions of nesting is to use the mean-field method which initially involves a concrete spin structure which removes the instability caused by nesting. This approach may reveal the real mechanism of renormalization of bands of a strongly correlated system. Such an approach corresponds to the treatment of antiferromagnetic or spiral spin structures as a zero approximation [18–20]. In this case, the spin order following from uniform solutions by the mean-field method should be more likely treated as local. The presence of the spin moment $d_n = \langle S_n \rangle$ localized at the site n causes a reduction of the on-site energy. A certain alignment of local spins under conditions of nesting, while removing the degeneracy, produces an additional gain in energy.

The main features of the Fermi surface of Sr_2RuO_4, obtained both from observations of magnetic quantum oscillations [21, 22] and from photoemission measurements (ARPES) [23–26], are largely reproduced by the three-band model suggested in [6–8] and by the results of calculation of the band structure in the LDA approximation [27, 28]. However, in addition to the main Fermi surfaces, the data of ARPES point to a clearly defined structure of shadow Fermi boundaries which are due to the umklapp processes. In particular, the sharp dispersionless peak, observed in the ARPES experiments [26] and corresponding to the shadow Fermi boundary along the lines $\Gamma(0, 0)$–$M(\pi, \pi)$, remains an enigma. In [26], it is attributed to the surface states arising in the distorted lattice of the surface layer. The shadow Fermi surfaces may point to the presence of periodic spin structures or lattice distortions.

It is our objective to study the manifestation of spin structures, in particular, with an incommensurate order, in the shape of a Fermi surface, in the structure of shadow Fermi surfaces, and in the magnetic susceptibility of the three-band model of Sr_2RuO_4. The calculations are performed in the mean-field approximation. In particular, the question is raised as to the possible bulk origin of the dispersionless peak along the lines Γ–M of the momentum space. The hypothesis of the spiral spin order with incommensurate momentum $Q = 2\pi(1/3, 1/3)$ is discussed. Here, we continue using the habitual but inaccurate term "incommensurate," although it would be more correct to refer to this structure with the period which is a multiple of the lattice period as "fractional quasi-momentum." Such a comparison may prove useful in revealing magnetic structures which, in the dynamic or static modes, may be present in the ground state of Sr_2RuO_4 rutinate.

2. THREE-BAND MODEL OF RuO_4 PLANE

In an approximate ionic model of $RuO_4 = Ru^{4+}(d^4)(O^{2-})_4$, four electrons take up three lower d orbitals d_{xy}, d_{xz}, and d_{yz} in the t_{2g}-symmetry field produced by oxygen ions [1, 2]. These orbits, involving p_π orbits of oxygen, generate three valence bands with the total population of four electrons per lattice site. In the previously suggested models of strong coupling [6–8], the lattice site states and the corresponding creation operators $c^\dagger_{\nu\sigma n}$ ($\nu = 1, 2, 3$) belong to Wannier combinations d–p_π of the respective α, β, and γ bands with the symmetry of xz, yz, and xy of d orbits, respectively. The Hamiltonian of the model has the form [8]

$$H = \sum_{\nu,\sigma}\sum_{k} \epsilon_{\nu k} c^\dagger_{\sigma\nu k} c_{\sigma\nu k} + T_{12} + \sum_{n,\nu}\Bigg\{ U n_{\uparrow\nu n} n_{\downarrow\nu n} + (2U_1 - J)\sum_{\nu'\neq\nu} \frac{1}{4} n_{\nu n} n_{\nu' n} - J\sum_{\nu'\neq\nu} \mathbf{S}_{\nu n}\mathbf{S}_{\nu' n}\Bigg\}. \tag{1}$$

Here, $n_{\nu\sigma n}$, $n_{\nu n} = n_{\nu\uparrow n} + n_{\nu\downarrow n}$, and $\mathbf{S}_{\nu n}$ denote the population and spin operators of the respective α, β, and γ orbits ($\nu = 1, 2, 3$). If the interband hopping interaction is disregarded, the energies of zero bands are

$$\epsilon^0_{\nu k} = -t^\nu_0 - 2t^\nu_x \cos k_x - 2t^\nu_y \cos k_y + 4t^\nu_{xy} \cos k_x \cos k_y. \tag{2}$$

The following values (in eV) found in [8] were used for the quantities t^ν_0 and the hopping integrals between the neighboring and next neighboring centers:

$$\{t_0, t_x, t_y, t_{xy}\} = \begin{cases} \{0.24, 0.31(0.045), 0.045(0.31), 0.01\} \\ \{0.50, 0.44, 0.44, 0.14\} \end{cases} \tag{3}$$

for the $\alpha(\beta)$ and γ bands ($\nu = 1(2), 3$). The values of $\{U, U_1 = U - 2J, J\} = \{1.2 \text{ eV}, 0.8 \text{ eV}, 0.2 \text{ eV}\}$ were taken for the interaction parameters. These parameters are defined by the intra-atomic Coulomb and exchange integrals related by the relation following from the Racah rules [8]. The operator of interband interaction $T_{\alpha\beta}$ in Eq. (1) (diagonal jumps with the change of the type of orbit) has the form

$$T_{\alpha\beta} = \sum_{k,\sigma} 4t_{\alpha\beta} \sin k_x \sin k_y (c^\dagger_{\sigma 1k} c_{\sigma 2k} + \text{H.c.}). \tag{4}$$

It is only such interaction that is allowed by the symmetry for an undistorted lattice (in a distorted lattice with alternating rotations of octahedrons, as in Sr_2RuO_4, the hopping between neighbors also make a contribution to interband interaction). The parameter $t_{\alpha\beta}$ is expected to be small, of the order of 0.01 eV.

Consider the scheme of derivation of mean-field equations for states with a certain order of local spins $\langle S_n \rangle$. If the spin–orbit interaction is ignored, the "spin" coordinate system defining the spin quantization axis is not related to the lattice coordinate system with the z axis along the c axis of crystal. However, for simplicity of description, we assume these coordinate systems to coincide, with the possibility of recovering the difference between them in the final result. With this reservation, we will select the following orientation of average spins:

$$\langle S_n^\nu \rangle = d_\nu [\mathbf{e}_x \cos Qn + \mathbf{e}_y \sin Qn]. \qquad (5)$$

For the spiral (SP), FM, and AFM states being treated, the vector Q is

$$Q = 2\pi\left(\frac{1}{3}, \frac{1}{3}\right)\text{—SP}, \quad Q = (0, 0)\text{—FM}, \qquad (6)$$
$$Q = (\pi, \pi)\text{—AFM}.$$

respectively. Given this choice of quantization axes, the form of the mean-field equations is the same for all three states, differing only by the helicity vector.

In the mean-field approximation, the wave function Φ_Q is determined by the population of electron eigenstates $b^\dagger_{\lambda k}$ of linearized Hamiltonian H_{lin},

$$[H_{\text{lin}}, b^\dagger_{\lambda k}] = E_{\lambda k} b^\dagger_{\lambda k};$$

$$b^\dagger_{\lambda k} = a^\dagger_{i\nu k} S_{i\nu}(k); \quad \lambda = 1, \ldots, 6; \qquad (7)$$
$$i = 1, 2, \quad \nu = 1, 2, 3.$$

The set of basis operators $a^\dagger_{i\nu k}$ is defined as

$$a^\dagger_{i\nu k} = \{a^\dagger_{1\nu k}, a^\dagger_{2\nu k}\} = \{c^\dagger_{\uparrow\nu k}, c^\dagger_{\downarrow\nu, k+Q}\}. \qquad (8)$$

The subscript $\nu = 1, 2, 3$ corresponds to the α, β, and γ bands, and $i = 1, 2$ corresponds to spin projections.

The average energy for the Φ_Q state per lattice site is

$$\bar{H}(\bar{T}, d_\nu, r_\nu) = \bar{T} + H_U(r_\nu, d_\nu),$$

$$H_U = \sum_\nu \left\{ U(r_\nu^2 - d_\nu^2) + \sum_{\nu' \neq \nu} [U_2 r_\nu r_{\nu'} - J d_\nu d_{\nu'}] \right\}. \qquad (9)$$

Here, $U_2 = 2U - 5J$. The energy given by Eq. (9) is a function of one-electron averages (average density, local spin, and kinetic energy per lattice site) which, in the basis given by Eq. (8), have the form

$$r_\nu = \frac{1}{2N} \sum_{i=1,2} \sum_k \langle a^\dagger_{i\nu k} a_{i\nu k} \rangle,$$

$$d_\nu = \frac{1}{2N} \sum_{i=1,2} \sum_k \langle a^\dagger_{i\nu k} a_{3-i,\nu,k} \rangle, \qquad (10)$$

$$\bar{T} = \frac{1}{N} \sum_k [\epsilon_{\nu k} \langle a^\dagger_{1\nu k} a_{1\nu k} \rangle + \epsilon_{\nu, k+Q} \langle a^\dagger_{2(\nu k)} a_{2\nu k} \rangle].$$

The linearized Hamiltonian of the mean field is

$$\hat{H}_{\text{lin}} = \hat{T} + \frac{\partial \bar{H}}{\partial r_\nu} \hat{r}_\nu + \frac{\partial \bar{H}}{\partial d_\nu} \hat{d}_\nu = \sum_k \hat{h}_k, \qquad (11)$$
$$\hat{h}_k = h_{i\nu, j\nu'} a^\dagger_{i\nu k} a_{j\nu' k}.$$

Here, $\hat{r}_\nu$, $\hat{d}_\nu$, and $\hat{T}$ are operators corresponding to the averages given by Eqs. (10), and the expressions for the matrices $h_{i\nu, j\nu'}$ follow from Eqs. (10) and (11). The proper operators $b^\dagger_{\lambda k}$ in Eq. (7) are determined by diagonalization of $h(k)$,

$$h_{i\nu, j\nu'}(k) S_{j\nu'\lambda} = S_{i\nu\lambda} E_\lambda(k). \qquad (12)$$

The quantities in Eqs. (10), in turn, are found by substitution of the expressions

$$\langle a^\dagger_{i\nu k} a_{j\nu' k} \rangle = \sum_\lambda S_{i\nu\lambda}(k) S^*_{j\nu'\lambda}(k) f(E_{\lambda k} - \mu) \qquad (13)$$

into (10). Here, f is the Fermi function, and $E_{\lambda k}$ denotes the eigenenergies of Hamiltonian h_k. The chemical potential μ is found from the condition $2\sum_\nu r_\nu = 4$. Note that, with $t_{\alpha\beta} = 0$, i.e., in the absence of hopping between the α and β orbits, the matrix $S_{i\nu\lambda}$ is factorized so that each one-electron state $b^\dagger_{\lambda k}$, $\lambda = \tilde{\lambda} + 2(\nu - 1)$, $\tilde{\lambda} = 1, 2$, belongs to a certain band ν. In this manner, we derive a self-consistent solution of mean field for each one of the types of spin order given in Eqs. (5) and (6).

In addition to the states given by Eqs. (5) and (6), we sought the mean-field states with charge and spin density waves that would have a period equal to three lattice constants in both the x and the y directions and a symmetry C_{4v} (see Fig. 8 below). Such structures, as well as the spiral state with $Q = 2\pi(1/3, 1/3)$, remove the degeneracy of levels which is due to nesting with the population $n = 4/3$ for each band. Centers of four sorts ($l = 1, \ldots, 4$) are characterized by the densities and values of local spins $r_{\nu l}$ and $d_{\nu l}$ of each one of the ν orbits (for the γ band, the lattice sites of the second and third sorts are equivalent). The local spin vectors were

Table 1. The average energies, the ratios of band populations, and the local moments for different structures, namely, for spiral, block-antiferromagnetic, ferromagnetic, antiferromagnetic, and paramagnetic structures. The states are listed in the order of increasing energies

	SP	BAFM	FM	AFM	PM
$\bar{H}$, eV	2.708	2.7104	2.7469	2.7552	2.7570
$2r_\alpha/r_\gamma$	1.935	1.4–3.2	2.15	2.137	2.04
$d_\alpha(=d_\beta)$	0.234	0.17–0.34	0.254	0.078	0
d_γ	0.160	0.11–0.20	0.181	0.093	0

assumed to be collinear, $\langle \mathbf{S}_n^\nu \rangle = d_\nu(l)\mathbf{e}_z$, and the search for a self-consistent solution by the mean-field method was performed for a block-ferromagnetic (BFM) or block-antiferromagnetic (BAFM) structure,

$$\langle S(n_x + 3m, n_y + 3m') \rangle = \zeta^{m+m'} \langle S(n_x, n_y) \rangle,$$

with $\zeta = +1$ or $\zeta = -1$, respectively. The mean-field equations for these structures are discussed in Appendix I. A self-consistent solution could be found only for a BAFM structure with $\zeta = -1$. It contains 18 sites in a unit cell and consists of almost ferromagnetic 3 × 3 blocks with opposite orientations of spins of neighboring blocks.

The investigation of main and shadow Fermi surfaces of each one of the bands involved the calculation of the intensity $A(k, \omega)|_{\omega = 0}$ with which these Fermi surfaces show up in photoemission. The required spectral functions are

$$A(k\omega) = \frac{1}{Z}\sum_{\sigma, m, n} |\langle m|c_{\sigma k}|n\rangle|^2 \times e^{-\beta E_n}\delta(\omega - E_n + E_m - \mu), \tag{14}$$

where Z is the partition function and $|n\rangle$ and $|m\rangle$ are the states of the entire system. In the mean-field approximation, with due regard for the definition of the basis given by Eq. (8), we derive

$$A(k\omega) = \sum_\sigma A_\sigma(k\omega), \tag{15}$$

$$A_\uparrow(k\omega) = \sum_{\nu, \lambda} |S_{1\nu\lambda}(k, \omega)|^2 \tilde{\delta}(E_\lambda(k) - \mu - \omega),$$

$$A_\downarrow(k\omega) = \sum_{\nu, \lambda} |S_{2\nu\lambda}(k - Q, \omega)|^2 \times \tilde{\delta}(E_\lambda(k - Q) - \mu - \omega). \tag{16}$$

Here, $E_\lambda(k)$ denotes one-electron energies, and $S_{i\nu\lambda}$ is the matrix of coefficients in the expression for the proper operators in Eq. (7) in the basis given by Eq. (8). The calculation of $A(k, \omega)$ involved the replacement of the δ function by the smeared function of one of two forms,

$$\delta(\epsilon) \longrightarrow \tilde{\delta}(\epsilon) = \frac{1}{4\gamma}\cosh^{-2}(\epsilon/2\gamma),$$

$$\delta(\epsilon) \longrightarrow \tilde{\delta}(\epsilon) = \frac{1}{\pi}\frac{\gamma}{\epsilon^2 + \gamma^2},$$

with the width $\Delta\omega \sim \gamma$.

The two-dimensional image of the function $A(k_x, k_y, \omega = 0)$ on the k_x, k_y plane gives a view of all sheets of the Fermi surface of three bands and of the intensity of the respective peaks in photoemission. A similar two-dimensional map

$$\tilde{A}(s, \omega) = A(k_x(s), k_y(s), \omega)$$

presents the dispersion of bands upon variation of k along the parametrically preassigned contour $\mathbf{k}(s) = \{k_x(s), k_y(s)\}$ on the phase plane.

3. FERMI SURFACES AND BAND DISPERSION

Table 1 gives the values of average energy (per lattice site), the ratios of densities, and the values of local spins on the α, β, and γ orbits for the spiral state given by Eqs. (5) and (6), and for the BAFM, FM, AFM, and PM states of a model with the parameters given in Eq. (3) and $t_{\alpha\beta} = 0$–0.03. The states are arranged in order of increasing energies.

The spiral and periodic (BAFM) states turn out to be lower with respect to energy. Both these states have a period corresponding to the nesting of bands at $n = 4$. The total spin moments $\langle S_t \rangle = \sum_\nu d_\nu \sim 0.6$ for three lower states are close to one another.

Despite a small difference in energies, the states have entirely different structures of the Fermi surfaces. The Fermi surfaces of different structures, i.e., images of the functions $A(k_x, k_y, \omega = 0)$, are given in Figs. 1–3. For the PM state, the well-known ([6–8]) pattern of the Fermi surface is observed (Fig. 1): quasi-one-dimensional sheets at $k_x(y) = \pm 2\pi/3$ correspond to the α and β bands, and the electron-type Fermi surface corresponds to the γ band. With $t_{\alpha\beta} \neq 0$, the Fermi surfaces for the α and β bands are split at the point of their intersection, which results in the emergence of two Fermi surfaces of the electron type and one of the hole type. Their parameters (the areas covered by them) are in good agreement with the characteristics determined from the measurements of magnetic quantum oscillations of the de Haas–van Alphen effect [21, 22]. A similar pattern of Fermi surface was obtained as a result of photoemission measurements (ARPES) for samples under conditions of cyclic variation of temperature in the range from 10 to 200 K [26]. It is assumed that such a procedure prevents the formation of surface and bulk superstructures of RuO_4 planes. The deviation of the

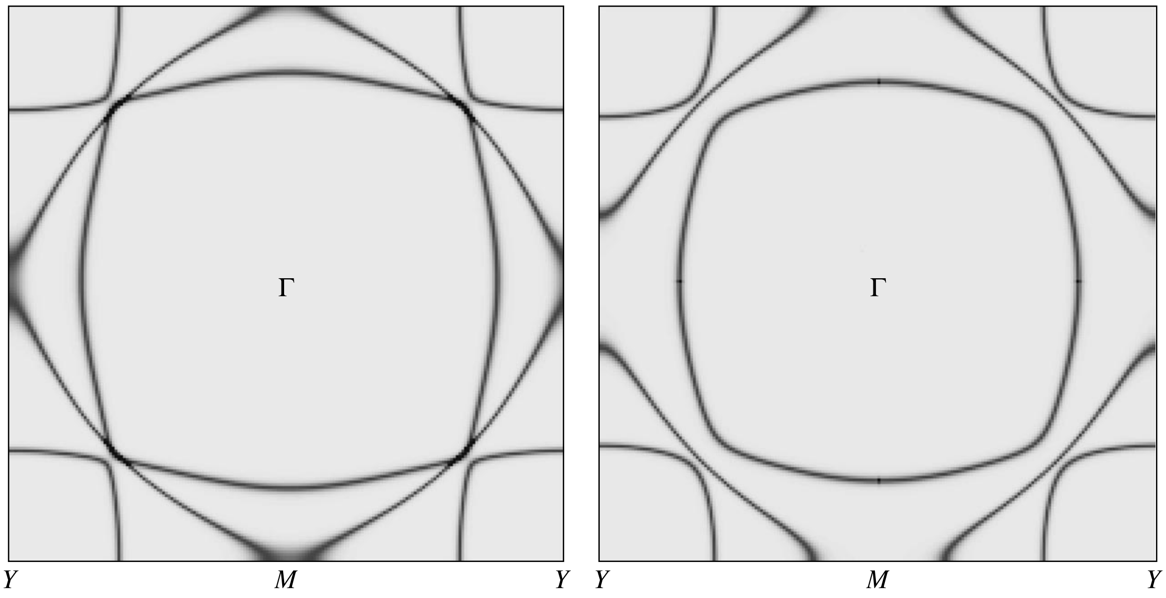


Fig. 1. The Fermi surfaces of the PM state for a model with parameters given in Eq. (3) and $t_{\alpha\beta} = 0$ (on the left). On the right, the same, but with the changed $t_{\alpha\beta} = 0.03$ and $t_0^\gamma = 0.6$ eV.

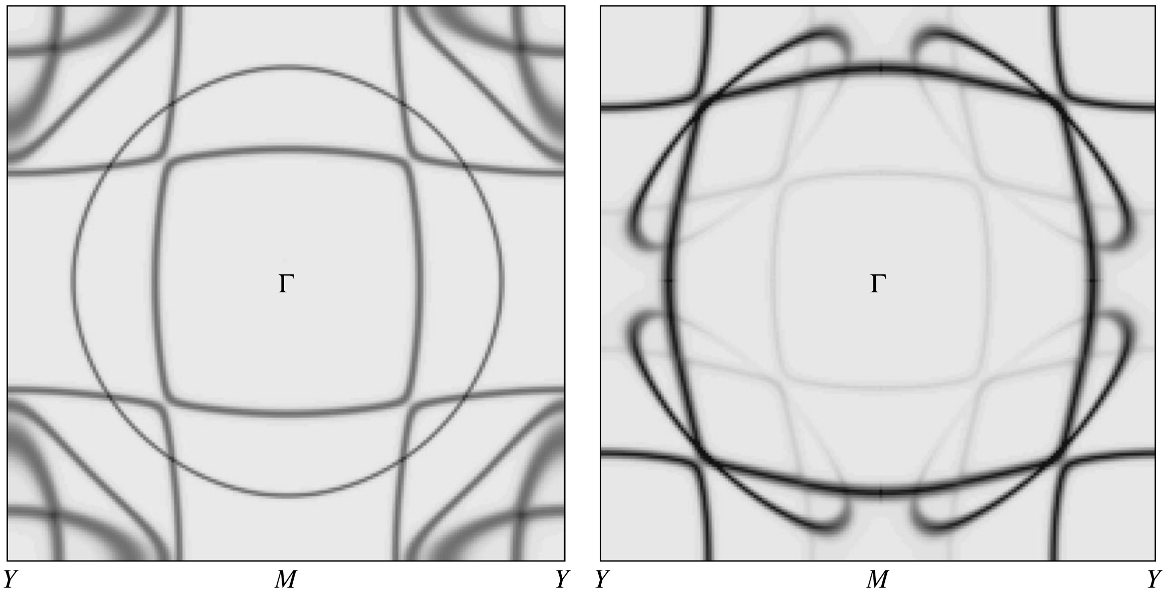


Fig. 2. The Fermi surfaces for the FM and AFM states with parameters given in Eq. (3) and $t_{\alpha\beta} = 0.01$ eV.

model parameters from the standard parameters given in Eq. (3) causes a variation of the Fermi surface. For example, an increase in $t_{\alpha\beta}$ draws apart the Fermi surfaces of the α and β bands in the region of their quasi-intersection, and a change of the difference $t_0^\gamma - t_0^{\alpha(\beta)}$ causes a variation of the relative arrangement of the sheets of the $\alpha(\beta)$ and γ bands (see Fig. 1).

Figure 2 gives the Fermi surfaces for the FM and AFM states. In both cases, they differ strongly from the Fermi surfaces of the PM state. A double set of Fermi surfaces for bands with spins $\sigma = \uparrow$ or $\downarrow$ is observed; distinction is made between bands with minimal and maximal population. Here, the Fermi surfaces are related to ferromagnetic states in an undistorted lattice, although the latter states, observed in the substituted compound $Ca_{2-x}Sr_xRuO_4$ [9, 10], correspond to a lattice with alternating rotations of RuO_4 octahedrons. For the AFM states, the Fermi surface of the γ band has a characteristic form of pockets familiar from the form of

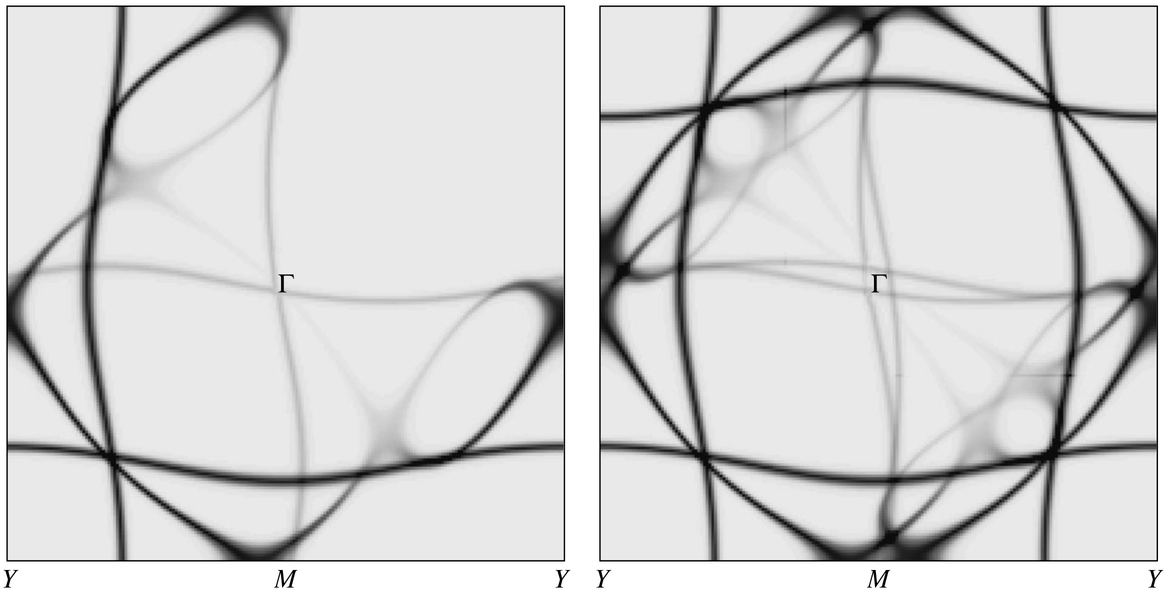


Fig. 3. The Fermi surfaces of a system in the spiral state with $Q = 2\pi(1/3, 1/3)$. On the right, the image of the total spectral function (15) of three bands, and on the left, the image of the spectral function $A_\uparrow(k, \omega = 0)$ for one of the polarizations ($\sigma = \uparrow$).

hole pockets in cuprate models. The difference consists in that the shadow part of the pocket is now facing the center $\Gamma(0, 0)$ of the Brillouin zone rather than the point $Y(\pi, \pi)$, because the band population $n_\gamma > 1$. Because of magnetic doubling of a unit cell in the AFM state, shadow Fermi surfaces arise for the α and β bands in addition to main sheets; these shadow surfaces are the reflection of the main Fermi surfaces relative to the $k_x \pm k_y = \pm\pi$ lines. Note that such reflected shadow Fermi surfaces are indeed observed in the ARPES spectra of Sr_2RuO_4 [26]. However, their origin may be associated both with the bulk spin AFM structure and with the lattice rearrangement of the surface layer with the cell doubling to the size of $\sqrt{2} \times \sqrt{2}$ or with surface antiferromagnetism. For example, Damascelli *et al.* [26] assumed the lattice distortion and the ferromagnetism of the surface layers in order to interpret the presence of dispersionless bands (see the discussion below).

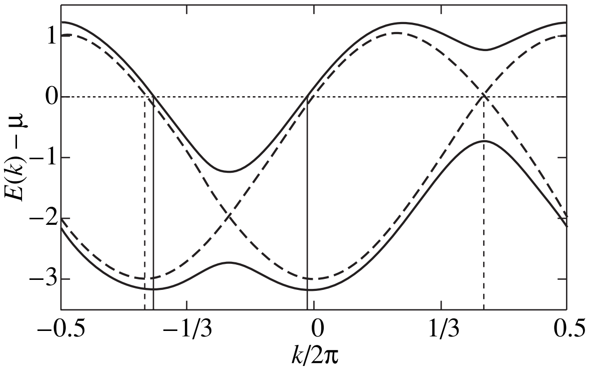


Fig. 4. The band energies (solid curves) for a one-dimensional system with the population $n = 2/3$ in the spiral state with $Q = 2\pi/3$. The vertical lines indicate the position of the Fermi boundaries corresponding to one of the polarizations, i.e., the image of the spectral function $A_\uparrow(k, \omega = 0)$. The dashed curves indicate the zeroth bands $\epsilon_k - \mu$ and $\epsilon_{k+Q} - \mu$, and the dashed vertical lines correspond to zeroth Fermi boundaries.

Figure 3 gives the Fermi surfaces of the spiral state with $Q = (2\pi/3, 2\pi/3)$. Note three singularities of a Fermi surface of the spiral state, namely: (1) the emergence of shadow Fermi surfaces along the lines $k_x = 0$ or $k_y = 0$ just at the locations of the observed dispersionless peaks in photoemission, (2) different images of the spectral function given by Eqs. (16) pertaining to different spin polarizations and contributing to the total intensity of photoemission, and (3) residual shadow Fermi surfaces of the γ band in the direction perpendicular to the helicity vector. They retain the form of electron pockets similar to those of the AFM state but shifted by $\pm Q/2$ for $A_{\uparrow(\downarrow)}(q, \omega = 0)$. In the representation of the function $A_\uparrow(k_x, k_y, \omega = 0)$, we are dealing with the spin polarization on the z' axis perpendicular to the plane of spin rotation of the spiral state. The possible orientation of z' relative to the lattice axis z remains unknown.

The first two features may be interpreted using a simple example of the spiral state of a one-dimensional one-band system with the occupancy $n = 2/3$ (see Fig. 4). In the spiral state with $Q = 2\pi/3$, the band energies $E_\lambda(k)$, $\lambda = 1, 2$, are asymmetric relative to $k = 0$ (with the basis $\{c^\dagger_{\uparrow k}, c^\dagger_{\downarrow(k+Q)}\}$). The image of the spectral function $A_\uparrow(k, \omega)$ will indicate the main (with a high peak in $\mathrm{Im}A_\uparrow$) Fermi boundary at $k \approx -2\pi/3$ and the shadow (with a low peak in $\mathrm{Im}A_\uparrow$) boundary at $k \approx 0$.

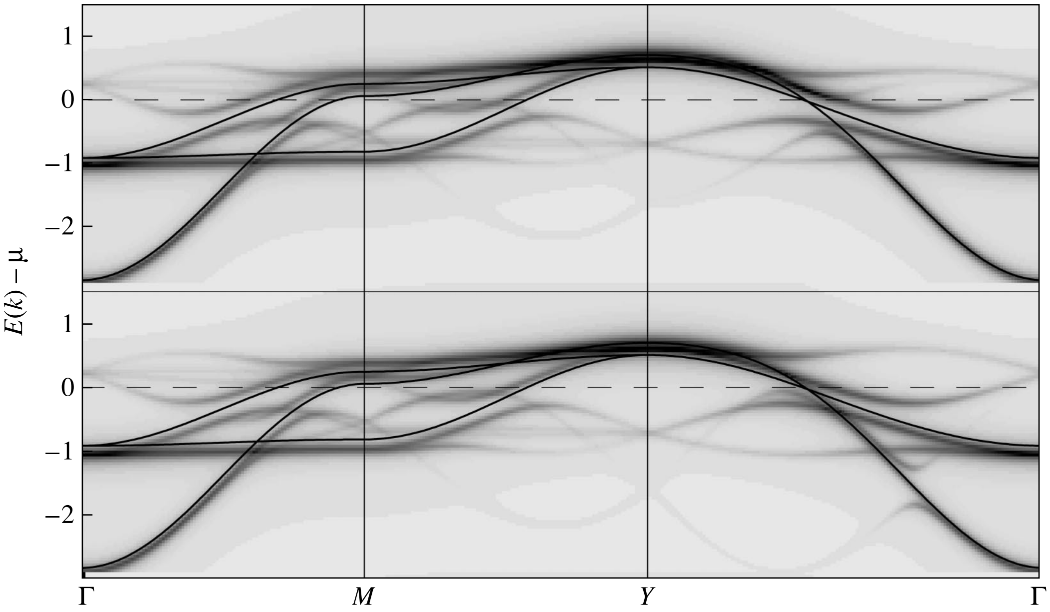


Fig. 5. The band dispersion obtained as the image of the total spectral function $A(k, \omega)$ for $k = (k_x(s), k_y(s))$ varying along two contours $\Gamma(0, 0)$–$M(0, \zeta\pi)$–$Y(\pi, \zeta\pi)$ with $\zeta = 1$ (at the top) and $\zeta = -1$ (at the bottom). The thin curves indicate zero band energies $(\epsilon_k - \mu)$ at $d_\nu = 0$.

The previous boundary at $k \approx 2\pi/3$ disappears due to band splitting. For the opposite polarization, an inverted pattern is observed: $A_\downarrow(k, \omega) = A_\uparrow(-k, \omega)$.

The shadow Fermi surfaces along the Γ–*M* lines are worthy of more detailed discussion (see Fig. 3). Their arrangement coincides with the position of dispersionless peaks observed in the ARPES experiments and referred to by Damascelli *et al.* [26] as "sharp Fermi surface" (SS). Previously [26], the origin of these peaks was associated with surface electron states. It was assumed that these states arose during the unit cell restructuring to the size of $\sqrt{2} \times \sqrt{2}$ in the surface layers of RuO_4 with rotations of oxygen octahedrons characteristic of ferromagnetic spin structures [9–11, 28]. One of the arguments in favor of this interpretation was the disappearance of SS peaks and of the respective Fermi surface under the action of cyclic temperature treatment of the material with a fast variation of T in the range from 10 to 200 K. It is assumed that such treatment causes the destruction of surface structures. The results of our calculation of the Fermi surface of the spiral state with $Q = 2\pi(1/3, 1/3)$ enable us to advance the hypothesis of the bulk origin of the SS peaks observed in the ARPES experiments along the Γ–*M* lines. We compare these peaks with the shadow Fermi boundaries arising along the lines $k_x = 0$ or $k_y = 0$ in the spiral state. In other words, these shadow Fermi boundaries may be interpreted as the main Fermi boundaries $k_{x(y)} = \pm 2\pi/3$ of the α and β bands, shifted by the vector $Q = 2\pi(1/3, 1/3)$ equal to the period of the spin structure in the system. In addition, the segment of the line Γ–*M* in the vicinity of the point *M* ($|k| > 2\pi/3$) touches the boundary of the electron pocket of the γ band, which causes an increase in the intensity of the total spectral density in this region. Therefore, we are dealing with umklapp processes for periodic structures with the period Q in the momentum space. To this, we must add the assumption that a cyclic variation of temperature causes the destruction of an equilibrium spiral (or periodic) structure with the period Q (as it was assumed to destroy the surface structure, according to the previous interpretation [26]). Conceivably, this process introduces so large a number of defects into the structure as to broaden this shadow Fermi boundary and make it unobservable.

One of the ways of testing this hypothesis could be provided by the search for the above discussed polarization asymmetry in the intensity of photoemission. The spin-polarization photoemission employed to identify the Fermi surface of ferromagnetic states may be useful for testing the spiral structure as well. However, we cannot predict the orientation of the "spin system of coordinates" of the spiral state (should this state be realized) if we fail to include the spin-orbit interaction in the treatment. The question also remains of the degree to which the spin structures may be regarded as static. It is also necessary to more thoroughly interpret other shadow Fermi boundaries observed in the photoemission of Sr_2RuO_4 which was not subjected to the procedure of cyclic variation of temperature. Yet another interesting theoretical question arises of whether the helicity of state may be reflected in the phase shifts of magnetic quantum oscillations and, as a result, be observed during measurements of such effects.

Figure 5 gives the dispersion of bands for the spiral state, obtained as the image of the total spectral density

$A(s, \omega)$ of all three bands (see Eq. (15)) upon variation of $k = (k_x(s), k_y(s))$ along the contours $\Gamma(0, 0) - M(\zeta\pi, 0) - Y(\zeta\pi, \pi)$, $\zeta = 1, -1$. For segments of the Y–Γ contour, which are parallel and perpendicular to the vector Q, the difference of the main (nonshadow) branches of band energies is small. Along the entire contour, the main branches almost repeat the band energies $(\epsilon_k - \mu)$ of linearized Hamiltonian (11) at $d_\nu = 0$, which are shown by the curves in Fig. 5.

The properties of solutions by the mean-field method with a periodic spin BAFM structure are described in Appendix II. We shall only note that, for this structure, Fermi surfaces exist for the γ band alone. For the α and β bands, the modulation of spins and electron densities brings about the emergence of a gap along all previous Fermi boundaries of zeroth α and β bands. This contradicts the observations of magnetic quantum oscillations [21], which reveal the Fermi surface of all three bands in a system. Neither does the intricate shape of the Fermi surface of the γ band correspond to the Fermi surface observed in photoemission [23–26]. For the spiral state, similar gaps are observed along separate regions of zeroth Fermi boundaries. This shows up as the disappearance of some regions of the main Fermi surfaces in the photoemission of electrons with one or another (certain) spin polarization (see Fig. 3).

4. SPIN SUSCEPTIBILITY IN THE SPIRAL STATE

Another method is known for studying spin structures with the aid of inelastic neutron scattering (INS). This method involves measurements of the dynamic spin susceptibility of a system. Recent INS measurements in Sr_2RuO_4 [12] revealed a magnetic peak for the incommensurate quasi-momentum $Q \sim (0.3, 0.3, 1)$ [in $(2\pi/a, 2\pi/a, 2\pi/c)$ units] at $T < 200$ K. Similar peaks were obtained in the case of susceptibility calculated for unperturbed bands and renormalized in accordance with the random phase approximation (RPA) [16, 17]. The matching of the absolute integral intensity of magnetic absorption with the quantity being observed requires parameters indicating that the system is close to instability [12]. Therefore, the results of calculation involving zeroth bands may be inadequate. In view of this, we performed an RPA calculation of $\chi(q\omega)$ using spiral states as a zeroth approximation. The spiral states with $Q = 2\pi(1/3, 1/3)$ represent the simplest of the possible structures removing the system instability due to nesting. Our treatment is similar to the calculations of $\chi(q, \omega)$ for antiferromagnetic (or spiral) states [18–20] in which the Hubbard band splitting is included even in the zeroth approximation. Unlike the one-band models in [18–20], Sr_2RuO_4 is described by a three-band model, and the deviation of the helicity vector $Q = 2\pi(1/3, 1/3)$ from $Q_{AFM} = (\pi, \pi)$ is significant.

We are interested in fluctuations of the q harmonics of both the total spin S_q and spins of different orbits,

$$S_{\nu, q} = \sum_n e^{iqn} S_{\nu n}, \quad S_q = \sum_\nu S_{\nu, q}. \tag{17}$$

Different Cartesian components $\chi_{ij}(q, \omega)$ ($i, j = x, y, z$ or 1, 2, 3) of the spin susceptibility tensor for the total spin or its orbital components are defined according to [30, 31] as

$$\chi_{ij}(q\omega) = -i\int_0^\infty dt e^{i(\omega + i\eta)t} \langle\langle S_{iq}(qt); S_{jq}^\dagger \rangle\rangle, \tag{18}$$

$$\chi_{ij}^{\nu\nu'}(q\omega) = -i\int_0^\infty dt e^{i(\omega + i\eta)t} \langle\langle S_{iq}^\nu(t); (S_{jq}^{\nu'})^\dagger \rangle\rangle, \tag{19}$$

where

$$\langle\langle A(t); B \rangle\rangle = \begin{cases} [A(t), B], & t > 0 \\ 0, & t < 0. \end{cases}$$

Orbit-dependent components $\chi^{\nu\nu}(q, \omega \longrightarrow 0)$ with different polarization for each one of the orbits ($\nu = \alpha, \beta, \gamma$) can be measured individually as well. For example, Imai *et al.* [32] extract them from the results of NMR measurements on ^{17}O oxygen atoms taking different positions in RuO_4 octahedrons. The total susceptibility given by Eq. (18) is $\chi = \sum_{\nu, \nu'} \chi^{\nu\nu'}$. For simplicity, we perform the calculation for a model of zero interband coupling $t_{\alpha\beta} = 0$ in Eq. (4). We use the simplest method of deriving $\chi(q\omega)$ from equations of motion [31]. This method is equivalent to the summation of ladder diagrams. For applying this method to spiral states, we will express the q components of spin and density in terms of band operators (7) of the spiral state of the mean field,

$$\begin{gathered} \{\rho_q^\nu, S_{1q}^\nu, S_{2q}^\nu, S_{3q}^\nu\}_l = \sum_{i, j} (\sigma_l)_{ij} r_{ij}^\nu(q), \\ l = 0, 1, 2, 3, \quad r_{ij}^\nu(q) = \frac{1}{N}\sum_k c_{i\nu, k+q}^\dagger c_{j\nu k}. \end{gathered} \tag{20}$$

Here, σ_0 and σ_l denote a unit matrix and Pauli matrices; the subscripts i and j are introduced instead of the $\pm 1/2$ spin projections. In the basis of band operators (7), we have

$$r_{ij}^\nu(q) = \sum_{k, i, j} (\sigma_l)_{ij} S_{i\nu\lambda}^*(k+q) S_{j\nu\lambda'}(k) b_{\lambda, k+q}^\dagger b_{\lambda' k}. \tag{21}$$

We will introduce the following notation for new operators dependent on space harmonics:

$$X_{\nu l}(q) = \begin{cases} \dfrac{1}{2}\sum_i (\sigma_l)_{ii} r^{\nu}_{ii}(q), \quad l = 0, 3, \\ \dfrac{1}{2}[(\sigma_l)_{12} r^{\nu}_{12}(q - Q) + (\sigma_l)_{21} r^{\nu}_{21}(q + Q)], \quad l = 1, 2. \end{cases} \tag{22}$$

We will characterize the thus defined operators by the common parameter q because of their common representation in terms of the band operators $b^{\dagger}_{\lambda k}$ of the spiral state,

$$X_{\nu l}(q) = \frac{1}{2N}\sum_k \sum_{i, j, \lambda, \lambda'} (\sigma_l)_{ij} S_{i\nu\lambda}(k + q) \times S_{j\nu\lambda'}(k)\theta_{\lambda\lambda'}(k, q), \tag{23}$$

$$\theta_{\lambda\lambda'}(k, q) = b^{\dagger}_{\lambda, k + q} b_{\lambda' k}, \tag{24}$$

where $l = 0, 1, 2, 3$. It follows from the representation of the Hamiltonian by Eq. (1) in the basis $\{ b^{\dagger}_{\lambda k} \}$ that, in the case of the spiral state, only the operator correlators $X(q)$ with the same values of q are nonzero,

$$\langle\langle X_{\nu l}(q\omega); X^{\dagger}_{\nu' l'}(q')\rangle\rangle = \delta_{qq'} G_{\nu l, \nu' l'}(q, \omega). \tag{25}$$

We calculate the correlators given by Eq. (25) in the RPA approximation using the method described by Izuyama *et al.* [31] (see also [18–20]). This method is fully equivalent to the summation of ladder diagrams. We proceed from the equations of motion for the operator given by Eq. (24),

$$i\frac{d}{dt}\langle\langle \theta_{\lambda\lambda'}(k, q, t); X^{\dagger}_{\nu' l'}(q)\rangle\rangle = \langle\langle [\theta_{\lambda\lambda'}(k, q, t), H]; X^{\dagger}_{\nu' l'}(q)\rangle\rangle + \delta(t)\langle [\theta, X^{\dagger}]\rangle. \tag{26}$$

In calculating the commutator $[\theta, H]$, we linearize the four-fermion contribution in the latter. We integrate Eq. (26) with respect to t with $e^{i\omega t}$ and with respect to k with the same weight functions as in Eq. (23) to derive the following algebraic equations for $G_{\nu l, \nu' l'}(q, \omega)$:

$$\{\delta_{ll'}\delta_{\nu\nu'} - Z^{\nu}_{ll'}[U\zeta_{l'}\delta_{\nu\nu'} + K(l')(1 - \delta_{\nu\nu'})]\} G_{\nu' l', \nu'' l''} = \delta_{\nu\nu''}\frac{1}{2}Z^{\nu}_{ll''}, \tag{27}$$

$$\zeta_l = \{-1, 1, 1, 1\}_l, \quad l, l', l'' = 0, 1, 2, 3,$$

$$K(l) = -\{U_2, J, J, J\}_l, \quad U_2 = 2U - 5J,$$

$$Z^{\nu}_{ll'}(q\omega) = -\frac{1}{2}\sum_{k, \lambda, \lambda'} \frac{f_{\lambda, k + q} - f_{\lambda' k}}{E_{\lambda, k + q} - E_{\lambda' k} + \omega + i\eta} \times F^{l}_{\lambda\lambda'}(k)(F^{l'}_{\lambda\lambda'}(k))^{*}, \tag{28}$$

$$F^{l}_{\lambda\lambda'}(k) = \sum_{i, j} S_{i\nu\lambda}(k + q)(\sigma_l)_{ij} S_{j\nu\lambda'}(k). \tag{29}$$

Here, E_{λ} and f_{λ} denote the band energies and Fermi functions of the spiral state. Compact expressions for functions (29) are given in Appendix II. In the absence of interactions ($U = U_1 = J = 0$), the functions $Z^{\nu}_{ll'}(q\omega)$ are nothing but Lindhard zero susceptibilities $\chi^{\nu\nu}_0$ for each one of the bands. As the case should be for a system of noninteracting bands ($t_{\alpha\beta} = 0$), the susceptibility $\chi^{\nu\nu}_0$ is diagonal over the band index, unlike that used in [17]. It is important that the susceptibility-testing interaction with the magnetic field or with neutrons is the sum of contributions from each band and contains no terms that are not diagonal with respect to ν. The diagonalizability of the functions $Z^{\nu}_{ll'} = Z^{\nu\nu}_{ll'}$ with respect to ν is retained when the interactions of U, U_1, and J are included in the mean-field approximation.

We compute $Z^{\nu}_{ll'}(q'\omega)$ from Eqs. (28) and (29) and solve the set of algebraic equations (27) to find the correlators $G_{\nu l, \nu' l'}(q'\omega)$ for each one of three values of the argument $q' = q - Q$, q, and $q + Q$. We recall definitions (20), (22), and (25) to derive expressions, for the sought spin susceptibilities given by Eq. (18). These expressions are similar to the results of [20] for a one-band model,

$$\chi_{zz} = G_{3, 3}(q\omega), \tag{30}$$

$$\chi_{xx} = \chi_{yy} = \frac{1}{4}\{[G_{11} + G_{22} - i(G_{12} - G_{21})]\big|_{q + Q} + [G_{11} + G_{22} + i(G_{12} - G_{21})]\big|_{q - Q}\}. \tag{31}$$

Here, the omitted subscripts ν in $G_{l, l'}|_{q'} \equiv G_{\nu l, \nu' l'}(q'\omega)$ are implied. It is demonstrated in Appendix II that, with $\omega \longrightarrow 0$ and $q' \longrightarrow 0$, the functions $G_{\nu l, \nu' l'}(q'\omega)$ have a pole corresponding to the singularities in susceptibility at $q = 0, \pm Q$. Note further that, even with nonmixed zones ($t_{\alpha\beta} = 0$), the elements $\chi^{\nu\nu'}$ nondiagonal with respect to ν are nonzero because of the exchange interaction of spins in the Hamiltonian given by Eq. (1).

Figure 6 gives the real and imaginary parts of total susceptibility $\bar{\chi}(q\omega) = (\chi_{xx} + \chi_{yy} + \chi_{zz})/3$ for q varying along the contour Γ–$Y(\pi, \pi)$–$M(0, \pi)$–Γ–$Y_2(\pi, -\pi)$. The calculation was performed for a model with the parameters given in Eq. (3) for $t_{\alpha\beta} = 0$ and finite parameter of broadening $\eta = 0.02$ eV in Eq. (28). In addition to the

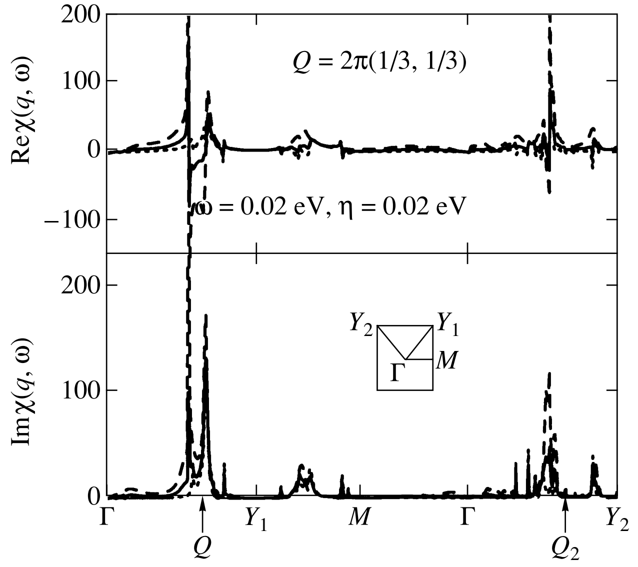


Fig. 6. The real and imaginary parts of total spin susceptibility $\chi(q, \omega)$ averaged over polarizations (solid curve) and its individual components, namely, χ_{zz} (dashed curves) and χ_{xx} (dotted curve), for q varying along the contour $\Gamma–Y_1–M–\Gamma–Y_2$ (see the inset). The frequency and parameter of broadening are $\omega = \eta = 0.02$ eV.

expected absorption peaks in $\chi''(q\omega)$ at $q = Q = 2\pi(1/3, 1/3)$, an almost symmetric peak is observed at $q = 2\pi(1/3, –1/3)$. Both peaks correspond to the observed incommensurate peaks in inelastic neutron scattering [12]. Figure 7 details the contributions to $\chi''(q, \omega)$ from different bands and different polarizations. Note that the integrated intensity $\tilde{\chi}(\omega) = \int\chi''(q\omega)\,d^2q/4\pi^2$ is of the order of 1 eV^{-1} which is comparable in order of magnitude with the similar value of about 2 eV^{-1} in cuprates [33]. Attempts at calculating the static uniform susceptibility ($\omega \longrightarrow 0$, $q \longrightarrow 0$) and comparing the results with those observed in the case of NMR on ^{17}O meet with the difficulty that the order of calculated quantities $\chi'(q, \omega = 0)$ varies abruptly at small values of q and ω. Under these conditions, the effect of spin fluctuations on the NMR characteristics must depend considerably on the nonuniformities and processes of scattering.

5. CONCLUSIONS

The results of calculation of the states of a three-band model with different magnetic structure by the mean-field method have shown the spiral state with the vector $Q = 2\pi(1/3, 1/3)$ corresponding to the nesting of

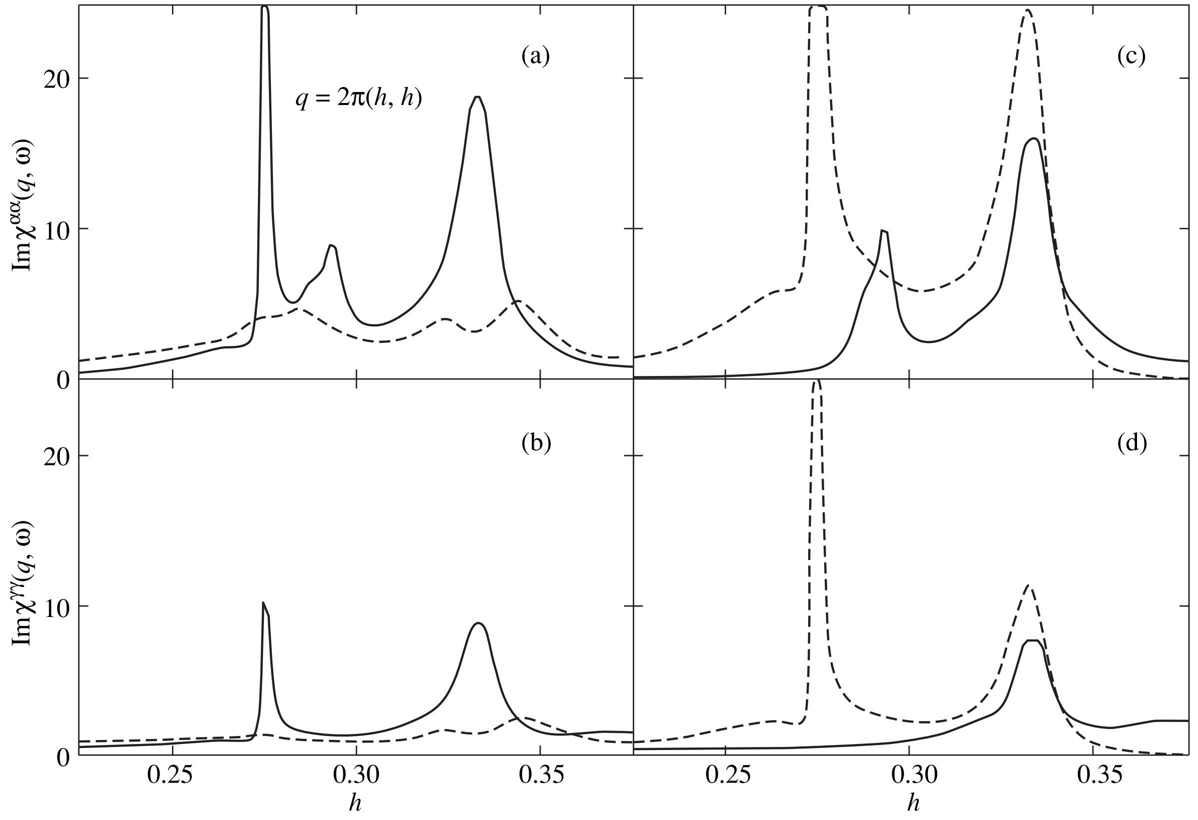


Fig. 7. The contributions (diagonal over ν) to the mean susceptibility from individual bands $\nu = \alpha$ and $\nu = \gamma$ (a, b) for the frequencies $\omega = 0.02$ eV (solid curves) and 0.06 eV (dashed curve) and $\eta = 0.02$ eV for $Q = 2\pi(h, h)$ in the vicinity of resonance $q \sim Q$. The plots (c) and (d) give the same for $\chi_{zz}^{\nu\nu}$ (dashed curve) and $\chi_{xx}^{\nu\nu}$ (solid curve) for $\omega = 0.02$ eV.

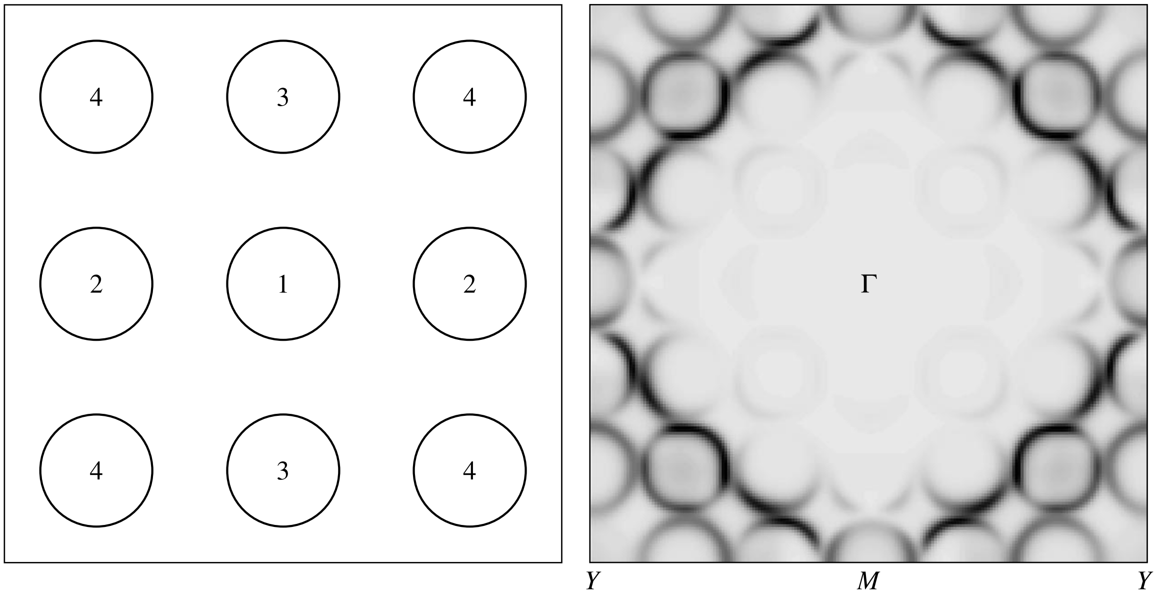


Fig. 8. On the left, atoms of four sorts, $l = 1, 2, 3, 4$, in a 3×3 block of a periodic block-antiferromagnetic structure. Local spins are collinear and of opposite signs for the centers of the neighboring blocks, $\zeta = -1$ in Eq. (33). On the right, a Fermi surface structure obtained as the image of the spectral function $A(k, \omega = 0)$ (see the text).

the α and β bands to be lower in energy among the studied structures (ferromagnetic, antiferromagnetic, spiral, periodic). The AFM structures in photoemission spectra must exhibit characteristic Fermi boundaries around electron pockets for the γ band and shadow Fermi boundaries of the α and β bands, which are the reflection of the main Fermi surfaces relative to the lines $k_x \pm k_y = \pm\pi$. Such Fermi surfaces were indeed observed in Sr_2RuO_4 [26]. However, their origin may be associated both with the bulk magnetic doubling of a unit cell of the AFM structure and with the magnetic or lattice rearrangement of the cell of surface layers.

For the spiral state with the vector $Q = 2\pi(1/3, 1/3)$, the main sheets of the Fermi surface for the α, β, and γ bands show up in photoemission with the intensity depending on the spin polarization. The search for such polarization asymmetry of photoemission spectra could serve as a test for the presence of spiral spin structures in Sr_2RuO_4. Shadow Fermi boundaries of the α and β bands along the $\Gamma(0, 0)$–$M(\pi, 0)$ (or $(0, \pm\pi)$) lines were revealed. They coincide with the main Fermi boundaries shifted by the vector Q. The position of sharp SS peaks along the Γ–M lines, observed in photoemission (ARPES) experiments, coincides with the position of the shadow Fermi surfaces of the α and β bands and with a segment of the boundary of the electron pocket of the γ band, formed by the spiral spin structure. This suggests that the SS features are due to the umklapp processes in spiral (or periodic) structures rather than to the surface states, as interpreted previously [26]. In this case, the disappearance of SS features as well as other shadow Fermi boundaries upon a cyclic variation of temperature [26] may be attributed to the destruction of spiral (instead of surface) structures. The results of calculations of spin susceptibility by the RPA method, in which the spiral spin structure is taken into account even in the zeroth approximation, conform the presence of peaks in $\chi''(q, \omega)$ at $Q = 2\pi(1/3, 1/3)$ in accordance with the magnetic peak observed in inelastic neutron scattering [9]. The arguments advanced above are indicative of the possible presence of spiral spin structures in Sr_2RuO_4 at low temperatures. This hypothesis poses new questions, for example, that of the effect of the spiral structure on the symmetry and properties of the superconducting order in rutinate or on the phase of magnetic quantum oscillations.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 00-03-32981, 00-15-97334). We are grateful to V.Ya. Krivnov for stimulating discussions of the problem and P. Fülde for the opportunity to work at the Max-Planck Institute for Physics of Complex Systems.

APPENDIX I

MEAN-FIELD EQUATIONS FOR A PERIODIC STRUCTURE

We will characterize the structure by the basis vectors $\mathbf{R}_{1(2)}$ of the Bravais cell and by the respective vectors of reciprocal lattice $\mathbf{Q}_{1(2)}$,

$$\begin{aligned} \mathbf{R}_{1(2)} &= (3a, \pm 3a), \\ \mathbf{Q}_{1(2)} &= (2\pi/6, \pm 2\pi/6). \end{aligned} \tag{32}$$

Table 2. The values of populations and local spins for atoms of four sorts in a 3 × 3 block of a periodic block-antiferromagnetic structure with a unit cell consisting of two blocks

l	1	2(3)	3(2)	4
$r_{n,\alpha(\beta)}$	0.808	0.596	0.784	0.608
$r_{n,\gamma}$	0.499	0.650	0.650	0.725
$d_{n,\alpha(\beta)}$	0.170	0.338	0.198	0.331
$d_{n,\gamma}$	0.118	0.202	0.202	0.164

The electron densities and local spins at the sites $n = i + (3L_x, 3L_y)$ serve as the order parameters of this structure,

$$\frac{1}{2}\langle n_{n\nu}\rangle = r_{i\nu}, \quad \langle S_{zn\nu}\rangle = \zeta^{L_x+L_y} d_{i\nu}. \tag{33}$$

These parameters depend on the position of the site $i = (i_1, i_2)$ within a 3 × 3 lattice block. The signs of equivalent spins of neighboring units are equal ($\zeta = +1$) or opposite ($\zeta = -1$) for block-ferromagnetic or block-antiferromagnetic periodic structures. With the symmetry of C_{4v}, we have lattice sites of four sorts, $l = 1, \ldots, 4$, according to Fig. 8.

Previously, in searching for paramagnetic solutions, the quasi-momentum k in the complete basis $\{c^{\dagger}_{\nu\sigma k}\}$ varied in the Brillouin zone G of the initial lattice ($|k_{x(y)}| < \pi$). The Brillouin zone $\kappa \in G_{18}$ of the new structure is confined within the $|\kappa_x \pm \kappa_y| \le 2\pi/3$ boundaries, and the complete set of operators is conveniently redesignated as

$$\begin{gathered}\{a^{\dagger}_{\nu\sigma m\kappa}\} = \{c^{\dagger}_{\nu\sigma k(m)}\},\\ m = (m_1, m_2), \quad k(m) = \kappa + Q_1 m_1 + Q_2 m_2.\end{gathered} \tag{34}$$

Here, κ is the quasi-momentum reduced to the new Brillouin zone, and $m = (m_1, m_2)$ is a pair of integers entering expressions for the shift $[k(m) - \kappa]$ in terms of reciprocal lattice vectors. We have 18 independent pairs of integers (m_1, m_2). They may be selected in different ways.

The mean energy of state is

$$\overline{H} = \sum_n \overline{H}_U(r_{n\nu}, d_{n\nu}) + \overline{T}. \tag{35}$$

The expression for $\overline{H}_U(r_{n\nu}, d_{n\nu})$ in terms of densities and local spins is given by the same formula (9) as before. The only difference is that the quantities $r_{n\nu}$ and $d_{n\nu}$ now depend on the site position in the cell according to Eqs. (33).

The linearized Hamiltonian obtained using Eqs. (34) and (35) has the form

$$\begin{gathered}H_{\mathrm{lin}} = \sum_{\kappa} h_{\kappa},\\ h_{\kappa} = \sum_{\nu,\sigma,m,m'} [\epsilon_{\nu}(k(m))\delta_{mm'} + V_{\nu\sigma}(m - m')] a^{\dagger}_{\sigma\nu m\kappa} a_{\sigma\nu m'\kappa},\end{gathered} \tag{36}$$

$$V_{\nu\sigma}(m) = U r_{\nu}(m) + \sum_{\nu' \ne \nu}\left\{U_2 r_{\nu'}(m) + \frac{\sigma}{|\sigma|} J d_{\nu'}(m)\right\}. \tag{37}$$

Here, $U_2 = 2U - 5J$; $k(m)$ was defined in Eqs. (34) above; the vector κ is the reduced quasi-momentum ($\kappa \in G_{18}$); $m = (m_1, m_2)$ and $m' = (m'_1, m'_2)$; and $r_{\nu}(\kappa, m)$ and $d_{\nu}(\kappa, m)$ are the Fourier harmonics of densities and spins, respectively,

$$\begin{aligned} r_{\nu}(m) &= \frac{1}{N}\sum_n e^{i(Q_1 m_1 + Q_2 m_2)n}\left\langle \frac{1}{2} n_{n\nu}\right\rangle,\\ d_{\nu}(m) &= \frac{1}{N}\sum_n e^{i(Q_1 m_1 + Q_2 m_2)n}\langle S_{zn\nu}\rangle.\end{aligned} \tag{38}$$

The diagonalization of h_{κ} defines the eigenenergies $E_{\kappa\lambda}$, $\lambda = 1, 2, \ldots, 18$, and the one-electron proper operators

$$b^{\dagger}_{\lambda\kappa} = a^{\dagger}_{\nu\sigma m\kappa} S_{\nu\sigma m\lambda}(\kappa)$$

and, in turn, enables one to calculate the mean values of $r_{n\nu}$ and $d_{n\nu}$ and their harmonics given by Eqs. (38) in terms of the matrix of the coefficients $S_{\nu\sigma m\lambda}(\kappa)$ and Fermi functions $f(E_{\kappa\lambda} - \mu)$ as this was done above for ferromagnetic, antiferromagnetic, and spiral structures. The Fermi boundaries for a self-consistent solution with such a structure were visualized, as before, using the image of the spectral function at $\omega = 0$,

$$A(k, \omega = 0) = \sum_{\sigma\nu\kappa m} |S_{\nu\sigma m\lambda}(\kappa)|^2 \tilde{\delta}(E_{\lambda}(\kappa) - \mu). \tag{39}$$

A self-consistent solution could be found only for the BAFM structure ($\zeta = -1$) in Eqs. (33). The respective densities and spins at sites of four sorts ($l = 1, \ldots, 4$) in a 3 × 3 block (see Fig. 8) are given in Table 2. Of all states treated by us, this BAFM structure is the closest to the lower spiral state (see Table 1).

Also given in Fig. 8 is the Fermi surface image corresponding to the given structure. It is remarkable that it is only the γ band that makes a contribution to the spectral function $A(k, \omega = 0)$. For the α and β bands, the mean field with the given periodicity leads to the emergence of a gap along all previous Fermi boundaries of zero α and β bands of the paramagnetic state.

APPENDIX II

SINGULARITIES OF $\chi(q\omega)$ AT $\omega \longrightarrow 0$

We will demonstrate that $\chi(q, \omega)$ exhibits singularities at $\omega \longrightarrow 0$ and $q \longrightarrow \pm Q$. In view of relation (31), we must prove that the functions $G_{ij}(q, \omega)$, $i, j = 1, 2$, have a singularity at $\omega \longrightarrow 0$ and $q \longrightarrow 0$, i.e., that the determinant of the set of algebraic equations (27) goes to zero at $q = 0$ and $\omega = 0$,

$$\|A\| = \left\|\delta_{ll'}\delta_{\nu\nu'} - Z^{\nu}_{ll'}(0, 0)[U\zeta_{l'}\delta_{\nu\nu'} + K(l')(1-\delta_{\nu\nu'})]\right\|\Big|_{q = 0,\, \omega = 0} = 0. \tag{40}$$

For this, we will find explicit expressions for the matrix $Z^{\nu}_{ll'}(q\omega)$ at $q = 0$ and $\omega \longrightarrow 0$. Before doing so, note that, at $t_{\alpha\beta} = 0$, a certain band ν corresponds to each pair of one-electron states $b^{\dagger}_{\lambda k}$ ($\lambda = 1, \ldots, 6$) of linearized Hamiltonian h_k, so that the subscript λ may be represented in the form $\lambda = \tilde{\lambda} + 2(\nu - 1)$, $\lambda = 1, 2$ and $\nu = 1, 2, 3$. Accordingly, the matrix of the coefficients in Eq. (12) is factorized into three independent second-order matrices $S^{\nu}_{i\nu,\,\tilde{\lambda}\nu}$. The functions

$$F^{l}_{\lambda\lambda'} = \tilde{F}^{l\nu}_{\tilde{\lambda}\tilde{\lambda}'}, \quad \tilde{\lambda}, \tilde{\lambda}' = 1, 2 \tag{41}$$

are factorized in just the same way, where $\lambda = \tilde{\lambda} + 2(\nu - 1)$ and is similar to λ'. As a result, for the second-order matrices $\tilde{F}^{l\nu}_{\tilde{\lambda}\tilde{\lambda}'}$, we derive the expressions

$$\begin{aligned}
\tilde{F}^{0\nu}_{\tilde{\lambda}\tilde{\lambda}'} &= [\sigma_0 c^{\nu}_{-} - i\sigma_2 s^{\nu}_{-}]_{\tilde{\lambda}\tilde{\lambda}'},\\
\tilde{F}^{1\nu}_{\tilde{\lambda}\tilde{\lambda}'} &= [-\sigma_1 c^{\nu}_{+} - \sigma_3 s^{\nu}_{+}]_{\tilde{\lambda}\tilde{\lambda}'},\\
\tilde{F}^{2\nu}_{\tilde{\lambda}\tilde{\lambda}'} &= [-\sigma_2 c^{\nu}_{-} + i\sigma_0 s^{\nu}_{-}]_{\tilde{\lambda}\tilde{\lambda}'},\\
\tilde{F}^{3\nu}_{\tilde{\lambda}\tilde{\lambda}'} &= [\sigma_1 s^{\nu}_{+} + \sigma_3 c^{\nu}_{+}]_{\tilde{\lambda}\tilde{\lambda}'},\\
c^{\nu}_{\pm} = \cos(\varphi^{\nu}_{k+q} &\pm \varphi^{\nu}_{k}), \quad s^{\nu}_{\pm} = \sin(\varphi^{\nu}_{k+q} \pm \varphi^{\nu}_{k}).
\end{aligned} \tag{42}$$

Here, σ_l denotes the Pauli matrices, and the quantities φ^{ν}_{k} are defined by the equations

$$\cos 2\varphi^{\nu}_{k} = -\frac{D^{\nu}_{k}}{g^{\nu}_{k}}, \quad \sin 2\varphi^{\nu}_{k} = -\frac{\delta\epsilon^{\nu}_{k}}{g^{\nu}_{k}}, \tag{43}$$

$$\begin{gathered}
D^{\nu}_{k} = \frac{1}{2}\frac{\partial \bar{H}}{\partial d_{\nu}}, \quad \delta\epsilon^{\nu}_{k} = \frac{1}{2}(\epsilon^{\nu}_{k} - \epsilon^{\nu}_{k+Q}),\\
g^{\nu}_{k} = \sqrt{(D^{\nu}_{k})^2 + (\delta\epsilon^{\nu}_{k})^2}.
\end{gathered} \tag{44}$$

All quantities in Eqs. (42) depend implicitly on the vector Q of helicity of the ground state. At $q = 0$, we have $s_{-} = 0$, $c_{-} = 1$, $s_{+} = \sin 2\varphi_k$, and $c_{+} = \cos 2\varphi_k$. According to Eqs. (28), the quantities $Z^{\nu}_{ll'}(q = 0, \omega)$ at $q = 0$ are determined only by contributions with $\lambda \neq \lambda'$. However, for such contributions at $q = 0$, one can prove that

$$\sum_{k} F^{\nu l}_{12}(F^{\nu l'}_{12})^{*}\Phi(k) \sim \delta_{ll'} \tag{45}$$

for any function $\Phi(k)$ exhibiting the symmetry $\Phi(k) = \Phi(-k - Q)$. This follows from expressions (42) and (43), from the values of $s_{\pm}$ and $c_{\pm}$ at $q = 0$, and from the parity of the functions $\cos 2\varphi_k$ (even) and $\sin 2\varphi_k$ (odd) in Eqs. (43) with respect to the operation $k \longrightarrow -(q + Q)$. It is easy to check that $\delta\epsilon_k = -\delta\epsilon_{-k-Q}$. As a result, with $q = 0$ and $\omega = 0$, we have

$$\begin{gathered}
Z^{\nu}_{ll'}(0, 0) = \delta_{ll'} Z_{ll}(0, 0),\\
Z^{\nu}_{22} = \frac{1}{2}\sum_{k}\frac{f^{\nu}_{1k} - f^{\nu}_{2k}}{g^{\nu}_{k}},
\end{gathered} \tag{46}$$

where f_{ik} denotes the Fermi functions for the levels E_{1k} and E_{2k}, and $g_k = (E_{1k} - E_{2k})/2 > 0$ is defined by expression (44).

Therefore, at $q = 0$ and $\omega = 0$, determinant (40) of the twelfth-order matrix A being factorized contains as a factor the determinant of the third-order matrix corresponding to the subscripts $l = l' = 2$ in Eq. (40). As a result, we have

$$\begin{gathered}
\|A\|_{q = 0,\, \omega = 0}\\
\sim \left\|\delta_{\nu\nu'} - Z^{\nu}_{22}(0, 0)[U\delta_{\nu\nu'} + J(1 - \delta_{\nu\nu'})]\right\|.
\end{gathered} \tag{47}$$

We will now make sure that the determinant in the right-hand side of Eq. (47) is zero. For this purpose, we will write the equations satisfied by the values of spin density d_{ν} of a self-consistent solution by the mean-field method with the spin structure. We will take into account the fact that, in the absence of direct interband coupling ($t_{\alpha\beta} = 0$), the matrix of the coefficients $S_{i\nu,\,\lambda}$ in the definition of eigen-operators (7) is factorized into three submatrices and expressed in terms of the quantities φ^{ν}_{k} defined in Eqs. (43) and (44). As a result, the equations for d_{ν} will be written as

$$d_{\nu} = \sum_{\nu'} Z^{\nu}_{22}(0, 0)[U\delta_{\nu\nu'} + J(1 - \delta_{\nu\nu'})]d_{\nu'} \tag{48}$$

or

$$\{\delta_{\nu\nu'} - Z_{22}^{\nu}(0,0)[U\delta_{\nu\nu'} + J(1-\delta_{\nu\nu'})]\}d_{\nu'} = 0, \quad (49)$$

where $Z_{22}^{\nu}(0,0)$ is defined by formula (46). From the fact that the values of d_ν in the solution by the mean-field method are nonzero follows the equality to zero of the respective determinant in the set of equations (49), which coincides with the determinant in the right-hand side of Eq. (47). So, $\det\|A\|_{q=0,\,\omega=0} = 0$, and the function $G_{22}(q\omega)$ exhibits a singularity at $q = 0$ and $\omega = 0$. In view of relation (31) of the susceptibilities $\chi_{xx(yy)}(q\omega)$ with $G_{22}(q \pm Q, \omega)$, we obtain proof of the fact that $\chi_{xx(yy)}(q\omega)$ exhibits a singularity at $q = \pm Q$ and $\omega = 0$.

REFERENCES

1. Y. Maeno, T. M. Rice, and M. Sigrist, Phys. Today **54**, 42 (2001).
2. Y. Maeno, H. Hashimoto, K. Yoshita, *et al.*, Nature **372**, 532 (1994).
3. T. M. Rice and M. Sigrist, J. Phys.: Condens. Matter **7**, L643 (1995).
4. K. Ishida, H. Mukuda, Y. Kitaoka, *et al.*, Nature **396**, 658 (1998).
5. M. Mukuda, K. Ishida, Y. Kitaoka, *et al.*, J. Phys. Soc. Jpn. **67**, 3945 (1998).
6. I. I. Mazin and D. J. Singh, Phys. Rev. Lett. **79**, 733 (1997).
7. I. I. Mazin and D. J. Singh, Phys. Rev. Lett. **82**, 4324 (1999).
8. A. Liebsch and A. Lichtenstein, Phys. Rev. Lett. **84**, 1591 (2000).
9. M. Braden, G. Andre, S. Nakatsyji, and Y. Maeno, Phys. Rev. B **58**, 847 (1998).
10. O. Friedt, M. Braden, G. Andre, *et al.*, Phys. Rev. B **63**, 174432 (2001).
11. R. S. Perry, L. M. Galvin, S. A. Grigera, *et al.*, Phys. Rev. Lett. **86**, 2661 (2001).
12. Y. Sidis, M. Braden, P. Bourges, *et al.*, Phys. Rev. Lett. **83**, 3320 (1999).
13. M. Braden, O. Friedt, Y. Sidis, *et al.*, cond-mat/0107579.
14. J. J. Neumeier, M. F. Hundley, M. G. Smith, *et al.*, Phys. Rev. B **50**, 17910 (1994).
15. S. Nishizaki, Y. Maeno, Z. Mao, *et al.*, J. Phys. Soc. Jpn. **69**, 572 (2000).
16. I. Mazin and D. J. Singh, Phys. Rev. Lett. **82**, 4324 (1999).
17. I. Eremin, D. Manske, C. Joas, and K. M. Bennemann, cond-mat/0102074.
18. J. R. Schrieffer, X. G. Wen, and F. C. Zhang, Phys. Rev. B **39**, 11663 (1989).
19. A. V. Chubukov and K. A. Mussaelian, Phys. Rev. B **51**, 12605 (1995).
20. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **116**, 1058 (1999) [JETP **89**, 564 (1999)].
21. A. P. Mackenzie, S. R. Julian, A. J. Diver, *et al.*, Phys. Rev. Lett. **76**, 3786 (1996).
22. P. G. Grigoriev, M. V. Kartsovnik, W. Biberacher, *et al.*, cond-mat/0108352.
23. A. Puchkov, Z. X. Shen, T. Kimura, and Y. Tokura, Phys. Rev. B **58**, R13322 (1998).
24. D. H. Lu, M. Schmidt, T. R. Cummins, *et al.*, Phys. Rev. Lett. **76**, 4845 (1996).
25. T. Yokoya, A. Chanani, T. Takahashi, *et al.*, Phys. Rev. B **54**, 13311 (1996).
26. A. Damascelli, D. H. Lu, K. M. Shen, *et al.*, Phys. Rev. Lett. **85**, 5194 (2000).
27. P. K. de Boer and R. A. de Groot, Phys. Rev. B **59**, 9894 (1999).
28. T. Oguchi, Phys. Rev. B **51**, 1385 (1995).
29. R. Matzdorf, Z. Fang, Ismail, *et al.*, Science **289** (5480), 746 (2000).
30. Yu. A. Izyumov, I. M. Katsnel'son, and Yu. N. Skryabin, *Magnetism of Collective Electrons* (Nauka, Moscow, 1994).
31. Y. Izuyama, D.-J. Kim, and R. Kubo, J. Phys. Soc. Jpn. **18**, 1025 (1963).
32. T. Imai, A. W. Hunt, K. R. Thurber, and F. C. Chou, Phys. Rev. Lett. **81**, 3006 (1998).
33. P. Bourges, H. Casalta, A. S. Ivanov, and D. Petitgrand, Phys. Rev. Lett. **79**, 4906 (1997).

Translated by H. Bronstein