Barrier Height and Tunneling Current in Schottky Diodes with Embedded Layers of Quantum Dots

A. I. Yakimov*, A. V. Dvurechenskiĭ, A. I. Nikiforov, and S. V. Chaĭkovskiĭ

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia
*e-mail: yakimov@isp.nsc.ru
Received December 20, 2001

Electrical characteristics of silicon Schottky diodes containing Ge quantum dot (QD) arrays are investigated. It has been found that the potential barrier height at the metal–semiconductor contact can be controlled by introducing dense QD layers, which is a consequence of the formation of a planar electrostatic potential of charged QDs. When the applied voltage is varied, the ideality factors of Schottky barriers exhibit oscillations due to the tunneling of holes through discrete levels in quantum dots. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 85.30.Kk; 73.21.La; 73.40.Gk; 73.20.Mf

The last decade has been marked by great progress in nanoelectronics. Its advances are largely associated with the introduction of nanostructures with quantum dots (QDs). QDs represent the limiting case of systems with a reduced dimensionality, because the motion of charge carriers in these systems is spatially confined to sizes smaller than the electron wavelength in all three dimensions. The dimensionality of the electron states in QDs is considered equal to zero, and quantum dots in this sense are artificial analogues of atoms [1]. The discrete energy spectrum of electron states localized in QDs serves as the characteristic feature of zero-dimensional systems that determines the particularity of physical phenomena in nanostructures with QDs [2]. A particular region of physical phenomena associated with the discreteness of the charge transfer by one electron and called single-electron phenomena is typical of electron transport processes in structures with QDs [3]. The attractive fact is that the characteristics of single-electron devices are universal in the sense that these are determined by only the mutual QD–drain, QD–source, and QD–gate capacitances and do not depend on the particular implementation of the diode or transistor.

Because of small sizes (~10 nm) and the high uniformity of their sizes and shape, self-organized QDs that form in the heteroepitaxy of elastically strained systems are most attractive from the practical point of view [4, 5]. Successful attempts have already been made at developing efficient heterolasers [4], photodetectors in the IR region [6], tunnel diodes [7], quantum transistors [8], and single-electron memory elements [9] based on arrays of such QDs. A broad range of fundamental physical problems associated with revealing the mechanisms and regularities of charge transfer in device structures with embedded QD layers arise in this connection.

This work is devoted to studying the potential distribution and electron transport processes in silicon Schottky diodes containing an array of germanium nanoclusters. Ge islands in Si represent potential wells for holes and can be charged with a positive charge, capturing holes from the surrounding volume and thus changing the potential in the vicinity of the Schottky barrier. In addition, the occurrence of discrete energy states in Ge QDs can enhance processes of the tunneling leakage of holes through the barrier. A knowledge of fundamental physical phenomena in such systems allows semiconductor diodes with required electrical characteristics to be developed purposefully.

The aim of this work was to find the regularities of the formation of the potential barrier and the variation of the ideality factor upon introducing QD layers into the region of the metal–semiconductor contact.

Formation of Schottky diodes with QDs. A schematic representation of a structure cross-section is shown in Fig. 1. Samples were grown by molecular beam epitaxy on phosphorus-doped Si(001) substrates with a resistivity of 7.5 Ω cm. The growth temperature of Si layers was 800 and 500°C before and after the deposition of the Ge layer, respectively. After cleaning the substrate, a Si buffer layer 50 nm thick was grown, on which a p^+-Si layer delta-doped with boron was subsequently deposited (the layer concentration of boron was 5×10^{13} cm^{-2}). Next, a p-Si layer was grown with the boron concentration at the level $N_B \sim 5 \times 10^{15}$ cm^{-3} and the thickness $L = 40$ nm. A Ge layer was introduced at the center of this layer at a temperature of 300°C with a varying equivalent thickness d_{eff}. To improve the properties of the resulting metal–semiconductor interface, samples were passed through a lock into another chamber and were held in an O$_2$ atmosphere at a pres-

sure of 10^{-4} Pa and a temperature of 500°C for 15 min. As a result of this procedure, a surface SiO_2 layer formed with a thickness of about 1 nm. Its role was in suppressing the formation of a static dipole layer at the interface, thus decreasing the reverse current of the diodes [10]. The ohmic contact to the buried delta-doped p^+-Si layer was formed by depositing Au followed by heating the structure at a temperature of 400°C for 10 min. The Schottky barrier was created by sputtering a Ti/Al contact on the epitaxial structure. The contact area was $A = 1.5 \times 10^{-4}$ cm^2. The samples were made in the variant of Schottky diodes with a short base in order to decrease the barrier height at the metal–semiconductor contact through the Schottky effect and, hence, to observe experimentally the change in the effective barrier height due to the electrostatic charging of QDs.

Four sets of samples were investigated. The samples of the first set did not contain Ge ($d_{eff} = 0$) and represented conventional silicon Schottky diodes. The equivalent thickness of Ge in the second set of samples comprised 5 monolayers (ML) (1 monolayer = 1.4 Å); in the third, $d_{eff} = 8$ ML; and in the fourth, $d_{eff} = 10$ ML. Under the growth conditions used in this work, a continuous Ge film grows at $d_{eff} \leq 5$ ML, and pyramidal Ge nanoclusters (QDs) faceted by {105} planes appear on the continuous film (wetting layer) at larger thicknesses [5]. For $d_{eff} = 8$ ML, the average size of the pyramid bases equals $a_{QD} = 10$ nm, the pyramid height is $h \sim 1$ nm, and the island density is 4×10^{11} cm^{-2}. For $d_{eff} = 10$ ML, these parameters are $a_{QD} = 15$ nm, $h \sim 1.5$ nm, and 3×10^{11} cm^{-2}, respectively.

Barrier height. One of the main characteristics of a Schottky diode is the potential barrier height at the metal–semiconductor interface. The potential barrier height can be found from an analysis of the temperature dependence of the I_s/T^2 ratio, where I_s is the diode saturation current and T is temperature [10]. In its turn, the saturation current can be found from a linear extrapolation of current–voltage characteristics at $V > 3$ kT/e to $V = 0$ (k is the Boltzmann constant; $e = |e|$ is the absolute value of the electron charge) [10].

Figure 2 shows experimental curves of $I_s/T^2(T^{-1})$ and the barrier height φ_B determined in this way. It was found that $\varphi_B = 0.33$ eV for $d_{eff} = 0$ and 5 ML, $\varphi_B = 0.34$ eV for $d_{eff} = 8$ ML, and φ_B increases up to 0.42 eV in a sample whose equivalent thickness of Ge is equal to 10 ML.

The observed growth of the barrier height can be explained based on the following model. Consider the energy diagram of a metal–p-type silicon contact (Fig. 3). The distribution of the potential due to the formation of a space-charge region (SCR) in Si along the
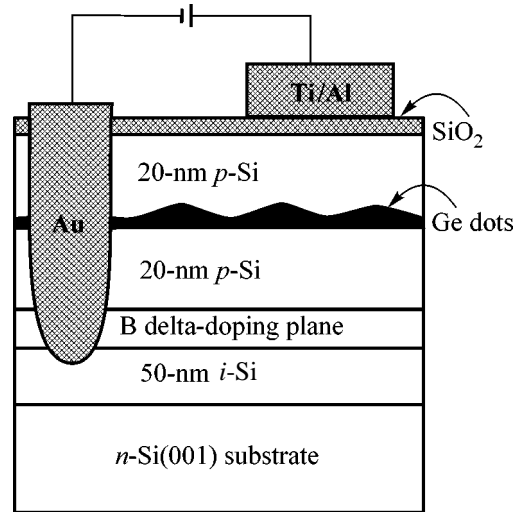
Fig. 1. Schematic representation of a cross-section of a silicon Schottky diode with Ge quantum dots.
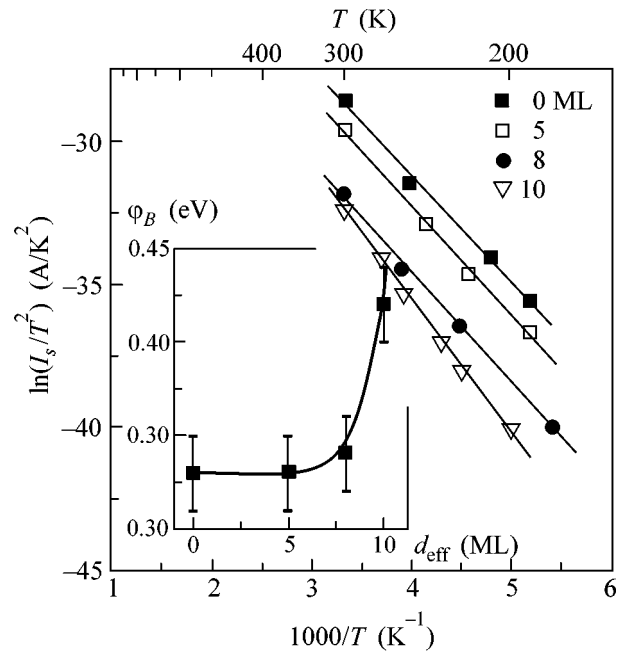

Fig. 2. Temperature dependence of the saturation current on coordinates used for determining the barrier height [10]. The inset shows the barrier height φ_B for various equivalent Ge thicknesses obtained from an analysis of $I_s/T^2(T^{-1})$ curves.

z axis perpendicular to the growth plane is given by the equation

$$\varphi(z) = \varphi_{BS} - \frac{eN_B}{\epsilon\epsilon_0}[w(V)z - z^2/2], \qquad (1)$$

where φ_{BS} is the height of the Schottky barrier, N_B is the impurity concentration, ϵ is the relative dielectric con-
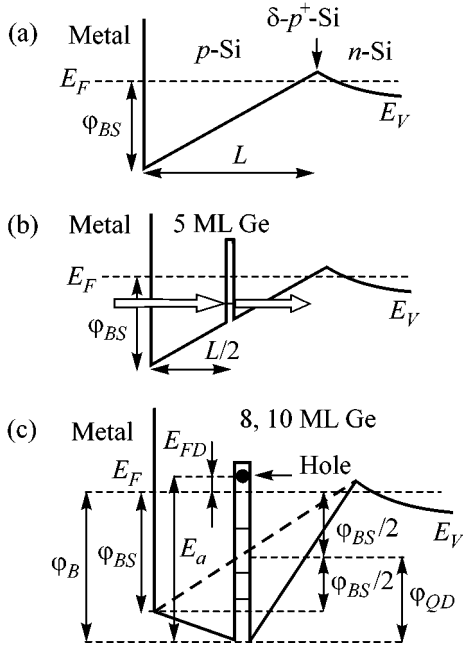
Fig. 3. Equilibrium valence band profile for a metal–*p*-type silicon contact along the growth direction. (a) The Ge layer is absent; (b) the Ge layer is neutral; and (c) the layer of Ge nanoclusters is charged by a positive charge of holes. E_F is the equilibrium Fermi level, φ_{BS} is the Schottky barrier height, φ_B is the effective barrier height in the presence of charged quantum dots, φ_{QD} is the change in the potential due to charged QDs, E_a is the depth of the hole energy level. The potential corresponding to the intermediate SiO_2 layer at the metal–semiconductor interface and to the buffer layer of i-Si is not shown in the figure.

stant of Si, ϵ_0 is the electric constant, and w is the SCR width. The length of the diode base ($L = 40$ nm) is significantly smaller than the value of w required for the formation of a depletion layer in Si with the impurity concentration $N_B = 5 \times 10^{15}$ cm^{-3} ($w(V = 0) \approx \sqrt{(2\epsilon\epsilon_0/eN_B)(\varphi_{BS} - kT)} \approx 300$ nm); therefore, first, the boundary of the depletion layer lies in the vicinity of the delta-doped p^+-Si layer already at a zero bias and hardly shifts at the reverse bias,[1] and, second, $\varphi(z)$ is a nearly linear function (Fig. 3a). The key issue in the understanding of the effect of Ge quantum dots on the electrostatic potential in the system is the possibility of QDs accepting holes from the metal and surrounding silicon. If the thickness of the wetting germanium layer (for samples of the second set) or the sizes of Ge islands (for samples of the third and fourth sets) are so small that the size quantization levels of holes in Ge lie below the Fermi level E_F, then the Ge layer is electrically neu-

tral and does not affect the barrier height (Fig. 3b). As the QD size increases, the hole energy level can rise above E_F and become filled with holes. In this case, the valence band edge in the plane $z = L/2$ will drop by the value φ_{QD}.

It is evident in Fig. 3c that the maximum height of the potential barrier φ_B is

$$\varphi_B = \begin{cases} \varphi_{BS}, & \text{if } \varphi_{QD} \le \varphi_{BS}/2 \\ \varphi_{QD} + \varphi_{BS}/2, & \text{if } \varphi_{QD} > \varphi_{BS}/2. \end{cases} \quad (2)$$

In its turn, $\varphi_{QD} + \varphi_{BS}/2 = E_a - E_{FD}$, where E_a is the energy of the shallowest level of the hole-filled levels in QDs reckoned from the Si valence band edge (QD ionization energy[2]), E_{FD} is the position of the Fermi level with respect to the hole level in QDs (Fig. 3c). For QDs in which the height h is much smaller than its size in the growth plane a_{QD}, $E_{FD} \approx \pi\hbar^2\langle N\rangle/m^* a_{QD}^2$ [11], where $\langle N\rangle$ is the average number of holes in each QD, and m^* is the effective mass of charge carriers. Assuming that $a_{QD} = 15$ nm and $m^* = 0.34m_0$ for heavy holes, we obtain $E_{FD} = 2.6\langle N\rangle$ meV. In QDs of such a small size, the maximum number of holes on size quantization levels ≤ 10 and the "ionization" energy we on the order of hundreds of meV [12]; therefore, $E_a - E_{FD} \approx E_a$ and $\varphi_{QD} + \varphi_{BS}/2 \approx E_a$. In this case, Eqs. (2) can be rewritten in a more demonstrative form:

$$\varphi_B \approx \begin{cases} \varphi_{BS}, & \text{if } E_a \le \varphi_{BS} \\ E_a, & \text{if } E_a > \varphi_{BS}. \end{cases} \quad (3)$$

The energy spectrum of holes in analogous layers of Ge/Si quantum dots was studied previously by photoconductivity spectroscopy [13], field-effect measurements [14], and deep level transient spectroscopy (DLTS) [15]. It was found that the ground-state energy of holes is $E_a = 0.34$ eV for Ge QDs forming at $d_{eff} = 8$ ML and $E_a = 0.40$–0.42 eV for $d_{eff} = 10$ ML. It is evident that both these values are, within good accuracy, equal to the effective barrier heights φ_B determined from an analysis of the temperature dependence of the saturation current in Fig. 2.

Ideality factor. The current–voltage characteristics of metal–semiconductor barriers are often written in the form $I = I_s[\exp(eV/nkT) - 1]$, where n is the ideality factor. At a low doping level and relatively high temperatures, n is close to unity. The deviation of n from unity in Schottky diodes is mainly associated with the occurrence of the tunneling current component [10]; there-

---

[1] Measurements of capacitance–voltage characteristics showed that the barrier capacitance of the diodes at the reverse bias in the voltage range $0 \le V \le 1$ V is actually independent of the applied voltage and equals $\epsilon\epsilon_0 A/L$.

[2] In fact, the term "ionization energy" is appropriate in full measure only for atoms, because the removal of an electron from an atom gives rise to an ion. As a rule, the situation is opposite in QDs: QDs are neutral when they contain no conduction electrons, and QDs acquire an excess charge only when they capture electrons or holes.

fore, an analysis of n provides information on tunneling processes in structures with QDs [16].

The ideality factor in the case of a reverse bias is determined by the equation [17]

$$n(V) = \frac{e}{kT}\partial V/\partial\ln\left[\frac{I\exp(eV/kT)}{\exp(eV/kT)-1}\right]. \qquad (4)$$

Figure 4 displays experimental curves of n versus the reverse bias for various samples. As the equivalent Ge thickness increases, the ideality factor grows, and peaks appear in curves $n(V)$ at $d_{eff} \geq 5$ ML, which points to a resonance character of the tunneling current. Resonance tunneling processes are a characteristic feature of charge carrier transport in double-barrier structures of reduced dimensionality and are due to the quantization of the energy spectrum of electrons or holes in the region confined between the barriers. As the reverse bias increases, the energy levels of holes in the QD layer, in turn, reach a resonance with the quasi-Fermi level in the metal. In this case, the probability of tunneling through the Schottky barrier and, hence, the ideality factor must increase, which is actually observed in our experiments.

A peak in curves $n(V)$ at voltages $V \simeq 1.1$ V is observed for all samples containing a Ge layer; therefore, we associate this peak with the penetration of holes through the energy level of a two-dimensional state in the wetting Ge layer (Fig. 3b), because this layer has the same thickness of 5 ML in all the samples containing Ge. The peaks at lower voltages in samples with $d_{eff} = 8$ and 10 ML are due to the tunneling of holes through discrete levels in QDs lying above the energy level in the wetting Ge layer.

The period of oscillations in curves $n(V)$ is reproduced sufficiently well at various temperatures (Fig. 4b). The average period at $d_{eff} = 10$ ML is $\Delta V \approx 160$ mV. Assuming that the QD layer is introduced exactly in the middle of the diode base and neglecting the band bending due to the potential of the ionized impurity in the diode base, one can estimate the energy gap between the hole levels in Ge nanoclusters at $\Delta E \approx e\Delta V/2 \approx 80$ meV. This value is in a reasonable agreement with the value of the energy gap between size quantization levels of holes in analogous Ge QDs (70–75 meV) determined by IR absorption [18] and resonance tunneling in p^+–i–p^+ structures [2].

In summary, it is important to note that the potential barrier height in structures with QDs can be increased only in the case of sufficiently dense packing of QDs in the layer. Otherwise, the array of charged QDs will not form a uniform planar barrier, which could efficiently control the transport of holes through the structure. In a sense, the phenomenon of a change in the height of the Schottky barrier upon introducing QD layers into the system is close in mechanism to the phenomenon that takes place in the case when surface states exist at the metal–semiconductor interface. However, if the density
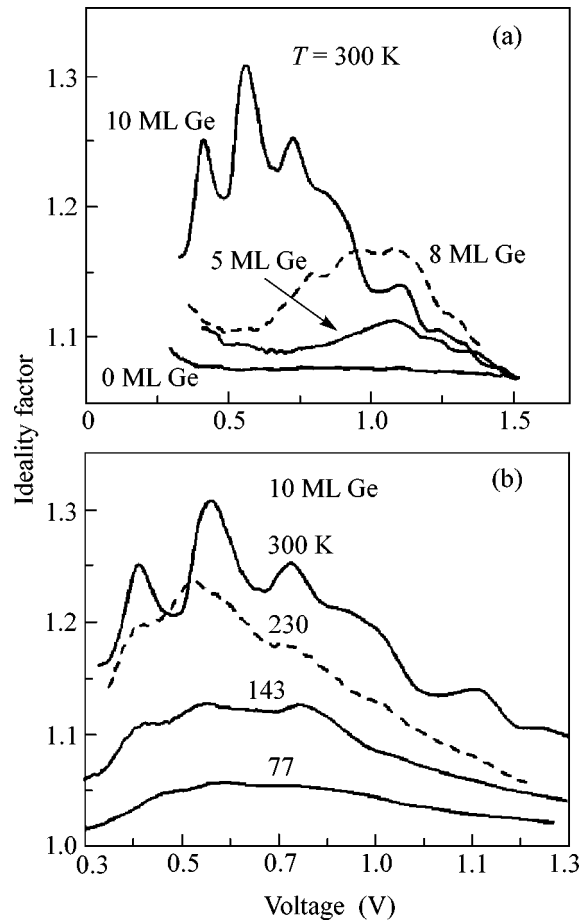

Fig. 4. Dependence of the ideality factor on the reverse bias. (a) Curves for various equivalent thicknesses of the Ge layer d_{eff} ($T = 300$ K). (b) Curves for various temperatures for a sample with $d_{eff} = 10$ ML.

of local levels and their energy spectrum are determined in the last case by the quality of the interface and are not controlled in practice, the parameters of QDs (their density, sizes, and energy spectrum of charge carrier states) are easily controllable given the level of modern technology. This allows effective control of the electrical characteristics of devices. The phenomenon of oscillations of the ideality factor in the case of a reverse bias in Schottky diodes with QDs may serve as a basis for the development of a new method of electron spectroscopy of energy levels in systems with reduced dimensionality.

REFERENCES

1. L. Jacak, P. Hawrylak, and A. Wojs, *Quantum Dots* (Springer-Verlag, Berlin, 1998).

2. A. V. Dvurechenskiǐ and A. I. Yakimov, Usp. Fiz. Nauk **171**, 7 (2001).

3. U. Meirav and E. B. Foxman, Semicond. Sci. Technol. **10**, 255 (1995).

4. N. N. Ledentsov, V. M. Ustinov, V. A. Shchukin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **32**, 385 (1998) [Semiconductors **32**, 343 (1998)].

5. O. P. Pchelyakov, Yu. B. Bolkhovityanov, A. V. Dvurechenskiǐ, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 1281 (2000) [Semiconductors **34**, 1229 (2000)].

6. A. I. Yakimov, A. V. Dvurechenskiǐ, A. I. Nikiforov, and Yu. Proskuryakov, J. Appl. Phys. **89**, 5676 (2001).

7. O. G. Schmidt, U. Denker, K. Eberl, *et al.*, Appl. Phys. Lett. **77**, 4341 (2000); H. W. Li and T. H. Wang, Physica B (Amsterdam) **304**, 107 (2001).

8. J. Phillips, K. Kamath, T. Brock, and P. Bhattacharya, Appl. Phys. Lett. **72**, 3509 (1998); K. H. Schmidt, M. Versen, U. Kunze, *et al.*, Phys. Rev. B **62**, 15 879 (2000).

9. S. Tiwari, F. Rana, H. Hanafi, *et al.*, Appl. Phys. Lett. **68**, 1377 (1996); L. Guo, E. Leobandung, and S. Chou, Appl. Phys. Lett. **70**, 850 (1997); N. Takahashi, H. Ishikuro, and T. Hiramoto, Appl. Phys. Lett. **76**, 209 (2000).

10. S. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981; Mir, Moscow, 1984).

11. V. Ryzhii, I. Khmyrova, V. Pipa, *et al.*, Semicond. Sci. Technol. **16**, 331 (2001); V. Ryzhii, J. Appl. Phys. **89**, 5117 (2001).

12. A. I. Nikiforov, V. A. Cherepanov, O. P. Pchelyakov, *et al.*, Thin Solid Films **380**, 158 (2000).

13. A. I. Yakimov, A. V. Dvurechenskiǐ, Yu. Proskuryakov, *et al.*, Appl. Phys. Lett. **75**, 1413 (1999).

14. A. I. Yakimov, A. V. Dvurechenskiǐ, V. V. Kirienko, *et al.*, Phys. Rev. B **61**, 10 868 (2000).

15. N. P. Stepina, R. Beyer, A. I. Yakimov, *et al.*, submitted to Phys. Low-Dimens. Struct.

16. T. H. Wang, H. W. Li, and J. M. Zhou, Appl. Phys. Lett. **79**, 1537 (2001).

17. S. Averine, Y. C. Chan, and Y. L. Lam, Appl. Phys. Lett. **77**, 274 (2000).

18. A. I. Yakimov, A. V. Dvurechenskiǐ, N. P. Stepina, and A. I. Nikiforov, Phys. Rev. B **62**, 9939 (2000).

Translated by A. Bagatur'yants

A Spin Polaron in a Two-Dimensional Antiferromagnet: From a Local Singlet to a Complex Quasiparticle

A. F. Barabanov, A. V. Mikheenkov*, and A. M. Belemuk

Institute for High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142092 Russia
*e-mail: mikheen@online.ru
Received December 7, 2001

We set forth basic theoretical ideas concerning the spin-polaron scenario for charge excitations in a two-dimensional antiferromagnet. A distinctive feature of the approach being developed consists in considering a local polaron (rather than a bare hole) as a zero approximation for the quasiparticle. At the next step, this excitation is dressed in antiferromagnetic spin waves to form a polaron of intermediate (or infinite) radius. The method allows us to continuously describe the transition from zero to finite temperatures and to consider a wide doping range. Our approach accounts for the main results of ARPES experiments in a CuO_2 plane.© *2002 MAIK "Nauka/Interperiodica".*

PACS numbers: 75.10.Jm; 75.30.Mb; 71.38.-k; 74.72.-h

1. INTRODUCTION

Many strongly correlated two-dimensional (2D) systems that exhibit a non-Fermi liquid behavior of the carriers are known to exist. Here, the principal difficulty of the theory lies in the fact that the operators of low-lying elementary excitations in such systems satisfy neither Bose nor Fermi commutation relations. One of the most glaring examples of such a situation is a doped 2D antiferromagnet (AFM), which reflects the salient features of the CuO_2 plane in high-temperature superconductors (HTSCs). This system has been best studied experimentally. In particular, angle-resolved photoemission spectroscopy (ARPES) shows a number of unusual properties; in particular, it allows the single-particle spectral function of the carries to be measured.

One approach to describing elementary excitations in a 2D AFM involves the concept of a spin magnetic polaron [1], whose consideration is the goal of this review.

The original idea behind this concept is that an elementary excitation in a 2D AFM can be represented as a bare particle (an electron or a hole) surrounded by a deformed spin insulating substrate. The simplest realization of such a quasiparticle, a local spin polaron (LSP), must be given by the solution of the cluster problem. Choosing energetically low states of a small cluster, we can then describe the LSP motion against the AFM background.

The LSP motion strongly depends on the magnetic substrate state and the spin–spin correlation functions. Below, our consideration of the magnetic subsystem is based on the spherically symmetric description of an $S = 1/2$ Heisenberg AFM on a square lattice [2, 3]. Note that the spherically symmetric state at any finite temperature and for sufficient frustration in the spin subsystem is known to be most realistic.

It is intuitively clear that the motion of a small polaron at $T = 0$ must depend on whether (or not) there is a long-range order in the spin subsystem. This implies that the second important step in developing the concept of a spin polaron is allowance for the LSP interaction with spin waves with quasi-momentum $\mathbf{Q} = (\pi, \pi)$; i.e., the need for introducing complex spin polaron (CSP) operators arises. The CSP is a local polaron dressed in spin waves with a quasi-momentum $\mathbf{q}$ close to $\mathbf{Q}$. The structure of the low-lying CSP spectrum is determined by the lower LSP band splitting. As a result, as will be seen below, the effective Fermi surface exhibits a large deviation from the Luttinger theorem. In addition, the following interesting ARPES experimental results are reproduced: a sharp drop in the intensity of the ARPES peaks when the quasi-momentum changes from $(\pi/2, \pi/2)$ to (π, π) or $(0, 0)$, the possible existence of a shadow band (shadow band effect), and the so-called pseudogap on the Fermi surface.

The spin-polaron approach works in any strongly correlated model. In this review, we demonstrate its basic ideas with a Kondo lattice, which is the simplest model used to describe the CuO_2 plane. An effective spin–fermion model with realistic features of the CuO_2 plane is used to describe the experimental results in detail.

2. A LOCAL SPIN POLARON ON A KONDO LATTICE

In this section, we consider a Kondo lattice with a vanishingly small exchange interaction between localized spins [4, 5]. The kinetic energy, which is much

lower than the Kondo exchange ($J \gg t$), is assumed to be the only small parameter of the problem.

A similar situation is observed in the CuO_2 planes in superconducting cuprates. Here, the Kondo exchange corresponds to the production energy of a Zhang–Rice singlet on a single CuO_4 plaquette [6]; it is of the order of 2–3 eV. The singlet kinetic energy t is about 0.5 eV and significantly exceeds the exchange between neighboring spins, which is about 0.15 eV.

Below, using a simple model as an example, we demonstrate the deviation from Fermi statistics (in the spin-polaron approach). A similar effect is also known as the spectral weight transfer [7], where bands with spectral weights that depend on filling and that are less than those for Fermi statistics are observed experimentally. This implies that fewer electrons than in the Fermi case are needed to fill the entire band.

2.1. The Spectral Weights of Bare Carriers

If we completely ignore the electron motion and the exchange interaction between neighboring localized spins, the model Hamiltonian takes the simplest form

$$\hat{H} = \sum_{\mathbf{r}} \hat{H}_{\mathbf{r}},$$

$$\hat{H}_{\mathbf{r}} = 2J\mathbf{S}_{\mathbf{r}} \cdot \mathbf{s}_{\mathbf{r}} \qquad (1)$$

$$= J(S_{\mathbf{r}}^z(n_{\mathbf{r}\uparrow} - n_{\mathbf{r}\downarrow}) + S_{\mathbf{r}}^+ a_{\mathbf{r}\downarrow}^+ a_{\mathbf{r}\uparrow} + S_{\mathbf{r}}^- a_{\mathbf{r}\uparrow}^+ a_{\mathbf{r}\downarrow}),$$

where $\mathbf{S}_{\mathbf{r}}$ is the operator of the localized spin $S = 1/2$ at lattice sites $\mathbf{r}$, $a_{\mathbf{r}\sigma}$ [$\sigma = \uparrow(+), \downarrow(-)$] is the annihilation operator of an electron with spin projection $s_{\mathbf{r}}^z = \sigma/2$, and $n_{\mathbf{r}\sigma} = a_{\mathbf{r}\sigma}^\dagger a_{\mathbf{r}\sigma}$.

The on-site Hamiltonian has eight eigenstates: two states without electrons that correspond to two localized spin directions, two states with two electrons (all these four states have zero energy, $E_0 = E_2 = 0$), and four states with one electron per site: the lower singlet state and the three triplet states with energies $E_{10} = \varepsilon_0 = -3J/2$ and $E_{11} = \varepsilon_1 = J/2$, respectively. For this model, the thermodynamic potential is given by

$$\Omega = -TN_L \ln Z,$$

$$Z = 2 + e^{-\beta(\varepsilon_0 - \mu)} + 3e^{-\beta(\varepsilon_1 - \mu)} + 2e^{2\beta\mu} \qquad (2)$$

(N_L is the number of lattice sites, μ is the chemical potential, and β is the inverse temperature). Hence, we can easily obtain the mean number of particles (determined by μ) and the mean energy for a given lattice site:

$$N = -\frac{1}{N_L}\frac{\partial\Omega}{\partial\mu} = \frac{1}{Z}(e^{-\beta(\varepsilon_0 - \mu)} + 3e^{-\beta(\varepsilon_1 - \mu)} + 4e^{2\beta\mu}),$$

$$\langle\hat{H}_r\rangle = \frac{1}{Z}(\varepsilon_0 e^{-\beta(\varepsilon_0 - \mu)} + 3\varepsilon_1 e^{-\beta(\varepsilon_1 - \mu)}). \qquad (3)$$

At high temperatures ($T \gg J$), the one-electron states are uniformly filled and then $N = 2/(e^{-\beta\mu} + 1)$. At low temperatures ($T \ll J$) and $\mu < 0$ (the energy scale zero is defined as $E_0 = E_2 = 0$), the two-electron states are not occupied. Moreover, since $e^{-\beta\varepsilon_0} \gg e^{-\beta\varepsilon_1}$, the one-electron triplet states are also empty and, in this limit,

$$N = \frac{1}{2e^{\beta(\varepsilon_0 - \mu)} + 1}, \quad \langle\hat{H}_r\rangle = \varepsilon_0 N. \qquad (4)$$

At $\mu > 0$, the two-electron states are filled. However, this case requires no special treatment, because Hamiltonian (1) is symmetric relative to the electron–hole transformation $a_\sigma = \sigma\tilde{a}_{-\sigma}^+$.

For comparison, let us consider free fermions at the level ε_0 with a two-fold spin degeneracy. Instead of Eq. (4), we then have $N^{\text{free}} = 2/(e^{\beta(\varepsilon_0 - \mu)} + 1)$. Therefore, we find from (4) that

$$N \cong (1/2)e^{-\beta(\varepsilon_0 - \mu)} = N^{\text{free}}/4 \qquad (5)$$

for $\mu < \varepsilon_0 - T$ and $\exp(\beta(\varepsilon_0 - \mu)) \gg 1$ and

$$N \cong 1 = N^{\text{free}}/2 \qquad (6)$$

for $\mu > \varepsilon_0 - T$ and $\exp(\beta(\varepsilon_0 - \mu)) \ll 1$.

Clearly, the filling will also differ from the Fermi type when the singlet states are smeared into a singlet zone due to electron hopping.

In order to subsequently generalize the above results to a more realistic situation, with allowance for the kinetic energy, let us now solve problem (1) using two-time Green's functions. Using the operators

$$a = a_{\mathbf{r}\uparrow}, \quad b = S_{\mathbf{r}}^z a_{\mathbf{r}\uparrow} + S_{\mathbf{r}}^- a_{\mathbf{r}\downarrow},$$
$$c = n_{\mathbf{r}\downarrow} a_{\mathbf{r}\uparrow}, \quad d = S_{\mathbf{r}}^z n_{\mathbf{r}\downarrow} a_{\mathbf{r}\uparrow} + S_{\mathbf{r}}^- n_{\mathbf{r}\uparrow} a_{\mathbf{r}\downarrow}, \qquad (7)$$

we can easily derive the closed system of equations of motion:

$$\tilde{\omega}\langle\langle a|a^+\rangle\rangle = 1 + J\langle\langle b|a^+\rangle\rangle, \quad \tilde{\omega} = \omega + \mu,$$

$$(\tilde{\omega} + J)\langle\langle b|a^+\rangle\rangle = \frac{3}{4}J\langle\langle a|a^+\rangle\rangle + 2J\langle\langle d|a^+\rangle\rangle,$$

$$\tilde{\omega}\langle\langle c|a^+\rangle\rangle = c_0 + J\langle\langle d|a^+\rangle\rangle, \qquad (8)$$

$$(\tilde{\omega} - J)\langle\langle d|a^+\rangle\rangle = d_0 + \frac{3}{4}J\langle\langle c|a^+\rangle\rangle.$$

Here, we introduce the means

$$c_0 = \langle\{c, a^+\}\rangle = \langle n_{r\downarrow}\rangle = N/2,$$

$$d_0 = \langle\{d, a^+\}\rangle = -\frac{1}{2J}\langle H_r\rangle = -\langle a^+ b\rangle. \qquad (9)$$

The last correlator can be easily derived from the Green's function $\langle\langle b|a^+\rangle\rangle$. Below, we restrict our analysis to the limit of low temperatures $T \ll J$ and $\mu < 0$. In

this limit, we can guess a self-consistent value of d_0 by taking it from the above thermodynamic analysis, which gives $d_0 = 3N/4$ [see Eq. (4)].

Solving the system of Eqs. (8) yields expressions for the electron Green's function $\langle\langle a|a^+\rangle\rangle$ and the Green's function $\langle\langle b|a^+\rangle\rangle$. Using (9), we obtain

$$\langle\langle a|a^\dagger\rangle\rangle = \frac{1+N}{4}\left(\tilde{\omega}+\frac{3}{2}J\right)^{-1}$$
$$+ \frac{3(1-N)}{4}\left(\tilde{\omega}-\frac{1}{2}J\right)^{-1} + \frac{N}{2}\left(\tilde{\omega}-\frac{3}{2}J\right)^{-1}, \tag{10}$$

$$\langle\langle b|a^\dagger\rangle\rangle = -\frac{3(1+N)}{8}\left(\tilde{\omega}+\frac{3}{2}J\right)^{-1}$$
$$+ \frac{3(1-N)}{8}\left(\tilde{\omega}-\frac{1}{2}J\right)^{-1} + \frac{3N}{4}\left(\tilde{\omega}-\frac{3}{2}J\right)^{-1}. \tag{11}$$

These three poles correspond to the transitions $E_0 \longrightarrow E_{10}$, $E_0 \longrightarrow E_{11}$, and $E_{10} \longrightarrow E_2$. At low temperatures $T \ll J$ and $\mu < 0$, the triplet states are unoccupied, which accounts for the absence of the corresponding pole in (10). The mean number of electrons at site r with the spin $s_{\mathbf{r}}^z = +1/2$ can be determined from (10) using the fluctuation–dissipation theorem:

$$\langle n_{r\uparrow}\rangle = \frac{1+N}{4}(e^{\beta(\varepsilon_0-\mu)}+1)^{-1}$$
$$+ \frac{3(1-N)}{4}(e^{\beta(\varepsilon_1-\mu)}+1)^{-1} + \frac{N}{2}(\varepsilon^{\beta(-\varepsilon_0-\mu)}+1)^{-1}, \tag{12}$$

$$\mu < 0.$$

For $T \ll J$ and $\mu < 0$, only the first term

$$\frac{N}{2} = \frac{1+N}{4}\frac{1}{e^{\beta(\varepsilon_0-\mu)}+1}, \tag{13}$$

must be retained on the right-hand side. Solving the last equation yields Eq. (4).

Thus, we have shown that only the lower pole of the Green's function (10) is significant for $T \ll J$ and $\mu < 0$. In what follows, it is convenient to determine the singlet operator and the corresponding Green's function, which has only one pole corresponding to the transition to a singlet state. Choosing the combination

$$\alpha = \frac{1}{\sqrt{2}}\left(\frac{a}{2}-b-\frac{c}{2}+d\right), \tag{14}$$

we may note that the equations of motion (8) [using (9)] lead to the simple relation

$$\left(\tilde{\omega}+\frac{3}{2}J\right)\langle\langle\alpha|\alpha^+\rangle\rangle = \langle\{\alpha,\alpha^+\}\rangle = \frac{1+N}{2} \tag{15}$$

for the singlet Green's function. Using Eq. (7), we may write the singlet operator (14) in a different form:

$$\alpha = \frac{1}{\sqrt{2}}(Z_r^{\downarrow\downarrow}X_r^{0\uparrow} - Z_r^{\downarrow\uparrow}X_r^{0\downarrow}), \tag{16}$$

where $Z_{\mathbf{r}}^{\sigma\sigma} = 1/2 + \sigma S_{\mathbf{r}}^z$, $Z_{\mathbf{r}}^{\downarrow\uparrow} = S_{\mathbf{r}}^-$, and $X_{\mathbf{r}}^{0\sigma} = a_{\mathbf{r}\sigma}(1 - n_{\mathbf{r}-\sigma})$ are the so-called Hubbard projection operators. Operator (16) is also a projection operator between the states without electrons and the exact local singlet polaron state. Using Eq. (13), it may be noted that the mean value of $\alpha^+\alpha$ gives the mean number N of electrons.

2.2. The Lower Polaron Band

Let us now add the kinetic energy operator to Hamiltonian (1):

$$\hat{T} = -t\sum_{\mathbf{g}\sigma}a_{\mathbf{r}+\mathbf{g},\sigma}^+ a_{\mathbf{r}\sigma}, \tag{17}$$

where g denotes the nearest neighbors and z is their number. In the limit $t \ll J$, $T \ll J$, and $\mu < 0$, we may restrict our analysis to the broadening of the singlet band alone and substitute Eq. (17) with the expression

$$\hat{T} = -t\sum_{\mathbf{g}\sigma}X_{\mathbf{r}+\mathbf{g}}^{\sigma 0}X_{\mathbf{r}}^{0\sigma}. \tag{18}$$

This substitution implies the exclusion of the transitions to (or from) two-electron states. Passing in Eq. (15) to momentum space and adding the contribution from the kinetic energy, we derive the equation

$$\left(\tilde{\omega}+\frac{3}{2}J\right)\langle\langle\alpha_{\mathbf{p}}|\alpha_{\mathbf{p}}^+\rangle\rangle = \frac{1+N}{2} + \langle\langle[\alpha_{\mathbf{p}},\hat{T}]|\alpha_{\mathbf{p}}^+\rangle\rangle. \tag{19}$$

Consider the last term of this equation in terms of the projection technique. See the next sections for detail on this technique. In the case under consideration, however, this merely implies that we project the result of the singlet operator commutation onto the singlet itself to give

$$[\alpha_{\mathbf{p}},\hat{T}] \cong E_{\mathbf{p}}\alpha_{\mathbf{p}}, \quad E_{\mathbf{p}} = \frac{\langle\{[\alpha_{\mathbf{p}},\hat{T}],\alpha_{\mathbf{p}}^+\}\rangle}{\langle\{\alpha_{\mathbf{p}},\alpha_{\mathbf{p}}^+\}\rangle}. \tag{20}$$

The last expression is calculated to the lower order in t; i.e., the correlation functions are uncoupled from different lattice sites. Using the results of the preceding section, we can then easily calculate all means:

$$\langle Z_r^{\uparrow\uparrow}X_r^{00}\rangle = \langle Z_r^{\downarrow\downarrow}X_r^{00}\rangle = (1-N)/2,$$
$$\langle Z_r^{\uparrow\uparrow}X_r^{\downarrow\downarrow}\rangle = \langle Z_r^{\downarrow\downarrow}X_r^{\uparrow\uparrow}\rangle = N/2,$$
$$\langle Z_r^{\uparrow\downarrow}X_r^{\downarrow\uparrow}\rangle = \langle Z_r^{\downarrow\uparrow}X_r^{\uparrow\downarrow}\rangle = -N/2, \tag{21}$$
$$\langle X_r^{\uparrow\uparrow}\rangle = N/2, \quad \langle X_r^{00}\rangle = 1-N.$$

Thus, we obtain

$$E_{\mathbf{p}} = -(tz\gamma_{\mathbf{p}})\frac{1+N}{4}, \quad \gamma_{\mathbf{p}} = \frac{1}{z}\sum_{\mathbf{g}} e^{i\mathbf{p}\mathbf{g}}. \quad (22)$$

An essential feature of our solution is the dependence of the band width on filling. When the number N of electrons changes from 0 to 1, the band width changes from 1/4 to 1/2 of the free fermion band width. We emphasize that the spin–spin correlations at neighboring lattice sites were disregarded ($\langle \mathbf{S_g S_{r+g}} \rangle = 0$) when deriving Eq. (22). This, in particular, implies that we ignored the superexchange between neighboring localized spins through conduction electrons, which is on the order of t^2/J. Apart from the superexchange, there are also other magnetic ordering processes, in particular, the processes that lead to ferromagnetic coupling. An analysis of the competition between the different magnetic ground states is beyond the scope of this section. Therefore, having determined the magnetic ordering temperature, $T_m \sim t^2/J$, we may only assert that result (22) holds for temperatures $T_m \ll T$.

At $t < T$, the singlet band width is insignificant and the problem reduces to that considered in the preceding section. In the temperature range $T_m \ll T \ll t$, the singlet zone filling is similar to but does not match the Fermi distribution. Let us call it a quasi-Fermi distribution. We find from Eqs. (19) and (20) that

$$N_p = \langle \alpha_p^\dagger \alpha_p \rangle = \frac{1+N}{2}\frac{1}{e^{\beta(\varepsilon_p - \mu)} + 1}, \quad (23)$$

$$\varepsilon_p = \varepsilon_0 + E_p.$$

For very small filling, $N \ll 1$, we obtain a distribution from Eq. (23) which is analogues to the Fermi one with a weight of 1/4:

$$N_p = \frac{1}{2}\frac{1}{e^{\beta(\varepsilon_p - \mu)} + 1}, \quad (24)$$

similar to result (5) of the preceding section. This distribution is valid for sufficiently small chemical potentials, $\mu < \min \varepsilon_p + O(tN/N_L)$, $N \ll 1$.

For large chemical potentials, the number $N(\mu)$ of electrons must be determined from Eq. (23), which yields

$$N = N_L^{-1}\sum_p N_p = F/(2 - F),$$

$$F = N_L^{-1}\sum_p (e^{\beta(\varepsilon_p - \mu)} + 1)^{-1}. \quad (25)$$

Using this relation, we can also determine the quasi-Fermi surface (FS) as a region where the quasi-Fermi distribution (23) falls from $(1 + N)/2$ to zero. Because of the weighting factor, which takes on values between 1/4 and 1/2 (depending on N), the quasi-FS for a given

number of electrons is larger than the ordinary FS. We see from Eq. (25) that the number of electrons changes from zero to unity as the chemical potential increases. For $\varepsilon_0 + tz < \mu < |\varepsilon_0| - tz$, the filling $N = 1$ is similar to the situation in the Hubbard model when the chemical potential lies between two Hubbard subbands.

Thus, even for the simple model considered here, a large deviation from the Fermi distribution is observed when the singlet band is filled with electrons. The carrier spectral weight in the band ranges from a mere one fourth to one half of the free-fermion spectral weight. Both the spectral weight and the band width depend on filling (spectral weight transfer [7]). This effect should be observed experimentally, for example, as a decrease of the spectral weight in photoemission experiments. One might also expect the spectral weight to decrease at temperatures lower than the magnetic ordering temperature. However, the dispersion relation (22) may change in this case.

3. A LOCAL POLARON DRESSED IN SPIN WAVES

In this section, we demonstrate the need to make the spin polaron structure more complex to adequately describe the low-energy spectrum of the carriers. First, we will dwell on the zero-temperature limit [4] and then show how we can continuously pass to finite temperatures [8].

3.1. Zero Temperature

Naturally, the Hamiltonian of a Kondo lattice for a realistic model should include not only the Kondo exchange and kinetic energy, as in the preceding section, but also the AFM interaction between neighboring localized spins. This corresponds to adding the term

$$H_2 = \frac{1}{2}I\sum_{rg} S_{r+g}^\alpha S_r^\alpha. \quad (26)$$

Let us first write out the first two equations from an infinite chain of equations for the retarded Green's functions that describe the LSP motion against the AFM spin background. In this section, we analyze only the quasiparticle spectrum. Therefore, to simplify the notation, we write the operators a instead of the Green's functions $\langle\langle a|b^+\rangle\rangle$ and omit the nonhomogeneous terms $\langle[a, b^+]_+\rangle$. The above equations then take the form

$$(\omega - \epsilon_p)a_p = Jb_p,$$

$$b_p = N^{-\frac{1}{2}}\sum_r b_r e^{-ipr}, \quad b_r = \tilde{S}_r a_r, \quad (27)$$

$$\omega \tilde{S}_{r+R} a_r = [\tilde{S}_{r+R} a_r, H_0 + H_1 + H_2]$$

$$= \sum_g t_g \tilde{S}_{r+R} a_{r+g} + J \tilde{S}_{r+R} \tilde{S}_r a_r \qquad (28)$$

$$+ I i e_{\alpha\beta\gamma} \sum_g \sigma_\alpha S^\beta_{r+R+g} S^\gamma_{r+R} a_r.$$

Here, the summation is performed over the sites r of a square lattice and over their nearest neighbors $|g| = 1$. For brevity, we omitted the spin subscripts on the creation ($a^+_{r\sigma}$) and annihilation ($a_{r\sigma}$) operators for Fermi particles (let these be electrons) and designated $S_r = S^\alpha_r \sigma^\alpha$. In what follows, we assume the electron spectrum to be

$$\epsilon_p = -\epsilon_{p+Q} = -2t(\cos p_x + \cos p_y). \qquad (29)$$

Considering the motion of a local magnetic polaron in the limit $N \ll 1$, we can see that terms proportional to the spin–spin correlation functions $C = \langle \mathbf{S}_r \mathbf{S}_{r+g} \rangle$ at neighboring lattice sites, i.e., those determined by the short-range magnetic order, emerge in the kinetic energy. However, it is clear that in the presence of a long-range order, the polaron motion must also depend on the parameter that corresponds to this long-range AFM order. In a magnetic substrate with a long-range order for the spherically symmetric state under consideration, $\langle S^\alpha_r \rangle = 0$, $C^{\alpha\beta}_R = \langle S^\alpha_r S^\beta_{r+R} \rangle = M\delta^{\alpha\beta} e^{iQR}$, $|R| \gg 1$, and $\mathbf{Q} = (\pi, \pi)$; i.e., in this state, the mean $\langle S_Q S_Q \rangle$ is a macroscopic quantity. The latter implies that the coupling of a local polaron with S_Q [$\mathbf{Q} = (\pi, \pi)$] corresponds to a new polaron state of infinite radius. In this case, it is important to take into account the quantum nature of the spin $\mathbf{Q}$ wave, because the transitions between local and depolarized polaron states essentially determine the polaron bands.

It is natural to introduce the following operators that describe the polaron states of an infinite radius:

$$c_r = \tilde{Q}_r a_r, \quad c_p = \tilde{Q}_0 a_{p+Q}, \qquad (30)$$

$$\tilde{Q}_r = N^{-1} \sum_R e^{iQR} \tilde{S}_{r+R} = e^{iQr} \tilde{Q}_0, \qquad (31)$$

$$d_r = \tilde{Q}_r \tilde{S}_r a_r, \quad d_p = \tilde{Q}_0 b_{p+Q}. \qquad (32)$$

The operators c_r and d_r allow for the long-range spin–electron correlations ($|R| \gg 1$). Note that our set of states is closed relative to the Kondo interaction. We can easily see from Eqs. (27), (30), and (32) that

$$(\omega - \epsilon_{p+Q})_{c_p} = J d_p. \qquad (33)$$

We use the standard Zwanzig–Mori projection technique [9] to close the system of equations for the Green's functions. Let us project the corresponding

commutators onto the following orthonormalized set of basis operators:

$$B_{1r} = a_r,$$

$$B_{2r} = f_2^{-1/2} b'_r, \quad b'_r = b_r - c_r, \qquad (34)$$

$$B_{3r} = f_3^{-1/2} c_r,$$

$$B_{4r} = f_4^{-1/2} d'_r, \quad d'_r = d_r - M a_r + M f_2^1 (b_r - c_r),$$

where f_i are the corresponding normalization factors. The coefficients in the projection relations $[B_i, H] = \sum_j a_{ij} B_j$ are defined as

$$a_{ij} = \langle [[B_{ir}, H], B^+_{jr}]_+ \rangle. \qquad (35)$$

Using Eqs. (27), (33), and (35), we derive the following simple system of four equations in momentum representation:

$$\omega B_{i,p} = \sum_j a_{ij}(p) B_{j,p}. \qquad (36)$$

We do not give any matrix elements of the spectral matrix $a_{ij}(p)$ (see [4]), which depend essentially on the parameters of long-range magnetic order M and correlator C.

Thus, in our approximation, the electron motion against the AFM background is described by the quasiparticle that is a coherent superposition of four Fermi fields: the bare electron field a_p; the delocalized polaron field c_p; and the two localized polaron fields b_p and d_p, which are mainly hybridized through the Kondo interaction. System (36) may be considered as the system of Schrödinger equations in which the matrix elements a_{ij} are the amplitudes of the transitions from state j to state i. The presence of nondiagonal matrix elements reflects the quantum nature of the spin S and Q waves. The eigenfunctions and eigenvalues of Eqs. (36) describe the elementary excitations that form the four bands. The band structure depends on the magnetic subsystem state (on M and C) and on the relations between the energy parameters t, J, and I. Below, we use the following typical values for M and C: $M = 0.1$ and $C = -0.335$.

In our model, all spectra depend on the momentum only via ϵ_p. Note that we used the approximation of a low Fermi-particle density, i.e., $\langle a^+_p a_p \rangle \longrightarrow 0$, when calculating the matrix elements (35).

Figure 1a shows four branches of the quasiparticle spectrum $E^{(i)}(p)$, $i = 1$–4, for $J \gg t$ and I (in the limit of strong correlation). Figure 1b shows the residues $Z^{(i)}_p$ of the electron Green's functions that correspond to the poles $E^{(i)}(p)$:

$$\langle\langle a_{p\sigma} | a^+_{p\sigma} \rangle\rangle = \langle\langle B_{1,p} | B^+_{1,p} \rangle\rangle = \sum_{i=1}^4 \frac{Z^{(i)}_p}{\omega - E^{(i)}(p)}. \qquad (37)$$
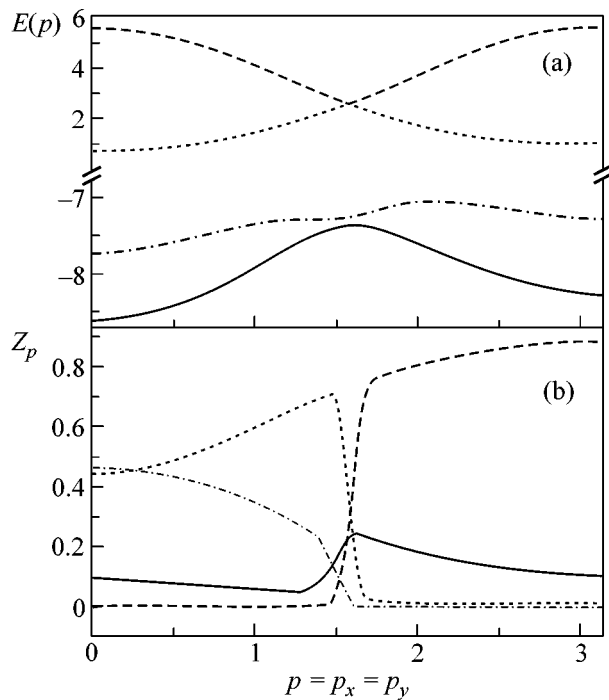
Fig. 1. (a) The quasiparticle spectrum $E^{(i)}(p)$ for $J = 5 \gg t$ and I, $t = 1$, and $I = 0.1$ along the symmetric direction $p = p_x = p_y$; (b) the residues $Z_p^{(i)}$ of the electron Green's function $\langle\langle a_{p\sigma} | a_{p\sigma}^+ \rangle\rangle$ at poles $E^{(i)}(p)$. Different lines correspond to four bands.

In the low-density limit, the residues $Z_p^{(i)}$ are determined by the solution of the equations of motion for the Green's functions $\langle\langle B_{1,p} | B_{1,p}^+ \rangle\rangle$ with the nonhomogeneous terms $W_{i,1} = \langle [B_{i,p}, B_{1,p}^+]_+ \rangle = \delta_{i,1}$. $Z_p^{(i)}$ determines the single-particle spectral function $A(p, \omega) = \sum_i Z_p^{(i)} \delta[\omega - E^{(i)}(p)]$. Since $Z_p^{(i)}$ characterizes the contribution (weight) of the bare particle state a_p to the quasiparticle state with energy $E^{(i)}(p)$, $\sum_i Z_p^{(i)} = 1$.

In Fig. 1a, the lower bands (centered at $E^{(s)} \cong -3J/2$) correspond to the motion of a local polaron in the singlet state $b^{(s)} = b - a/2$ and to the motion of this polaron coupled to a Q wave. The upper bands (centered at $E^{(t)} \cong J/2$) describe the motion of a local polaron in the triplet state $b^{(t)} = b + 3a/2$ and its coupling to a Q wave. The terms "singlet" and "triplet" mean that, when acting on the spin subsystem state, the operators $b_r^{(s)+}$ and $b_r^{(t)+}$ produce, respectively, singlet or triplet spin–electron pairs at site r.

It is important to note that the above four bands are only roughly symmetric about the boundary of the Brillouin magnetic zone. In momentum space, the separa-

tion between two quasiparticle states with the same energy can differ from the AFM vector $\mathbf{Q}$; i.e., the shadow band effect attributable to these bands can differ from the standard effect governed by the vector $\mathbf{Q}$ in a two-sublattice Néel AFM [10].

The lower part of the elementary-excitation spectrum is largely determined by the states b_p and d_p. At the same time, as we see from Fig. 1b (solid curve), the possible filling of each p state with a particle of a given spin for the lower band is not that small, being close to 0.1. Note that, because of such a large deviation of the filling from unity, the FS is large even at small filling.

If the three energy parameters t, I, and J are of the same order of magnitude, then the elementary excitation in any of the four bands is a coherent superposition of the states a, b, c, and d; these states enter into this superposition with comparable weights. We emphasize that disregarding the $\mathbf{Q}$ polarons, i.e., the states c and d, results in the disappearance of two bands and can significantly change the remaining bands.

3.2. The Polaron Structure at Finite Temperatures

In the preceding section, we described the band splitting of a local polaron through its dressing in an AFM spin wave $S_\mathbf{Q}$ at $T = 0$. Our description was based on the fact that the nonzero quantity $\langle \mathbf{S_Q S_Q} \rangle$ is macroscopic. However, it is clear that the spectrum must change continuously with increasing temperature and the above splitting cannot disappear suddenly, although $\langle \mathbf{S_Q S_Q} \rangle$ is no longer a macroscopic quantity at finite temperatures.

We now present the method of describing the spectrum at finite temperatures [8]. In this method, the spin polaron structure is assumed to be more complex: the polaron is a superposition of a local spin polaron, a local spin polaron dressed in a continuum of spin waves $S_\mathbf{q}$ with $\mathbf{q}$ close to $\mathbf{Q}$, and, at $T = 0$, a local spin polaron dressed in $S_\mathbf{Q}$. As previously, we use a Kondo lattice as an example to demonstrate the approach. Here, however, we take into account the hoppings not only to the first nearest neighbors but also to the second and third ones. This model is more realistic, because it gives a minimum for the lower band close to $\mathbf{k} = (\pi/2, \pi/2)$.

Thus, the Hamiltonian takes the form

$$\hat{H} = \hat{J} + \hat{T} + \hat{I}, \quad \hat{J} = J \sum_\mathbf{r} \mathbf{a}_\mathbf{r}^+ \tilde{\mathbf{S}}_\mathbf{r} \mathbf{a}_\mathbf{r},$$

$$\hat{T} = \tau_g \sum_{\mathbf{r},\mathbf{g}} a_{\mathbf{r}+\mathbf{g}}^+ a_\mathbf{r} + \tau_d \sum_{\mathbf{r},\mathbf{d}} a_{\mathbf{r}+\mathbf{d}}^+ a_\mathbf{r} + \tau_{2g} \sum_{\mathbf{r},\mathbf{g}} a_{\mathbf{r}+2\mathbf{g}}^+ a_\mathbf{r}, \quad (38)$$

$$\hat{I} = \frac{I}{2} \sum_{\mathbf{r},\mathbf{g}} \hat{\mathbf{S}}_\mathbf{r} \hat{\mathbf{S}}_{\mathbf{r}+\mathbf{g}}.$$

Here, $\mathbf{g} = \pm\mathbf{g}_x \pm \mathbf{g}_y$ are the vectors of the nearest neighbors; $\mathbf{d}$ and $2\mathbf{g}$ are the vectors of the second and third
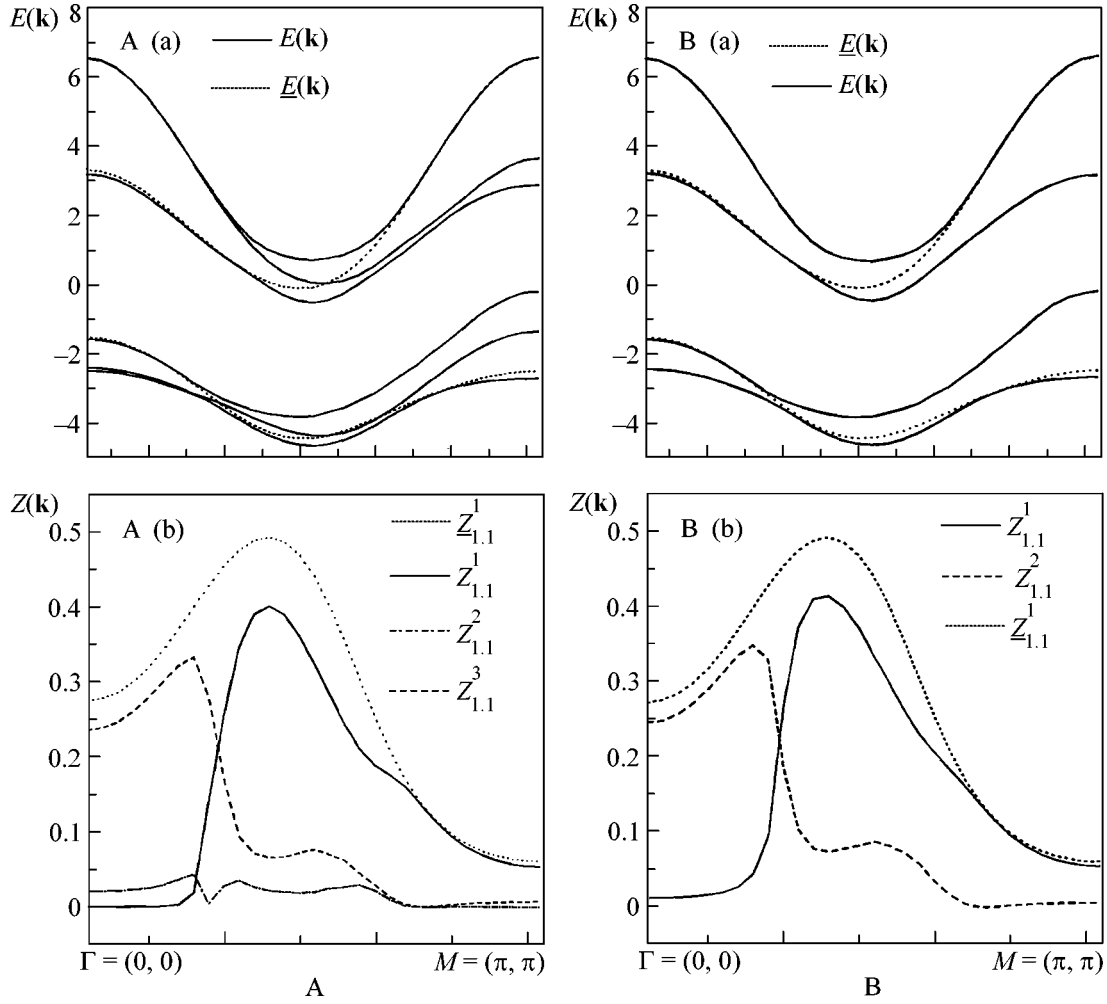
Fig. 2. The spectrum $E^i(k)$ at $T = 0$ along the direction $k = k_x = k_y$. $\underline{E}^{(i)}(k)$ is the same in the local polaron approximation. (b) The residues $Z^{(i)}_{1,1}(k)$ of the Green's function $\langle\langle a_{k\sigma} | a^+_{k\sigma}\rangle\rangle$ at poles $E^{(i)}(k)$. (a) The quasiparticle excitation spectrum $E^{(i)}(k)$ at $T = 0.1I$ along $k = k_x = k_y$. $E^{(i)}(k)$ is the same in the local polaron approximation. (b) The residues $\underline{Z}^{(1)}_{1,1}(k)$ of the Green's function $\langle\langle a_{k\sigma} | a^+_{k\sigma}\rangle\rangle$ at poles $E^{(i)}(k)$. $\underline{Z}^{(1)}_{1,1}(k)$ corresponds to $\underline{E}^{(1)}(k)$ in the local polaron approximation.

nearest neighbors, respectively. The Fermi operator $a^+_{r,g}$ creates a hole with the spin $S = 1/2$ and projection $\sigma/2$ ($\sigma = \pm1$). As previously, the designation $\tilde{S}_r = S^\alpha_r \sigma^\alpha$ is used in the Kondo term $\hat{J}$.

To analyze the excitation spectrum, we introduce the basis of on-site operators

$$A_{r,1} = a_r, \quad A_{r,2} = \tilde{S}_r a_r. \tag{39}$$

To allow for the spin polaron of infinite radius at $T = 0$, we also add the operators

$$A_{r,j} = \tilde{Q}_r A_{r,i}, \quad \tilde{Q}_r = N^{-1}\sum_\rho e^{iQ\rho}\tilde{S}_{r+\rho}, \tag{40}$$

$$j = i+2, \quad i = (1\text{–}2),$$

which describe the local polaron coupling to spin excitations (with the long-range order). Note that the contribution from these operators vanishes at $M = \langle S_Q S_Q\rangle = 0$. Finally, let us introduce the operators

$$A_{r,j} = \tilde{Q}^{(1)}_r A_{r,i}, \quad \tilde{Q}^{(1)}_r = N^{-1}\sum_{\rho,\, q\in\Omega} e^{iq\rho}\tilde{S}_{r+\rho}, \tag{41}$$

$$j = i+4, \quad i = (1\text{–}2).$$

Here, Ω is a small region of the Brillouin zone near the point (π, π) (see Fig. 3). The summation does not include $q = Q$.

The operators $A_{r,5}$ and $A_{r,6}$ describe the polaron of intermediate radius, i.e., the local polaron coupling with the spin waves whose quasimomentum lies in a square region around Q with linear sizes L.
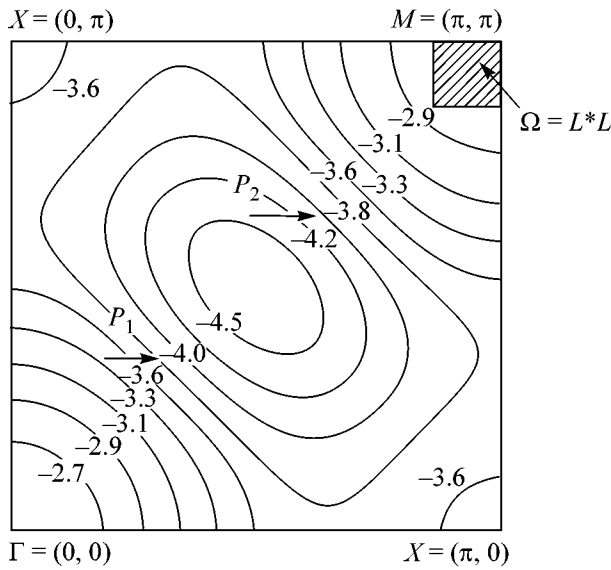
Fig. 3. Surfaces of constant energy for the lower band $E^{(1)}(k)$ at $T = 0.1I$.

The spin-polaron spectrum $\varepsilon_i(\mathbf{k})$ (where i is the band number) can be determined by using the matrix of two-time retarded Green's functions $G_{i,j}(t, \mathbf{k})$ for the operators $A_{\mathbf{k}, \sigma, i}$. We solved the system of equations of motion for $G_{i,j}(t, \mathbf{k})$ in terms of the standard Zwanzig–Mori projection technique, where we restricted ourselves to the above basis of operators $\{A_{\mathbf{k}, \sigma, i}\}$.

Here, we do not give the corresponding matrix elements, because these are cumbersome.

At $T = 0$ (case 1), we use a basis of six operators, $i = 1$–6. At finite temperatures, there are no operators $A_{\mathbf{k}, \sigma, 3}$ and $A_{\mathbf{k}, \sigma, 4}$, because these describe the polaron of infinite radius, which can exist only in the presence of a long-range order.

Thus, the Green's functions take the form

$$G_{i, j}(\omega, \mathbf{k}) = \sum_{l = 1}^{12} \frac{Z_{(i, j)}^{(l)}(\mathbf{k})}{\omega - \varepsilon_l(\mathbf{k})}. \qquad (42)$$

The quantity $Z_{(1, 1)}^{(l)}(\mathbf{k})$ is the number of bare holes with a given spin σ and momentum $\mathbf{k}$ in the state $|\mathbf{k}, \sigma\rangle$ of the quasiparticle band $\varepsilon_l(\mathbf{k})$. Note that the spectral weights (residues) $Z_{(i, j)}(\mathbf{k})$ satisfy the sum rule $\sum_s Z_{(1, 1)}^{(s)}(k) = 1$. The dependence of the spin-polaron excitation spectrum ε_l on the correlation function $M = \langle \mathbf{S_Q S_Q} \rangle$ shows up only when the Q-polaron states $A_{\mathbf{r}, 3}$ and $A_{\mathbf{r}, 4}$ are taken into account. To calculate the spectrum $\varepsilon(\mathbf{k})$ requires knowledge of the spin correlation functions $\langle \mathbf{S}_{R_1} \mathbf{S}_{R_2} \rangle$. The method of their calculation is outlined in [3, 11]. In particular, the following numerical values were obtained for the spin correlation functions at $T = 0$: $\langle \mathbf{S}_R \mathbf{S}_{R + g} \rangle = -0.332$, $\langle \mathbf{S}_R \mathbf{S}_{R + d} \rangle = -0.145 + M$, and

$\langle \mathbf{S}_R \mathbf{S}_{R + 2g} \rangle = -0.144 + M$, where $M = 0.0577$. The following model parameters were used: $\tau = 1$, $\tau_{2g} = 0.5\tau$, $\tau_d = \tau_{2g}$, $\tau_d = -\tau_{2g}$, $J = 5\tau_{2g}$, and $I = 0.5\tau_{2g}$; these were chosen in such a way so that the minimum of the lower zone did not lie far from the point $(\pi/2, \pi/2)$. The parameter L, which specifies the region Ω for the operators $A_{\mathbf{q}, 5(6)}$, was calculated for each $\mathbf{k}$ in the Brillouin zone by the variational method. The size of the region Ω strongly depends on $\mathbf{k}$. For example, $L(0, 0) = 0.05\pi$ but $L(\pi, \pi) = 0.35\pi$. The calculations were performed in the limit of small filling, $n \ll 1$, where n is the total number of carriers per cell.

Case 2 corresponds to the temperature $T = 0.4I$, which corresponds to the correlation length $\xi = 40$ lattice constants.

Our results for the most interesting lower bands $\varepsilon_l(\mathbf{k})$ for $\mathbf{k}$ along the symmetric direction from $\Gamma = (0, 0)$ to $M = (\pi, \pi)$ (in cases 1 and 2) are shown in Figs. 2A and 2B, respectively. Figure 3 shows surfaces of constant energy for the first band. As we see from Figs. 2A(a) and 2B(a), the minimum of the lower band lies near the point $(\pi/2, \pi/2)$ in both cases, which is characteristic of the Emery model and corresponds to the experimentally observed spectrum [12]. We also see from the figures that the cases $T = 0$ and $T \neq 0$ are similar, except that the second band vanishes at nonzero temperature, when there is no long-range order.

In Figs. 2a(b) and 2B(b), the bare-carrier weights $Z_{1, 1}$ are plotted against $\mathbf{k}$ for $T = 0$ and $T \neq 0$, respectively. At zero temperature, the residue of the second band is negligible; otherwise, the figures are identical. Thus, the set of operators required at $T = 0$ may be adequately replaced by the set of operators $A_{1, 2, 5, 6}$ when the long-range order disappears.

We emphasize that the results obtained in this section, as in the case of zero temperature, describe the shadow-band effect and the strong violation of the Lattinger theorem.

Interestingly, as the temperature rises (i.e., as the correlation length ξ decreases), the band shape does not change qualitatively up to the temperatures that correspond to a correlation length of 3–6 lattice constants.

4. A SPIN POLARON FOR A REALISTIC MODEL

The most recent ARPES studies indicate that the spectra of optimally doped cuprates differ from those of insulating cuprates. Experimental data for undoped compounds [12, 13] suggest an isotropic band bottom near the point $(\pi/2, \pi/2)$ in momentum space, a large energy difference between $N = (\pi/2, \pi/2)$ and $X = (\pi, 0)$, and the so-called remnant FS [14] (the surface on which the single-particle spectral weight decreases sharply). A flat-band region, a large FS centered at the point $M = (\pi, \pi)$, and the so-called shadow FS, which results from AFM spin correlations [10], were found in optimally doped cuprates [15–21]. The flat-band region has the

shape of an extended saddle point in the direction $(\pi/2, 0)$–$(\pi, 0)$. The shadow FS resembles the main FS displaced by the AFM vector $\mathbf{Q} = (\pi, \pi)$. In addition, a high-energy pseudogap is observed near the point X with an energy on the order of 0.1–0.2 eV for intermediate doping [13, 22–24]. In the rigid-band model, the isotropic minimum of the spectrum for undoped compounds must result in small hole pockets near the point N as the doping increases. However, experimental data reveal no such pockets [24], and the FS for optimal and intermediate dopings appears to be arc-shaped [25]. This obvious contradiction can be resolved only by considering the evolution of the spectral density during doping.

So far we have used only particular modifications of the Kondo lattice model. However, this model (just as, e.g., the t–J model) does not contain any important realistic features of the CuO$_2$ plane, in particular, the specific lattice form and direct oxygen–oxygen hoppings. The effective spin-fermion model [26, 27] (obtained by reducing the even more complex and complete Emery model) may be assumed to be sufficiently realistic for the CuO$_2$ plane.

All of the above ARPES results in a wide doping range can be explained in terms of the spin-polaron approach for the spin-fermion model [28] when the complex spin-polaron structure (the superposition of polaron states with different radii [29]) and frustration in the spin subsystem are taken into account. To this end, we must include both the Kondo hoppings of oxygen holes attended by the copper spin flip with amplitude τ and the direct oxygen–oxygen hoppings with amplitude h in the spin-fermion model Hamiltonian. For the spin subsystem, we must take into account the AFM interaction between the first (J_1) and second (J_2) nearest neighbors of a square copper lattice. The constants J_1 and J_2 can be expressed in terms of the frustration parameter p as follows: $J_1 = (1 - p)J$ and $J_2 = pJ$, $0 \leq p \leq 1$, $J > 0$. The frustration parameter p increases with hole concentration x.

As previously, we solve the problem in terms of the projection technique with allowance for the spherically symmetric state of the spin subsystem. The spin-polaron basis operators at finite temperatures are constructed in a way similar to that used in the preceding section (given the two possible oxygen-hole positions in an elementary cell).

We present our results only for weak doping, $x = 0.02$ ($p = 0.1$), and for the temperature $T = 0.2J$ (the typical parameters $\tau \sim 0.4$ eV, $J = 0.4\tau$, and $h = 0.4\tau$ were chosen).

Figure 4 shows the spectrum $\varepsilon_1(\mathbf{k})$ and the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ of a bare hole in the lower band with $\varepsilon_1(\mathbf{k}) = $ const and $n_{h,\sigma}^{(1)}(\mathbf{k}) = $ const isolines for the frustration parameter $p = 0.1$. The circles are plotted on the contours of constant energy that correspond to a FS
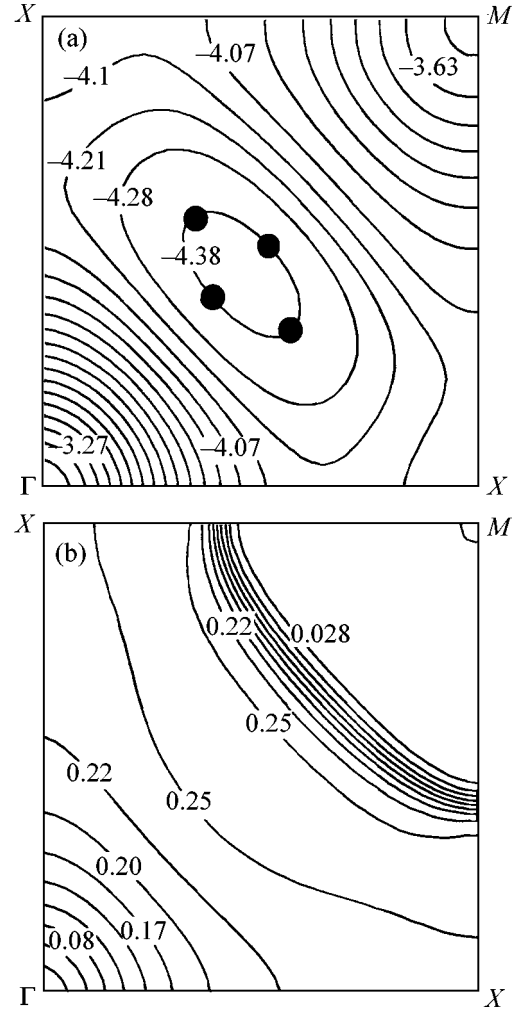

Fig. 4. (a) The hole spectrum $\varepsilon_1(\mathbf{k})$ and (b) the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ of a bare hole for the lower band are represented by $\varepsilon_1(\mathbf{k}) = $ const (in units of $\tau = 1$) and $n_{h,\sigma}^{(1)}(\mathbf{k}) = $ const isolines. The circles in (a) are plotted on the contour of constant energy that corresponds to the FS calculated for the hole doping $x = 0.02$ (the frustration parameter $p = 0.1$). The circle diameter is proportional to the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ on the FS. The results are presented in the first quadrant of the Brillouin zone. For the symmetry points, we use the notation $\Gamma = (0, 0)$, $X = \{(\pi, 0), (0, \pi)\}$, and $M = (\pi, \pi)$.

with $x = 0.02$. The circle diameter is proportional to the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ on the FS.

As can be seen from Fig. 4a, the minimum of $\varepsilon_1(\mathbf{k})$ is close to the point N and the spectrum is fairly isotropic near the band bottom. The spectrum along the Γ–M and Γ–X–M directions reproduces the ARPES results (cf., e.g., Fig. 3 in [13]). The width of the lower hole (upper electron) band, $W_1 \approx (4.4 - 3.3)\tau = 0.44$ eV, also corresponds to the ARPES results: $W_1 \approx 0.2$ eV for

$Bi_2Sr_2CaCuO_{6+0.5}$ [13], $W_1 \approx 0.3$ eV for $Ca_2CuO_2Cl_2$ [14], and $W_1 \approx 0.35$ eV for La_2CuO_4 [12]. Some uncertainty in these experimental data is attributable to the drop in spectral weight near the point Γ.

The most important result for very small doping is a sharp decrease in the spectral weight of the lower band when moving along **k** from point *N* to point *M* (see Fig. 4b). We see that the **k** line of this sharp decrease in spectral weight is close to the line that gives the remnant FS in the ARPES experiments [12, 14]. The spin-polaron spectrum in Fig. 4a exhibits a symmetry similar to the symmetry of the Brillouin magnetic zone, but the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ has the symmetry of the true Brillouin zone (see Fig. 4b).

A detailed analysis [28] indicates that taking into account the complex spin-polaron structure and the actual features of the CuO_2 plane makes it possible to describe the main ARPES experimental results in terms of the same approach and in a wide hole-doping range. In particular, these results include the existence of a pseudogap for intermediate doping and a large FS, as well as the displacement of the band bottom to point $M = (\pi, \pi)$ when passing to optimal doping.

5. SUPERCONDUCTIVITY IN THE SPIN-POLARON APPROACH

So far we have considered only the normal properties of strongly correlated systems. Let us now turn to superconductivity. In this section, we show that it would be not suffice to consider the pairing of bare holes $a_{\mathbf{k}\sigma}^+$ when studying the superconducting properties [30]. A more complex excitation is required, i.e., a spin polaron that is generally described by the operator

$$p_{\mathbf{k}\sigma}^+ = a_{\mathbf{k}\sigma}^+ + b_{\mathbf{k}\sigma}^+, \tag{43}$$

where $b_{\mathbf{k}}^+$ is the sum of terms in the following symbolic form: $a_{\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2-\dots-\mathbf{k}_n}^+ \hat{S}_{\mathbf{k}_1}\hat{S}_{\mathbf{k}_2}\dots\hat{S}_{\mathbf{k}_n}$, $\hat{S}_{\mathbf{k}_i}$ (see, e.g., [31]).

Let us return to a simple Kondo lattice, as in subsection 3.1. Below, we show that for this model it is important to take into account anomalous Green's functions of the form $\langle\langle b_{-\mathbf{k},-\sigma}^+ | a_{\mathbf{k},\sigma}^+ \rangle\rangle$. Moreover, we show that no superconductivity is possible in the approach under consideration without allowance for such Green's functions. In fact, this implies that the strong interaction of electrons with excitations of the antiferromagnetically ordered spin subsystem is the superconducting coupling mechanism in this model.

Recall that the Hamiltonian of a (square) Kondo lattice is

$$H = \sum_{\mathbf{r},\mathbf{g}} t_{\mathbf{g}} a_{\mathbf{r}+\mathbf{g},\sigma}^+ a_{\mathbf{r},\sigma}$$

$$+ J\sum_{\mathbf{r},\sigma_1,\sigma_2} a_{\mathbf{r},\sigma_1}^+ S_{\mathbf{r}}^\alpha \hat{\alpha}_{\sigma_1,\sigma_2}^\alpha a_{\mathbf{r},\sigma_2} + \frac{1}{2}I\sum_{\mathbf{r},\mathbf{g}} S_{\mathbf{r}+\mathbf{g}}^\alpha S_{\mathbf{r}}^\alpha. \tag{44}$$

All of the designations have been explained above.

Let us now write the first two equations from an infinite chain of equations for the Green's functions describing the hole motion against the AFM background [cf. (27) and (28)]:

$$\omega\langle\langle a_{\mathbf{r}\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle = \delta_{\mathbf{r},\mathbf{r}_1} + \sum_{\mathbf{g}} t_{\mathbf{g}}\langle\langle a_{\mathbf{r}+\mathbf{g},\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle$$

$$+ J\langle\langle b_{\mathbf{r}\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle, \tag{45}$$

$$\omega\langle\langle b_{\mathbf{r}\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle = \sum_{\mathbf{g},\sigma_1} t_{\mathbf{g}}\langle\langle S_{\mathbf{r}}^\alpha \hat{\sigma}_{\sigma,\sigma_1}^\alpha a_{\mathbf{r}+\mathbf{g},\sigma_1} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle$$

$$+ iIe_{\alpha\beta\gamma}\sum_{\mathbf{g},\sigma_1} \hat{\sigma}_{\sigma,\sigma_1}^\alpha \langle\langle S_{\mathbf{r}+\mathbf{g}}^\beta S_{\mathbf{r}}^\gamma a_{\mathbf{r},\sigma_1} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle$$

$$+ J\frac{3}{4}\langle\langle a_{\mathbf{r}\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle - J\langle\langle b_{\mathbf{r}\sigma} | a_{\mathbf{r}_1,\sigma}^+ \rangle\rangle$$

$$+ J\sigma\left\langle\left\langle 2b_{\mathbf{r},-\sigma}^+ \sum_{\sigma_1}(\sigma_1 a_{\mathbf{r},\sigma_1} a_{\mathbf{r},-\sigma_1}) \middle| a_{\mathbf{r}_1,\sigma}^+ \right\rangle\right\rangle. \tag{46}$$

Here, $b_{\mathbf{r}\sigma} = S_{\mathbf{r}}^\alpha \hat{\sigma}_{\sigma,\sigma_1}^\alpha a_{\mathbf{r},\sigma_1}$. As previously, we take a spherically symmetric singlet state with a zero mean spin projection at any lattice site ($\langle S_{\mathbf{r}}^\alpha \rangle = 0$) and non-zero AFM spin correlations $\langle S_0^\alpha S_{\mathbf{r}}^\alpha \rangle$ as the ground state of the spin subsystem. The last term in Eq. (46) is proportional to the carrier concentration *x*, which is a small parameter of the problem ($x \leq 0.2$ in the region important for HTSC).

Let us first consider the system in a normal state. To close the chain of equations, we make use of the repeatedly mentioned Zwanzig–Mori projection technique [9]. For the system of Eqs. (45) and (46), this implies that the first two terms on the right-hand side of Eq. (46) must be approximated by their projections onto some chosen operator basis. When describing the system in a normal state, the simplest operator basis is formed by the two operators that emerge in the first Eq. (45). These are the annihilation operators for a bare electron, $a_{\mathbf{r}\sigma}$, and the on-site spin-polaron operator $b_{\mathbf{r}\sigma}$. The choice of such a basis implies that the spin polaron is constructed as a coherent superposition of the bare-hole operator $a_{\mathbf{k}\sigma}$ and the operator $b_{\mathbf{k}\sigma}$ for the hole associated with the local spin. This local polaron is a coun-

terpart of the Zhang–Rice singlet in the Hubbard three-band model. In the limit of strong Kondo interaction J, such local polarons best describe the true lower quasiparticle band.

To describe superconductivity, we must expand the operator basis $a_{\mathbf{r}}$ and $b_{\mathbf{r}}$ to include the operators $a_{\mathbf{r}}^+$ and $b_{\mathbf{r}}^+$ when projecting. The only term with a nonzero projection onto this additional subspace is the last term on the right-hand side of Eq. (46). Its projection onto $a_{\mathbf{r}}^+$ is zero, while its projection onto $b_{\mathbf{r}}^+$ takes the following form (note that in this case, the projection technique yields exactly the same result as does the simplest uncoupling):

$$
J\sigma\left\langle\!\!\left\langle 2b_{\mathbf{r},-\sigma}^+\sum_{\sigma_1}(\sigma_1 a_{\mathbf{r},\sigma_1}a_{\mathbf{r},-\sigma_1})\Big| a_{\mathbf{r}_1}^+\right\rangle\!\!\right\rangle
$$
$$
\approx J\eta^*\langle\langle\sigma b_{\mathbf{r},-\sigma}^+|a_{\mathbf{r}_1}^+\rangle\rangle,
\tag{47}
$$

$$
\eta^* = \left\langle\sum_{\sigma_1}(\sigma_1 a_{\mathbf{r},\sigma_1}a_{\mathbf{r},-\sigma_1})\right\rangle.
\tag{48}
$$

As a result, the anomalous Green's function $F^2 = \sigma\langle\langle b_{-\mathbf{k},-\sigma}^+|a_{\mathbf{k},\sigma}^+\rangle\rangle$ and the anomalous mean η^* emerge on the right-hand side of Eq. (46). Performing further standard calculations, we can derive a closed system of equations for the superconducting case [30] by using the projection technique described above. Determining the anomalous Green's functions from this system allows us to derive the standard equation for a superconducting gap with inclusion of the upper (ω_2) and lower (ω_1) polaron band spectra. Omitting the corresponding intermediate calculations, we give an estimate for the gap Δ at $T = 0$ in the simplest logarithmic approximation by disregarding the dependence of $\Delta_{\mathbf{k}}$ on $\mathbf{k}$. The equation for the gap has the standard form for the BKS theory, but with the effective coupling constant $g \sim J^3/(\omega_2^2 - \omega_1^2) \sim J$ rather than J^2, as would be the case for a weak Kondo interaction.

We assume that the chemical potential lies near the center of the lower band. In that case, we can easily verify that $\omega_2 \sim J$ for the typical parameters $t_g = -t$, $J = 3t$, $I = 0.4t$, $t = 1$, and $c_g = \langle\mathbf{S}_{\mathbf{r}}\mathbf{S}_{\mathbf{r}+\mathbf{g}}\rangle = -0.33$. The cutoff energy is determined either by the lower-band width W_1 or by the characteristic spin-excitation energy I. In our case, these two quantities are on the order of t. As a result, the following estimate holds for Δ:

$$
\Delta = W_1\exp\left(-\frac{\pi W_1\omega_2^2}{2\left(\frac{3}{4}\right)^2 J^3}\right) \sim t\exp(-t/J).
\tag{49}
$$

For HTSC, W_1 is typically 0.3–0.5 eV, which gives a reasonable estimate $\Delta \approx 100$ K for the gap.

The model under consideration is the most dramatic demonstration of a situation where only the term (47), which results from the projection onto the purely polaron operator b^+, leads to superconducting coupling (both of spin polarons and bare holes).

6. CONCLUSION

We have demonstrated the basic ideas behind the spin-polaron approach for strongly correlated 2D systems. The method of constructing complex polaron states (only such states can adequately describe elementary excitations), is common to all strongly correlated models. This method is much simpler and more transparent than, for example, the self-consistent Born approximation. In our view, it may also prove to be efficient in the subsequent development of the theory for strongly correlated systems, in particular, in studying their kinetic properties [32].

REFERENCES

1. A. F. Barabanov, L. A. Maksimov, and A. V. Mikheyenkov, in *Lectures on the Physics of Highly Correlated Electron Systems IV*, Ed. by F. Mancini (American Inst. of Physics, New York, 2000), AIP Conf. Proc. **527**, 1 (2000).

2. H. Shimahara and S. Takada, J. Phys. Soc. Jpn. **60**, 2394 (1991).

3. A. F. Barabanov and V. M. Berezovskiĭ, Zh. Éksp. Teor. Fiz. **106**, 1156 (1994) [JETP **79**, 627 (1994)].

4. L. A. Maksimov, A. F. Barabanov, and R. O. Kuzian, Phys. Lett. A **232**, 286 (1997).

5. L. A. Maksimov, R. Hayn, and A. F. Barabanov, Phys. Lett. A **238**, 288 (1998).

6. F. C. Zhang and T. M. Rice, Phys. Rev. B **41**, 7243 (1990).

7. H. Eskes, M. B. J. Meinders, and G. A. Sawatzky, Phys. Rev. Lett. **67**, 1035 (1991).

8. A. F. Barabanov, O. V. Urazaev, A. A. Kovalev, and L. A. Maksimov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 386 (1998) [JETP Lett. **68**, 412 (1998)].

9. H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

10. A. P. Kampf and J. R. Schrieffer, Phys. Rev. B **42**, 7967 (1990).

11. A. F. Barabanov and O. A. Starykh, J. Phys. Soc. Jpn. **61**, 704 (1992).

12. B. O. Wells, Z.-X. Shen, A. Matsuura, *et al.*, Phys. Rev. Lett. **74**, 964 (1995).

13. D. S. Marshall, D. S. Dessau, A. G. Loeser, *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).

14. F. Ronning, C. Kim, D. L. Feng, *et al.*, Science **282**, 2067 (1998).

15. J. G. Tobin, C. G. Olson, C. Gu, *et al.*, Phys. Rev. B **45**, 5563 (1992).

16. K. Gofron, J. C. Campuzano, H. Ding, *et al.*, J. Phys. Chem. Solids **54**, 1193 (1993).

17. A. A. Abrikosov, J. C. Campuzano, and K. Gofron, Physica C (Amsterdam) **214**, 73 (1993).

18. D. S. Dessau, Z.-X. Shen, D. M. King, *et al.*, Phys. Rev. Lett. **71**, 278 (1993).

19. D. M. King, Z.-X. Shen, D. S. Dessau, *et al.*, Phys. Rev. Lett. **73**, 3298 (1994).

20. P. Aebi, J. Osterwalder, P. Schwaller, *et al.*, Phys. Rev. Lett. **72**, 2757 (1994).

21. V. Borisenko, M. S. Golden, S. Legner, *et al.*, Phys. Rev. Lett. **84**, 4453 (2000).

22. A. G. Loeser, Z.-X. Shen, D. S. Dessau, *et al.*, Science **273**, 325 (1996).

23. H. Ding, T. Yokoya, J. C. Campuzano, *et al.*, Nature **382**, 51 (1996).

24. H. Ding, M. R. Norman, T. Yokoya, *et al.*, Phys. Rev. Lett. **78**, 2628 (1997).

25. M. R. Norman, H. Ding, M. Randeria, *et al.*, Nature **392**, 157 (1998).

26. V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987).

27. V. J. Emery and G. Reiter, Phys. Rev. B **38**, 4547 (1988).

28. A. F. Barabanov, A. A. Kovalev, O. V. Urazaev, *et al.*, Zh. Éksp. Teor. Fiz. **119**, 777 (2001) [JETP **92**, 677 (2001)].

29. A. F. Barabanov, A. A. Kovalev, O. V. Urazaev, and A. M. Belemouk, Phys. Lett. A **265**, 221 (2000).

30. A. F. Barabanov, L. A. Maksimov, and A. V. Mikheenkov, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 362 (2001) [JETP Lett. **74**, 328 (2001)].

31. O. F. de Alcantara Bonfim and G. F. Reiter, in *High-Temperature Superconductivity: Physical Properties, Microscopic Theory and Mechanisms: Proceedings of the University of Miami Workshop on Electronic Structure and Mechanisms for High-Temperature Superconductivity*, Ed. by J. Ashkenazi (Plenum, New York, 1991).

32. S. A. Gordyunin, A. M. Belemouk, A. E. Karakozov, and A. F. Barabanov, Phys. Lett. A **285**, 197 (2001).

Translated by V. Astakhov

Momentum Space Topology of Fermion Zero Modes on Brane[1]

G. E. Volovik

Low Temperature Laboratory, Helsinki University of Technology, FIN-02015 HUT, Finland
Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117940 Russia
Received December 10, 2001

We discuss fermion zero modes within the 3 + 1 brane, i.e., the domain wall between the two vacua in 4 + 1 spacetime. We do not assume relativistic invariance in 4 + 1 spacetime or any special form of the 4 + 1 action. The only input is that the fermions in bulk are fully gapped and are described by a nontrivial momentum-space topology. Then the 3 + 1 wall between such vacua contains chiral 3 + 1 fermions. The bosonic collective modes in the wall form the gauge and gravitational fields. In principle, this universality class of fermionic vacua can contain all the ingredients of the Standard Model and gravity. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 04.50.+h; 11.25.Mj; 11.27.+d; 73.43.-f

INTRODUCTION

The idea that our Universe lives on a brane embedded in higher dimensional space [1] is popular at the moment. It is the further development of old ideas of extra compact dimensions introduced by Kaluza [2] and Klein [3]. In a new approach, compactification occurs because the low-energy physics is concentrated within the brane, for example, in a flat 4-dimensional brane embedded in a 5-dimensional anti-de Sitter space with a negative cosmological constant [4]. Branes can be represented by topological defects, such as domain walls (membranes) and strings. It is supposed that we live inside the core of such a defect. This new twist in the idea of extra dimensions is fashionable, because by accommodation of the core size one can bring the gravitational Planck energy scale close to the TeV range. That is why there is hope that the deviations from Newton's law may become observable even at a distance on the order of 1 mm. At the moment, the Newton's law has been tested for distances >0.2 mm [5].

The particular mechanism us to why the matter is localized on the brane is that the low-energy fermionic matter is represented by the fermion zero modes, whose wave function is concentrated in the core region. Outside the core, the fermions are massive and thus are frozen out at a low temperature T. An example of such topologically induced Kaluza–Klein compactification of the multidimensional space is provided by the condensed-matter analogs of branes, i.e., domain walls and vortices. These topological defects do contain fermion zero modes which can live only within the core of defects. These fermions form the 2 + 1 world within the domain wall and 1 + 1 world within the domain wall in

quasi-two-dimensional thin films or in the core of the linear defects—quantized vortices.

Recently, an attempt was made to "construct" the 4 + 1 condensed-matter system with gapped fermions in bulk and gapless excitations on the 3 + 1 boundary, which include gauge bosons and gravitons [6]. This is the 4 + 1-dimensional generalization of the 2 + 1 quantum Hall effect (QHE) in the presence of the external $SU(2)$ gauge field, where the low-energy fermions are the analogs of the so-called edge states [7] on the boundary of the system.

Here we show that there is a natural scenario in which the chiral fermions emergently appear in the brane together with gauge and gravitational field. Instead of the QHE system, we consider the system in which the quantization of Hall conductivity occurs without external magnetic field. In this scenario, the topology of momentum space [8, 9] plays the central role determining the universality classes. We consider the domain wall which separates two 4 + 1 quantum vacua with nontrivial topology in the momentum space. If the momentum-space topological invariants are different on the two sides of the wall, such a 3 + 1 brane contains fermion zero modes, that is, the gapless 3 + 1 fermions. Close to the nodes in the energy spectrum, (Fermi points) these fermions are chiral. The collective bosonic modes within the brane correspond to gauge and gravitational fields acting on fermion modes.

As distinct from the relativistic theories [4], in this scenario the existence of gauge and gravitational fields in the brane does not require the existence of the corresponding 4 + 1 fields in the bulk. The 3 + 1 fields in the brane emergently arise as collective modes of fermionic vacuum, much as they arise in quantum liquids belonging to the universality class of Fermi points [8]. Thus, the brane separating the 4 + 1 vacua with differ-

---

[1] This work was submitted by the author in English.

ent momentum-space topologies is one more universality class of the quantum vacua, whose properties are dictated by the momentum-space topology.

WALLS IN 2 + 1 SYSTEMS

Let us first consider how all this occurs in 2 + 1 systems, after which it can be easily generalized to the 4 + 1 case. For the 2 + 1 systems, it is known that the quantization of Hall or spin-Hall conductivity can occur even without an external magnetic field. This quantization is provided by the integer-valued momentum-space topological invariant [10]:

$$
N_3 = \frac{1}{24\pi^2} e_{\mu\nu\lambda}
$$
$$
\times \mathrm{tr} \int dp_x dp_y dp_0 \, \mathcal{G} \partial_{p_\mu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\nu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\lambda} \mathcal{G}^{-1}. \tag{1}
$$

Here, $\mathcal{G}$ is the fermionic propagator expressed in terms of the momentum $p_\mu = (\mathbf{p}, p_0)$, where $\mathbf{p} = (p_x, p_y)$ and p_0 is the frequency on the imaginary axis. In the most simple examples, which occur, for example, in thin films of ^{3}He and probably in the atomic layers of some superconductors, one has $\mathcal{G}^{-1} = z - \mathcal{H}(\mathbf{p})$, where $z = ip_0$, and the 2×2 Hamiltonian $\mathcal{H}(\mathbf{p}) = \tau^i g_i(p_x, p_y)$ is expressed in terms of Pauli matrices τ^i. In this case, Eq. (1) is simplified:

$$
N_3 = \frac{1}{4\pi} \int \frac{dp_x dp_y}{|\mathbf{g}|^3} \mathbf{g} \cdot \left(\frac{\partial \mathbf{g}}{\partial p_x} \times \frac{\partial \mathbf{g}}{\partial p_y} \right). \tag{2}
$$

The invariant exists only if the fermions are gapped, i.e., if their energy $E(p_x, p_y) = |\mathbf{g}| \neq 0$. The value of Hall or spin-Hall conductivity depends on this invariant, and that is why the quantization of conductivities occurs without external field. The invariant N_3 can be varied by varying the film thickness, instead of varying the magnetic field in conventional QHE. Similar invariants have been used in conventional QHE too, see [11, 12].

Another important property of the conventional QHE, which is reproduced by the system under discussion, is the existence of the edge states on the boundary of the system, or on the boundary separating vacua with different values of quantized conductivity. Let us consider the domain wall (the 1 + 1 brane) separating vacua with different topological invariants on the left and on the right side of the wall: $N_3(\mathrm{right})$ and $N_3(\mathrm{left})$. If $N_3(\mathrm{right}) \neq N_3(\mathrm{left})$, one finds that there are fermion zero modes. These are the gapless branches $E(p_\parallel)$, where $p_\parallel$ is the linear momentum along the wall. These branches cross zero energy when $p_\parallel$ varies. Close to zero energy, the spectrum of the ath fermion zero mode is linear:

$$
E_a(p_\parallel) = c_a(p_\parallel - p_a). \tag{3}
$$

These fermion zero modes correspond to the chiral (left-moving and right-moving) gapless edge states in QHE. There is an index theorem which determines the algebraic number ν of the fermion zero modes, i.e., the number of modes crossing zero with positive slope (right-moving) minus the number of modes with negative slope (left-moving):

$$
\nu = \sum_a \mathrm{sgn}\, c_a. \tag{4}
$$

According to this theorem, which is similar to the Atiyah–Singer index theorem [13] relating the number of fermion zero modes to the topological charge of the gauge field configuration, one has [14]

$$
\nu = N_3(\mathrm{right}) - N_3(\mathrm{left}). \tag{5}
$$

The crossing point p_a on each branch is nothing more than the Fermi surface in 1D momentum space $p_\parallel$. In general, the Fermi surfaces can be described by the topological invariant N_1 expressed in terms of Green's function [8]

$$
N_1 = \mathrm{Tr} \oint_C \frac{dl}{2\pi i} \mathcal{G}(p_0, p_\parallel) \partial_l \mathcal{G}^{-1}(p_0, p_\parallel) = \nu. \tag{6}
$$

Here, $\mathcal{G}$ is the propagator for the 1 + 1 fermion zero modes and the contour C embraces all the points ($p_0 = 0$, $p_\parallel = p_a$) where Green's function is singular. In the simplest case, the propagator for the 1 + 1 fermion zero modes has the form

$$
\mathcal{G}^{-1} = ip_0 - E_a(p_\parallel), \tag{7}
$$

and the contour C embraces the point ($p_0 = 0$, $p_\parallel = p_a$) in momentum space. The equation $N_1 = N_3(\mathrm{right}) - N_3(\mathrm{left})$ illustrates the topology of the dimensional reduction in the momentum space: the momentum-space topological invariant N_3 of the bulk 2 + 1 system gives rise to the 1 + 1 fermion zero modes described by the momentum-space topological invariant N_1.

BRANES IN 4 + 1 SYSTEMS

Now we can move from the 2 + 1 to the 4 + 1 dimension. Let us suppose that we have a quantum liquid in 4 + 1 spacetime which contains a 3 + 1 domain wall separating two domains, each with fully gapped fermions. Then, everything can be obtained from the case of the quantum liquid in 2 + 1 spacetime just by increasing the dimension.

According to the analogy with 2 + 1 systems, the 4 + 1 gapped fermions must have nontrivial momentum-space topology. Such topology is described by the invariant N_5 instead of N_3:

$$
N_5 = C_5 e_{\mu\nu\lambda\alpha\beta} \mathrm{tr} \int d^5 p \, \mathcal{G} \partial_{p_\mu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\nu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\lambda}
$$
$$
\times \mathcal{G}^{-1} \mathcal{G} \partial_{p_\alpha} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\beta} \mathcal{G}^{-1}. \tag{8}
$$

Here, $p_\mu = (p_0, \mathbf{p})$, where $\mathbf{p} = (p_1, p_2, p_3, p_4)$ is the momentum in 4D space; p_0 is the energy considered at imaginary axis $z = ip_0$; and C_5 is proper normalization. It is the difference $N_5(\text{right}) - N_5(\text{left})$ of invariants on both sides of the domain wall [extension of $N_3(\text{right}) - N_3(\text{left})$] which must give rise to the $3 + 1$ fermion zero modes within the brane.

The relativistic example of the propagator with nontrivial invariant N_5 is provided by $\mathcal{G}^{-1} = ip_0 - \mathcal{H}$, where the Hamiltonian in the 4D space is $\mathcal{H} = M\Gamma^5 + \sum_{i=1}^{4} \Gamma^i p_i$, and Γ^{1-5} are 4×4 Dirac matrices satisfying the Clifford algebra $\{\Gamma^a, \Gamma^a\} = 2\delta^{ab}$. In this example, the proper domain wall containing the fermion zero modes separates the domains with opposite signs of mass parameter M, since for such a wall $N_5(\text{right}) = -N_5(\text{left})$. The existence of fermion zero modes in such a domain wall is a well-known fact in relativistic theories. We would like to stress, however, that the existence of fermion zero modes does not require the relativistic theory in the bulk. It will suffice to have the nontrivial invariant N_5, which determines the universality class of a fermionic vacuum.

In the $1 + 1$ wall, the energy spectrum of fermion zero modes in the wall crosses zero at points in 1D momentum space. Thus, the energy spectrum of fermion zero modes in the $3 + 1$ brane must be zero at points in 3D momentum space. This means that the spectrum has Fermi points. Fermi points are described by the momentum-space topological invariant N_3, which is now the difference between the number of right-handed and left-handed fermions [8]:

$$N_3 = \frac{1}{24\pi^2} e_{\mu\nu\lambda\gamma}$$

$$\times \text{tr} \int_{\sigma_3} dS^{\gamma} \mathcal{G} \partial_{p_\mu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\nu} \mathcal{G}^{-1} \mathcal{G} \partial_{p_\lambda} \mathcal{G}^{-1}. \tag{9}$$

Here, the integral is over the 3-dimensional surface σ_3 embracing the singular points ($p_0 = 0$, $\mathbf{p} = \mathbf{p}_a$) of the spectrum. This is the analog of invariant N_1 in Eq. (6). Close to the ath Fermi point, the fermion zero modes represent $3 + 1$ chiral fermions, whose propagator has the general form expressed in terms of the tetrad field:

$$\mathcal{G}^{-1} = \sigma^\nu e^\mu_{\nu a}(p_\mu - p_{\mu a}). \tag{10}$$

Here, $\sigma^\nu = (1, \sigma)$ and σ are Pauli matrices.

In the same manner as in Eqs. (5) and (6), which relate the number of fermion zero modes to the topological invariants in bulk $2 + 1$ domains, the total topological charge of the Fermi points within the domain wall is expressed through the difference of the topological invariants in bulk $4 + 1$ domains:

$$N_3 = N_5(\text{right}) - N_5(\text{left}). \tag{11}$$

The quantities $e^\mu_{\nu a}$ and $p_{\mu a}$, which enter the fermionic spectrum, are dynamical variables. These are the low-energy collective bosonic modes which play the part of the effective gravitational and gauge fields, correspondingly, acting on chiral fermion [8]. These fields emergently arise in the fermionic vacuum with nontrivial momentum-space topology. The brane between the topologically different vacua thus represents one more universality class of the "emergent behavior" [15].

In a similar manner, the gauge and gravity fields arise as collective modes on the boundary of the $4 + 1$ system exhibiting the quantum Hall effect [6]. Both systems have similar topology: in [6], the nontrivial topology is provided by the external field, while in our case it is assumed that the vacuum itself has a nontrivial topology, $N_5 \neq 0$, even without the gauge field.

CONCLUSION

We showed that, if the momentum-space topology of the fermionic vacuum in $4 + 1$ spacetime is nontrivial, the $3 + 1$ domain wall between the two such vacua contains chiral fermions, while bosonic collective modes in the wall are the gauge and gravitational fields. This emergent behavior does not depend on the details of action in the bulk $4 + 1$ systems, or on the details of the brane structure. Neither Lorentz invariance nor the gravity in $4 + 1$ bulk system are required for the emergency of chiral fermions and collective fields in the brane. However, the nontrivial topology alone does not guarantee that the gravitational field will obey Einstein equations: the proper (maybe discrete) symmetry and the proper relations between different "Planck" scales in the underlying fermionic system are required [8]. The energy scale which marks the cutoff for the integrals over fermions must be much smaller than the energy scale at which the Lorentz invariance is violated. We hope that within this universality class one can obtain all the ingredients of the Standard Model and gravity.

REFERENCES

1. V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **125**, 136 (1983); K. Akama, in *Lecture Notes in Physics*, Vol. 176: *Gauge Theory and Gravitation*, Ed. by K. Kikkawa, N. Nakanishi, and H. Nariai (Springer-Verlag, Berlin, 1983), pp. 267–271; K. Akama, hep-th/0001113.

2. T. Kaluza, Sitzungsber. K. Preuss. Akad. Wiss. **K1**, 966 (1921).

3. O. Klein, Z. Phys. **37**, 895 (1926).

4. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).

5. C. D. Hoyle, U. Schmidt, B. R. Heckel, *et al.*, Phys. Rev. Lett. **86**, 1418 (2001).

6. S. C. Zhang and J. Hu, Science **294**, 823 (2001).

7. B. I. Halperin, Phys. Rev. B **25**, 2185 (1982); X. G. Wen, Phys. Rev. Lett. **64**, 2206 (1990); M. Stone, Phys. Rev. B **42**, 8399 (1990).

8. G. E. Volovik, Phys. Rep. **351**, 195 (2001).

9. G. E. Volovik, in *Proceedings of the XVIII CFIF Autumn School "Topology of Strongly Correlated Systems", Lisbon, 2000* (World Scientific, Singapore, 2001), p. 30; cond-mat/0012195.

10. G. E. Volovik and V. M. Yakovenko, J. Phys.: Condens. Matter **1**, 5263 (1989).

11. M. Kohmoto, Ann. Phys. **160**, 343 (1985).

12. K. Ishikawa and T. Matsuyama, Z. Phys. C **33**, 41 (1986); Nucl. Phys. B **280**, 523 (1987); T. Matsuyama, Prog. Theor. Phys. **77**, 711 (1987).

13. M. F. Atiyah and I. M. Singer, Ann. Math. **87**, 484 (1968); **93**, 119 (1971).

14. G. E. Volovik, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 492 (1997) [JETP Lett. **66**, 522 (1997)].

15. R. Laughlin and D. Pines, Proc. Natl. Acad. Sci. USA **97**, 28 (2000).

Possible Pseudogap Phase in QCD[1]

K. Zarembo

Department of Physics and Astronomy and Pacific Institute for the Mathematical Sciences,
University of British Columbia, Vancouver, V6T 1Z1 Canada

Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia

e-mail: Konstantin.Zarembo@teorfys.uu.se

Received December 14, 2001

Thermal pion fluctuations, in principle, can completely disorder the phase of quark condensate and thus restore chiral symmetry. If this happens before the quark condensate melts, strongly interacting matter will be in the pseudogap state just above the chiral phase transition. The quark condensate does not vanish locally, and quarks acquire constituent masses in the pseudogap phase, despite the fact that chiral symmetry is restored. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 12.38.Mh; 12.39.Fe

The physics of light hadrons is to a large extent controlled by the approximate chiral symmetry of QCD. The order parameter of the chiral symmetry, the quark condensate $\langle \bar{\psi}\psi \rangle$, acquires a nonzero expectation value in the QCD vacuum and chiral symmetry proves to be spontaneously broken. Chiral symmetry breaking yields quark constituent masses of order 350–400 Mev and thus sets the scale of hadron masses. Pions arise as low-energy pseudo-Goldstone excitations of the chiral condensate. If strongly interacting matter is heated to the temperature that exceeds a critical value of 150–200 Mev, chiral symmetry is restored. Usually, the restoration of chiral symmetry is associated with melting of quark condensate. I will discuss a less familiar mechanism of symmetry restoration by phase decoherence, which, if realized, implies the existence of an intermediate phase, similar to the pseudogap phase of high-T_c superconductors, the analogy with which I will extensively use. In the pseudogap phase, quarks still condense and acquire constituent masses, but chiral symmetry is not broken because the condensate phase is completely disordered. The potential relevance of the pseudogap phenomenon to QCD was pointed out by Babaev and Kleinert [1, 2], who examined low-dimensional toy models of chiral symmetry breaking [1–3] and the Nambu–Jona–Lasinio model [2, 4].

A continuous symmetry associated with the complex order parameter $\Psi = \rho e^{i\varphi}$ is usually restored when the free energy of the symmetry-breaking state with nonvanishing condensate $\rho \neq 0$ starts to exceed the free energy of the symmetric state with $\rho = 0$. However, phase decoherence can restore the symmetry even if $\rho \neq 0$. When the phase of the condensate is completely

disordered, $\langle e^{i\varphi} \rangle = 0$, the expectation value of the order parameter, apparently, vanishes: $\langle \Psi \rangle = 0$. There is growing evidence that such a mechanism is realized in some high-T_c superconductors [5, 6], whose normal nonsuperconducting state possesses many features characteristic of superconductivity. Most notably, the energy gap does not close above the point at which superconductivity is destroyed and gradually disappears only at much higher temperature. A theoretical explanation of the pseudogap phenomenon [5] relies on the essentially two-dimensional nature of high-T_c superconductivity. In two dimensions, topological phase fluctuations of the superconducting condensate (vortices) are extremely important. Depending on the temperature and on the phase stiffness (the energy cost of phase fluctuations), vortices either are bound in pairs or form a plasma. These two phases are separated by the Berezinsky–Kosterlitz–Thouless (BKT) transition [7]. In high-T_c superconductors, the temperature of a BKT transition is lower than the temperature at which the condensate of Cooper pairs melts and the superconducting gap, accordingly, shrinks. The superconductivity is then destroyed by vortices that unbind at the BKT transition and completely disorder the condensate phase.

In QCD, the chiral order parameter is an $N_f \times N_f$ matrix (N_f is the number of light quark species): $\Sigma_{ff'} = \bar{\psi}_f \psi_{f'}$. Its vacuum expectation value is diagonal, because light quarks have small current masses, which align the condensate in a particular direction. Throughout this paper, I will discuss the chiral limit and neglect quark masses. The coherent fluctuations of the condensate phase cost no energy in this approximation and correspond to the Goldstone modes of the broken chiral

---

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

symmetry. The chiral phase of the quark condensate is a unitary $N_f \times N_f$ matrix:

$$e^{i\gamma^5 \lambda^a \pi^a} = \frac{1 - \gamma^5}{2} U + \frac{1 + \gamma^5}{2} U^\dagger,$$

where λ^a are $SU(N_f)$ generators and π^a are the Goldstone fields associated with pions.

The low-energy dynamics of pions is described by the chiral Lagrangian:

$$\mathscr{L}_\chi = \frac{F_\pi^2}{4} \mathrm{tr} \partial_\mu U^\dagger \partial^\mu U, \tag{1}$$

where $F_\pi = 93$ Mev is the pion decay constant. The classical thermodynamics of the nonlinear sigma model was extensively studied by Monte Carlo simulations [8, 9], and, indeed, the phase transition associated with the disordering of the chiral field was found in numerical simulations: $\langle \mathrm{tr} U \rangle \neq 0$ below the critical point, and $\langle \mathrm{tr} U \rangle = 0$ above it. The large-N analysis [10] suggests that the phase transition does not disappear if one goes from classical to quantum thermodynamics. Since

$$\langle \bar{\psi} \psi \rangle = \langle |\Sigma| (\mathrm{tr} U + \mathrm{tr} U^\dagger) \rangle, \tag{2}$$

the chiral condensate turns to zero when $\langle \mathrm{tr} U \rangle = 0$ and chiral symmetry is restored. The question is whether the temperature of pion disordering is lower or higher than the temperature at which the quark condensate melts. If it is lower, as suggested by the smallness of the pion decay constant (the analog of phase stiffness in QCD), then the pseudogap phase exists in the intermediate range of temperatures. The results of [11, 10] seem to support the hypothesis that the chiral phase transition is essentially driven by the angular fluctuations of quark condensate.

If the chiral transition is caused by pion decoherence, the phase diagram of QCD will look as shown in the figure, with the pseudogap phase sandwiched between the hadron and the quark–gluon plasma phases. It should be mentioned that no symmetry and no order parameter can distinguish the pseudogap phase from the quark–gluon plasma, since the separation of the quark condensate in phase and modulus only makes sense within the low-energy approximation, when phase fluctuations are sufficiently light compared to all other modes. The pseudogap phase cannot be separated from quark–gluon plasma by a phase transition, inasmuch as there is no phase transition between the pseudogap state and the normal state in high-T_c superconductors. The dashed line in the figure thus denotes a smooth crossover, which can be rather broad. The distinctive feature of the pseudogap phenomenon is that the symmetry restoration affects only the Goldstone modes, while parameters associated with other excitations are continuous or almost continuous across the phase transition. In particular, masses of all excitations in the pseudogap phase, except for pions, will be determined by constituent quark mass and thus will be rather large. If there is no pseudogap phase and the constituent quark mass disappears above the chiral transition, masses of non-Goldstone modes are expected to decrease at the critical temperature.

A particular mechanism that can lead to pion decoherence is the disordering of the chiral condensate by baryons [9, 12]. The idea behind this mechanism closely follows the analogy with high-T_c superconductivity. In the scenario proposed in [9, 12], baryons disorder the chiral condensate in the same way as vortices disorder the phase of a superconducting gap in two dimensions. The key point is that baryons can be associated with topological excitations of the chiral field [13]. Inside the baryon, $U(x)$ winds around an $SU(2)$ subgroup of $SU(N_f)$. Therefore, in a sufficiently dense random ensemble of baryons and antibaryons, the chiral field is randomly distributed over $SU(N_f)$, which results in $\langle \mathrm{tr} U \rangle = 0$. Numerical simulations of 3D sigma model [9] indicate that this picture is probably correct. In particular, baryon susceptibility undergoes a dramatic rise in the vicinity of phase transition [9]. It was independently observed in [12] that the thermal density of an ideal baryon gas is comparatively large already at $T \sim 150$–200 MeV, despite a small Boltzmann factor associated with large baryon masses. The latter is compensated by a large entropy due to a large number of baryon resonances. From this point of view, the phase transition in the nonlinear sigma model resembles the BKT transition [12]. The pseudogap phase is then analogous to the high-temperature plasma phase of the 2D XY model, in which vortices are liberated and all correlations are screened. Similarly, there are no pions in
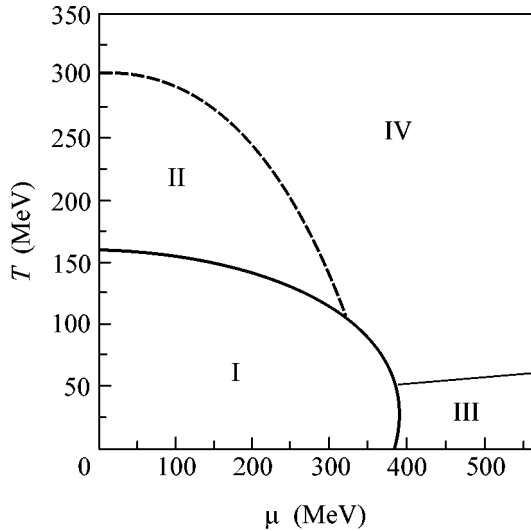
QCD phase diagram: I hadron phase, II pseudogap phase, III color superconducting phases, and IV quark–gluon plasma.

the pseudogap phase and all excitations of the chiral field must be massive.

The whole idea of the pseudogap mechanism relies on the assumption that the low-energy approximation is still accurate near the chiral phase transition. Therefore, the effective chiral Lagrangian (1) should still make sense in pseudogap phase, though the chiral perturbation theory, at least in its straightforward implementation, should break down. Since the linear and the nonlinear sigma models share the same symmetries and belong to the same universality class, the order of the transition is determined by universality arguments and depends on N_f [14]. For realistic quark masses, Monte-Carlo simulations indicate [15] that the transition becomes a sharp but smooth crossover.

The phase transition in the nonlinear sigma model is a nonperturbative phenomenon which is very hard to describe analytically. Nevertheless, some information can be deduced from simple dimensional arguments. The only dimensional parameter of the nonlinear sigma model is the pion-decay constant. Consequently, the critical temperature should be proportional to the pion-decay constant with some numerical coefficient:

$$T_c \propto F_\pi. \qquad (3)$$

Here, F_π stands for the "bare" pion-decay constant, that is, the coefficient of the kinetic term in the chiral Lagrangian obtained after integrating over all heavy degrees of freedom. Taking into account the thermal pion loops, which effectively reduce F_π [16], would be a double counting. Once the dependence of the decay constant on the temperature and the chemical potential is known, condition (3) can be used to locate the critical line in the T–μ plane:

$$\frac{F_\pi^2(T_c(\mu), \mu)}{T_c^2(\mu)} = \frac{F_\pi^2(T_c(0), 0)}{T_c^2(0)}. \qquad (4)$$

I will calculate F_π as a function of temperature and chemical potential within the framework of the semiphenomenological constituent quark model of [17, 18], which was rather successful in describing chiral dynamics and the nucleon properties in [19, 18]. Pions arise in this model as chiral phases of the constituent quark mass:

$$\mathcal{L} = \overline{\psi}(i\partial - Me^{i\gamma^5\lambda^a\pi^a})\psi$$
$$= \overline{\psi}\left[i\partial - M\left(\frac{1-\gamma^5}{2}U + \frac{1+\gamma^5}{2}U^\dagger\right)\right]\psi. \qquad (5)$$

This form of constituent quark Lagrangian was motivated by the instanton liquid model [17]. In fact, such an interaction of constituent quarks with pions will arise, after a Hubbard–Stratonovich transformation, from any four-quark interaction, local or nonlocal, which retains the symmetries of QCD. The chiral Lagrangian in this model is obtained after integration over quark fields and subsequent derivative expansion

of the fermion determinant. The pion-decay constant is the coefficient of the first term with the smallest number of derivatives. The expression for the pion-decay constant obtained in this way [19, 18] can easily be generalized to the case of nonzero temperature and chemical potential:

$$F_\pi^2(T, \mu) = 4N_cM^2T$$
$$\times \sum_n \int \frac{d^3p}{(2\pi)^3} \frac{1}{\{[(2n-1)\pi T - i\mu]^2 + p^2 + M^2\}}$$
$$= F_\pi^2 - \frac{N_cM^2}{2\pi^2} \int_M^\infty \frac{d\omega}{\sqrt{\omega^2 - M^2}} \qquad (6)$$
$$\times \left[\frac{1}{e^{(\omega-\mu)/T} + 1} + \frac{1}{e^{(\omega+\mu)/T} + 1}\right],$$

where $N_c = 3$ is the number of colors. Both the temperature and the chemical potential tend to decrease F_π, so that a larger chemical potential requires a lower critical temperature to satisfy Eq. (4), as expected.

Lattice simulations give $T_c(0) = 150$ MeV at $N_f = 3$ and $T_c(0) = 170$ Mev at $N_f = 2$ in the chiral limit [15]. The boundary of the hadron phase in the figure is obtained by solving Eq. (4) with $T_c(0) = 160$ MeV, $M = 350$ MeV, and $F_\pi = 93$ MeV. Other lines in the figure are drawn somewhat arbitrarily. Since the critical temperature changes slowly in a rather wide range of chemical potentials and is always smaller than the constituent quark mass, Boltzmann statistics should be a good approximation unless μ is close to M. The expansion in T/M then yields the following analytic expression for the critical line at $\Delta T \equiv T_c(\mu) - T_c(0) \ll T_c(0)$:

$$\frac{\Delta T}{T_0} \approx \frac{N_cM^2}{2\pi F_\pi^2}\sqrt{\frac{T_0}{2\pi M}}e^{-M/T_0}\left(\cosh\frac{\mu}{T_0} - 1\right), \qquad (7)$$

where $T_0 \equiv T_c(0)$

When T is small and μ is sufficiently large, the transition to the color superconducting state is expected to occur [20]. This transition cannot be driven by pion disordering, so that condition (4) is expected to work only for $\mu < M$. Nevertheless, the equation $F_\pi(0, \mu_c) = 0$, to which Eq. (4) reduces at zero temperature, gives a reasonable value of μ_c, which is roughly consistent with various estimates of a critical chemical potential for the color superconducting phase transition [21].

The chiral symmetry is broken in a color superconductor by the diquark condensate, whose phase excitations are similar to ordinary pions [22] and are massless in the chiral limit. However, the pion-decay constant in a color superconductor is rather large [23], and it is very unlikely that pion decoherence can emerge in the color superconducting phase.

Finally, I should mention that the pseudogap phase, if it exists at $N_c = 3$, disappears in the large-N_c limit.[2] The pion-decay constant (chiral phase stiffness) grows as $F_\pi^2 = O(N_c)$ at large N_c and suppresses fluctuations of the chiral field that could drive pion decoherence. Condition (3) then gives $T_c = O(\sqrt{N_c})$, which is smaller than the estimate based on the ideal gas approximation for baryons [12], $T_c = O(N_c/\ln N_c)$, but still grows with N_c, unlike the temperature at which the chiral condensate completely melts, which is supposed to be finite at $N_c \longrightarrow \infty$.

REFERENCES

1. H. Kleinert and E. Babaev, Phys. Lett. B **438**, 311 (1998); hep-th/9809112; E. Babaev, Phys. Lett. B **497**, 323 (2001); hep-th/9907089.

2. E. Babaev, Int. J. Mod. Phys. A **16**, 1175 (2001); hep-th/9909052.

3. R. MacKenzie, P.K. Panigrahi, and S. Sakhi, Int. J. Mod. Phys. A **9**, 3603 (1994); hep-th/9309047.

4. E. Babaev, Phys. Rev. D **62**, 074020 (2000); hep-ph/0006087.

5. V. J. Emery and S. A. Kivelson, Nature **374**, 434 (1995).

6. J. Corson, R. Mallozzi, J. Orenstein, *et al.*, Nature **398**, 221 (1999); Z. A. Xu, N. P. Ong, Y. Wang, *et al.*, Nature **406**, 486 (2000).

7. V. L. Berezinsky, Zh. Éksp. Teor. Fiz. **59**, 907 (1970) [Sov. Phys. JETP **32**, 493 (1971)]; J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973).

8. J. B. Kogut, M. Snow, and M. Stone, Nucl. Phys. B **200**, 211 (1982); P. Dreher, Phys. Lett. B **281**, 127 (1992); Nucl. Phys. B **389**, 265 (1993).

9. C. DeTar, Phys. Rev. D **42**, 224 (1990).

10. A. Bochkarev and J. Kapusta, Phys. Rev. D **54**, 4066 (1996); hep-ph/9602405.

11. C. DeTar and S. Dong, Phys. Rev. D **45**, 4681 (1992).

12. I. I. Kogan, A. Kovner, and B. Tekin, Phys. Rev. D **63**, 116007 (2001); hep-ph/0101040.

13. T. H. R. Skyrme, Nucl. Phys. **31**, 556 (1962); E. Witten, Nucl. Phys. B **223**, 433 (1983).

14. R. D. Pisarski and F. Wilczek, Phys. Rev. D **29**, 338 (1984); F. Wilczek, Int. J. Mod. Phys. A **7**, 3911 (1992); Erratum: **7**, 6951 (1992).

15. F. Karsch, E. Laermann, and A. Peikert, hep-lat/0012023; F. Karsch, hep-ph/0103314.

16. J. Gasser and H. Leutwyler, Phys. Lett. B **184**, 83 (1987); S. Jeon and J. Kapusta, Phys. Rev. D **54**, 6475 (1996); hep-ph/9602400.

17. D. Diakonov and V. Y. Petrov, Nucl. Phys. B **272**, 457 (1986).

18. D. Diakonov, hep-ph/9802298.

19. D. Diakonov, V. Y. Petrov, and P. V. Pobylitsa, Nucl. Phys. B **306**, 809 (1988).

20. K. Rajagopal and F. Wilczek, hep-ph/0011333.

21. J. Berges and K. Rajagopal, Nucl. Phys. B **538**, 215 (1999); hep-ph/9804233; G. W. Carter and D. Diakonov, Phys. Rev. D **60**, 016004 (1999); hep-ph/9812445; R. Rapp, T. Schafer, E. V. Shuryak, and M. Velkovsky, Ann. Phys. **280**, 35 (2000); hep-ph/9904353.

22. M. Alford, K. Rajagopal, and F. Wilczek, Nucl. Phys. B **537**, 443 (1999); hep-ph/9804403; T. Schafer and F. Wilczek, Phys. Rev. Lett. **82**, 3956 (1999); hep-ph/9811473.

23. D. T. Son and M. A. Stephanov, Phys. Rev. D **61**, 074012 (2000); hep-ph/9910491; Erratum: **62**, 059902 (2000); hep-ph/0004095; K. Zarembo, Phys. Rev. D **62**, 054003 (2000); hep-ph/0002123; S. R. Beane, P. F. Bedaque, and M. J. Savage, Phys. Lett. B **483**, 131 (2000); hep-ph/0002209; V. A. Miransky, I. A. Shovkovy, and L. C. Wijewardhana, Phys. Rev. D **62**, 085025 (2000); hep-ph/0009129; **63**, 056005 (2001); hep-ph/0009173.

[2] [1, 2] contain a detailed discussion of what happens to the pseudogap phase in the large-N limit of $(2 + \varepsilon)$-dimensional Gross–Neveu model.)

Interaction of Nonlinear Resonances in Cavity QED

S. V. Prants

Il'ichev Pacific Oceanological Institute, Far East Division, Russian Academy of Sciences,
ul. Baltiĭskaya 43, Vladivostok, 690041 Russia
Received December 13, 2001

Strong coupling of the internal and external degrees of freedom of a cold atom to each other and to the spatially periodic field of the standing light wave in a high-finesse cavity is responsible for the dynamic instability of the atomic center-of-mass motion. Due to a weak interaction of the internal nonlinear resonances in the standard model of cavity QED, a stochastic layer appears, whose width in the semiclassical approximation is estimated in terms of the main parameters of the system: atomic recoil frequency, mean number of excitations, and detuning from the resonance. As a result, the atomic motion in the absolutely regular potential has the fractal character, with long Lévy flights alternating with small chaotic oscillations in potential wells. © *2002 MAIK "Nauka/Interperiodica".*

PACS numbers: 05.45.Mt; 42.50.Vk; 12.20.Ds

Experiments with individual atoms and photons in high-finesse Fabry–Perot cavities [1, 2] provide new possibilities for controlling the internal and external quantum states of atoms, their cooling, and for processing quantum information. Modern cavity QED is well-developed for the verification of the main postulates of quantum mechanics (entangled quantum states, nonlocality of interaction, etc.) and for the study of the fundamental problem of quantum–classical correspondence—quantum chaos. Along with the theoretical and numerical studies of the strongly coupled atomic–field systems with Hamiltonians generating classical chaos [3–6], the QED variant was experimentally implemented in [7, 8] for the theoretically well-studied quantum rotator model, i.e., a cold atom periodically excited by standing-wave pulses with a frequency detuned far from the atomic transition frequency. For large detunings (compared to the natural linewidth), the atomic center-of-mass dynamics in the field of a modulated standing wave is described by the Hamiltonian with 3/2 degrees of freedom [9]. Generally speaking, the interaction of atoms with photons in a high-finesse cavity is the interaction of three degrees of freedom—field, intra-atomic, and translational—even in the simplest one-dimensional case. This model was proposed in [10] for the Hamiltonian chaos.

The purpose of this study is to reveal and analyze a mechanism responsible for the appearance of the Hamiltonian chaos in the interaction of cold atoms with the standing-wave field in a high-finesse cavity. It is shown in the semiclassical approximation that chaos arises due to the modulation of the slow atomic translational motion by fast Rabi oscillations, resulting in the formation of a narrow stochastic layer in the vicinity of the separatrix of unperturbed motion even in the absence of any standing-wave modulation. On the basis

of the reduced equations of motion, the width of the stochastic layer is analytically found in terms of the main parameters of the system: detuning, atomic recoil frequency, and mean number of excitations.

In the rotating-wave approximation, the interaction of a cold two-level atom with a certain harmonic mode of quantized electromagnetic field of a perfect cavity with allowance for the atomic recoil is described by the Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}\hbar\omega_a\hat{\sigma}_z + \hbar\omega_f\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right)$$
$$- \hbar\Omega_0(\hat{a}^\dagger\hat{\sigma}_- + \hat{a}\hat{\sigma}_+)\cos(k\hat{x}), \tag{1}$$

where $\hat{x}$ and $\hat{p}$ are the operators of the atomic coordinate and momentum, respectively; $\hat{\sigma}$ are the Pauli matrices; and $\hat{a}$ and $\hat{a}^\dagger$ are the field-mode operators. Taking the quantum averages $\xi = k_f\langle\hat{x}\rangle$, $\rho = \langle\hat{p}\rangle/\hbar k$, $u = \langle\hat{a}^\dagger\hat{\sigma}_- + \hat{a}\hat{\sigma}_+\rangle$, $v = i\langle\hat{a}^\dagger\hat{\sigma}_- - \hat{a}\hat{\sigma}_+\rangle$, and z as dynamic variables, one can easily demonstrate that the Heisenberg equations generate the following closed nonlinear set of five semiclassical equations of motion:

$$\dot{\xi} = \alpha\rho, \quad \dot{\rho} = -u\sin\xi, \quad \dot{u} = \delta v,$$
$$\dot{v} = -\delta u + \left[(2N-1)z + \frac{3}{2}z^2 + \frac{1}{2}\right]\cos\xi, \tag{2}$$
$$\dot{z} = -2v\cos\xi,$$

where the dot means the differentiation with respect to the dimensionless time $\tau = \Omega_0 t$ and the control parameters $\alpha = \hbar k_f^2/m\Omega_0$, $\delta = (\omega_f - \omega_a)/\Omega_0$, and $N = \langle\hat{a}^\dagger\hat{a} + (\hat{\sigma}_z + 1)/2\rangle$ are the normalized atomic recoil frequency, the normalized resonance detuning, and the

mean number of excitations, respectively. The set of Eqs. (2) generalizes corresponding equations discussed in [10] to the case of arbitrary N. The integral of motion

$$W = \frac{\alpha\rho^2}{2} - u\cos\xi - \frac{\delta}{2}z \qquad (3)$$

corresponds to the energy conservation in the problem.

It follows from Eqs. (2) that, at the exact resonance, the slow translational variables ξ and ρ are separated from the fast intra-atomic and field variables u, v, and z. In the spatially periodic optical potential $U = -u_0\cos\xi$, the motion of the atomic center of mass is described by the simple equation $\ddot{\xi} + \alpha u_0\sin\xi = 0$ of a free nonlinear pendulum and, depending on energy W, is either regular oscillations in a potential well or regular flights over the peaks of U. The variations in the internal energy z of atom and in the energy of its interaction with the mode u for $\delta = 0$ are the Rabi oscillations modulated by the standing wave and having periodically varying frequency [10].

To reveal the mechanism of Hamiltonian chaos for $\delta \neq 0$, let us analyze Eq. (2) in the limit of large photon number, $N \gg 1$, and moderate detuning, $|\delta| > 1$. The normalized Rabi frequency is a value on the order of $\sqrt{N} > 1$ and is much higher than the frequency $\sqrt{\alpha u_0} \ll 1$ of small-amplitude translational oscillations (the estimate of the dimensionless recoil frequency of an atom strongly coupled to the field mode of a high-finesse cavity gives a value on the order of $\alpha < 10^{-2}$). As a result, the equations for fast variables can be represented as the Bloch-type equations

$$\dot{u} = \delta v, \quad v = -\delta u + 2Nz\cos\xi,$$
$$\dot{z} = -2v\cos\xi, \qquad (4)$$

where the function $\cos\xi$ can be taken to be a constant c in the time interval covering many Rabi oscillations. In this approximation, one can easily find the general solution to Eqs. (4). We are only interested in the solution for the interaction energy between the internal degrees of freedom and the mode

$$u = u(0)\left[N\left(\frac{2c}{\Omega_N}\right)^2 + \left(\frac{\delta}{\Omega_N}\right)^2\cos\Omega_N\tau\right]$$
$$+ \frac{\delta}{\Omega_N}v(0)\sin\Omega_N\tau + \frac{2N\delta c}{\Omega_N^2}s_z(0)(1 - \cos\Omega_N\tau), \qquad (5)$$

where $\Omega_N = \sqrt{\delta^2 + (2c)^2 N}$ is the Rabi frequency. Since the function $\cos\xi$ varies with time much more slowly compared to the rapidly oscillating variables u, v, and z, the variable u can be regarded as a spatially independent amplitude- and frequency-modulated signal controlling the motion of the atomic center of mass according to the equation of a frequency-modulated nonlinear oscillator

$$\ddot{\xi} + \alpha u(\tau)\sin\xi = 0. \qquad (6)$$

Without loss of generality, the analysis of this equation can be simplified for $c = 1$ and the special initial condition $u_0 = v_0 = 0$ and $z_0 = 1$ corresponding to the completely excited atom at $\tau = 0$ and to the arbitrary field state. In this case, Eq. (6) is derived from the following classical Hamiltonian:

$$\mathcal{H} = \frac{1}{2}\dot{\xi}^2 - \omega^2\cos\xi + \omega^2\cos\Omega_N\tau\cos\xi$$
$$= \mathcal{H}_0 + V, \qquad (7)$$

where $\mathcal{H}_0$ is the unperturbed Hamiltonian of a free nonlinear oscillator with the frequency $\omega = \sqrt{2\alpha N|\delta|}/\Omega_N$ of small-amplitude oscillations. Representing the perturbation V in the form

$$V = \frac{\omega^2}{2}[\cos(\xi + \Omega_N\tau) + \cos(\xi - \Omega_N\tau)], \qquad (8)$$

one can treat Eq. (7) as the Hamiltonian of a particle moving in the field of three plane waves in the coordinate system moving with the phase velocity of the first wave, while the phase velocities of the second and third waves are equal to Ω_N and $-\Omega_N$ respectively.

Near the separatrix, the variation in the translational atomic energy $E = \mathcal{H}_0$ is calculated using the Poisson brackets [11]

$$\Delta E = \int_{-\infty}^{\infty}\{\mathcal{H}_0, V\}d\tau = \omega^2\int_{-\infty}^{\infty}\dot{\xi}\sin(\xi - \Omega_N\tau)d\tau. \qquad (9)$$

This integral can be calculated using the known values of the nonlinear pendulum variables ξ and $\dot{\xi}$ taken at the separatrix, where the pendulum energy is $E_s = \omega^2$. These values are $\xi_s = 4\arctan\exp[\pm\omega(\tau - \tau_n)]$ and $\dot{\xi}_s = \pm2\omega/\cosh[\omega(\tau - \tau_n)]$, where τ_n corresponds to the center of velocity soliton. As a result, one arrives at the value

$$\Delta E = \pm4\pi\Omega_N^2\frac{\exp(\pi\Omega_N/2\omega)}{\sinh(\pi\Omega_N/\omega)}\sin\Omega_N\tau_n, \qquad (10)$$

which determines the dimensionless width of a stochastic layer of a modulated nonlinear oscillator (7) in the vicinity of the unperturbed separatrix $\delta E \simeq \max|\Delta E|/E_s$. Taking into account that $\Omega_N \gg \omega$ for real atoms, one finally obtains the expression

$$\delta E \simeq 8\pi\left(\frac{\Omega_N}{\omega}\right)^3\exp(-\pi\Omega_N/2\omega), \qquad (11)$$

where the ratio Ω_N/ω is on the order of $\sqrt{N/\alpha|\delta|}$. Note that the above simplified analysis yields the lower

bound for the width of the stochastic layer. Thus, for arbitrary parameters (under the above assumptions), the strongly coupled atomic–field system with Hamiltonian (1) always contains the domain where classical chaos originates. This domain is localized in the vicinity of the separatrix and has an exponentially small width (11). This conclusion is corroborated by the numerical experiments [10], where the Lyapunov exponents and Poincaré sections were calculated in various ranges of α, N, and δ.

The phase portrait of a parametric nonlinear oscillator described by Eqs. (7) and (8) consists of a sequence of separatrix loops in the plane $d(\xi \pm \Omega_N\tau)/d\tau - (\xi \pm \Omega_N\tau)$, with finite trajectories inside and infinite trajectories outside [11]. The width of each loop and the spacing between loops are ω and $2\Omega_N$, respectively. If $\Omega_N \gg \omega$, these nonlinear resonances interact with each other but do not overlap [12]. As a result, stochastic layers appear at the separatrix sites, with the width on the order of at least δE, which can be very small for the typical values of parameters $N \simeq 1$–10, $\alpha < 10^{-2}$, and $|\delta| \simeq 1$. The atomic motion is very sensitive to small variations of control parameters, especially of the detuning δ. The phase-space topology of a strongly coupled atomic–field system changes due to bifurcations. This opens up new possibilities for manipulating cold atoms, their cooling and acceleration, and trapping in the field one or several photons.

A typical chaotic trajectory (characterized by a small positive Lyapunov exponent) of an atom in a periodic standing-wave field consists of long portions of regular ballistic motions, so-called Lévy flights, alternating with the chaotic small-amplitude atomic oscillations in potential wells. The Lévy flights strongly affect the global transport properties of the Hamiltonian systems and lead to an anomalous or "strange" kinetics [13]. Recent numerical experiments [14] performed for an atomic–field system with three coupled degrees of freedom have demonstrated that the statistical motional characteristics such as the distribution of Poincaré recurrence times and the moments of the center of mass of a cold atom testify to anomalous atomic diffusion [14]. The Lévy flights can be used for laser cooling of atoms below the recoil limit, as was demonstrated in the experiments reported in [15, 16].

REFERENCES

1. J. Ye, D. W. Vernooy, and H. J. Kimble, Phys. Rev. Lett. **83**, 4987 (1999).
2. P. Münstermann, T. Fischer, P. Maunz, *et al.*, Phys. Rev. Lett. **82**, 3791 (1999).
3. P. I. Belobrov, G. M. Zaslavskii, and G. Kh. Tartakovskii, Zh. Éksp. Teor. Fiz. **71**, 1799 (1976) [Sov. Phys. JETP **44**, 945 (1976)].
4. G. P. Berman, E. N. Bulgakov, and D. D. Holm, *Crossover Time in Quantum Boson and Spin Systems* (Springer-Verlag, Berlin, 1994).
5. S. V. Prants, L. E. Kon'kov, and I. L. Kirilyuk, Phys. Rev. E **60**, 335 (1999).
6. S. V. Prants and L. E. Kon'kov, Phys. Rev. E **61**, 3632 (2000).
7. F. L. Moore, J. C. Robinson, C. F. Bharucha, *et al.*, Phys. Rev. Lett. **75**, 4598 (1995).
8. H. Ammann, R. Gray, I. Shvarchuck, and N. Christensen, Phys. Rev. Lett. **80**, 4111 (1998).
9. R. Graham, M. Schlautmann, and P. Zoller, Phys. Rev. A **45**, R19 (1992).
10. S. V. Prants and V. Yu. Sirotkin, Phys. Rev. A **64**, 033412 (2001).
11. G. M. Zaslavsky, R. Z. Sagdeev, D. A. Usikov, and A. A. Chernikov, *Weak Chaos and Quasiregular Patterns* (Nauka, Moscow, 1991; Cambridge Univ. Press, Cambridge, 1991).
12. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
13. M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, Nature **363**, 31 (1993).
14. S. V. Prants, M. Edelman, and G. M. Zaslavsky, Phys. Rev. E (in press).
15. F. Bardou, J. P. Bouchaud, O. Emile, *et al.*, Phys. Rev. Lett. **72**, 203 (1994).
16. J. Reichel, F. Bardou, M. Ben Dahan, *et al.*, Phys. Rev. Lett. **75**, 4575 (1995).

Translated by R. Tyapaev

Coherent Anti-Stokes Raman Scattering of Slow Light in a Hollow Planar Periodically Corrugated Waveguide

S. O. Konorov*, D. A. Akimov*, A. N. Naumov*, A. B. Fedotov*, R. B. Miles**, J. W. Haus***, and A. M. Zheltikov*

* International Laser Center, Physics Faculty, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
e-mail: zheltikov@top.phys.msu.su
** Department of Mechanical and Aerospace Engineering, Princeton University, 08544-5263 Princeton, NJ, USA
*** Electro-Optics Program, University of Dayton, 45469-0245 Dayton, OH, USA
Received December 5, 2001

Enhancement of coherent anti-Stokes Raman scattering (CARS) by molecular nitrogen in a hollow planar periodically corrugated waveguide is experimentally detected. The measured dependence of CARS efficiency on the thickness of the waveguide layer indicates that CARS enhancement under these conditions is at least partially due to the decrease in the group velocity of pump pulses around the photonic band gap. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 42.65.Dr; 42.65.Wi; 42.65.Hw

Remarkable properties of structures with a periodically modulated refractive index are opening ways to modify and control the dispersion of optical materials [1, 2]. Periodic and quasi-periodic structures, such as multilayer mirrors, microstructure fibers, as well as two- and three-dimensional periodic arrays generally referred to as photonic crystals [2, 3], are now actively used to shape, transfer, and control ultrashort light pulses [4, 5], as well as to create optical fibers of new types [6–8] and to develop optical switches, couplers, filters, and other optical components [2]. Structures with a periodically modulated refractive index often allow nonlinear-optical interactions to be phase-matched [9–12]. Physically, phase matching in photonic band-gap (PBG) structures is based on the generalized momentum conservation [9, 10] involving the reciprocal lattice vector of a periodic structure. Waveguides with a periodically perturbed refractive index offer a convenient way of extending this approach to large lengths of nonlinear-optical interaction [13], providing more degrees of freedom in reducing the phase mismatch of light pulses involved in a nonlinear-optical process.

In this paper, we will demonstrate the possibility to considerably increase the efficiency of four-wave mixing (FWM) in a gas medium filling a hollow planar corrugated waveguide. The main difference of our experimental approach from the methods used in earlier nonlinear-optical experiments in PBG waveguides (see, e.g., [13]) is that a gas filling the waveguide layer of a hollow planar waveguide serves as a nonlinear medium in our experiments. The coherence length under these conditions considerably exceeds the waveguide length (which was on the order of several centimeters in our

studies). Four-wave mixing efficiency can be improved in such a situation, as will be shown below, due to the field enhancement in a PBG waveguide related to group-velocity lowering for one or several pump waves around the PBG edge.

To experimentally implement the main idea of this work, we employed a waveguide structure consisting of a mirror and a diffraction grating. Both optical elements forming this waveguide were aluminum-coated. As shown in our previous studies [14, 15], such a structure integrates a hollow waveguide and a one-dimensional photonic crystal, combining the advantages of both of these optical components. On the one hand, high-power laser radiation can be coupled into such a waveguide, allowing ultrashort pulses to be produced due to self-phase modulation and high-order stimulated Raman scattering using the approaches similar to those developed in [16, 17] and permitting high-order harmonic generation experiments [18]. Similar to gas-filled hollow fibers, the waveguide regime of nonlinear-optical interactions in our structure improves the efficiency of wave-mixing and harmonic-generation processes relative to the regime of tightly focused pump beams due to a radical increase in the interaction length. On the other hand, a periodic perturbation of the refractive index introduced by the diffraction grating gives rise to photonic band gaps (Fig. 1), which substantially change the dispersion properties of light fields with respect to the case of a gas medium in a conventional gas cell or waveguide modes in a gas-filled hollow fiber. The created waveguide opens new ways of phase and group-velocity matching in nonlinear-optical interactions through the independent control of three main dispersion components: material dispersion, dispersion of

waveguide modes, and dispersion of a periodic structure. The material dispersion of the gas filling the waveguide can be varied by changing the gas composition and the gas pressure. The waveguide dispersion can be changed by varying the thickness of the waveguiding layer and by choosing appropriate materials for waveguide walls and a set of waveguide modes involved in the nonlinear-optical process. Finally, the period and the profile of the grating are the main knobs to control the dispersion of the periodic structure.

Our analysis of the dispersion of light waves involved in nonlinear-optical interactions in a hollow planar PBG waveguide was based on the equations of coupled-mode theory [1]. The use of this approach allowed us to find complex propagation constants for a hollow planar PBG waveguide by searching for the eigenvalues of $2N \times 2N$ square matrices of the relevant characteristic equations, where N is the number of modes of an unperturbed planar waveguide (a waveguide with no corrugation), determined from the cut-off condition. The inset in Fig. 1 presents the spectral dependences of the transmission coefficient and the effective refractive index for the lowest order bulk mode TM_2 of the hollow planar PBG waveguide calculated with the use of coupled-mode equations. The photonic band gap arises in this spectral region, as follows from the results of our calculations, due to the strong coupling of the lowest order bulk mode TM_2 with surface plasmon modes TM_0 and TM_1. The results of these calculations, as can be seen from Fig. 1, qualitatively agree with the experimental data. Our method of calculations allows the position of the photonic band gap in the transmission spectrum of a hollow planar PBG waveguide to be reproduced with reasonable accuracy (Fig. 1). However, the absolute values of the transmission coefficient, effective refractive index, and group velocity obtained with the use of the above-described approach can be considered as very rough estimates only, since these absolute values are highly sensitive to the coupling coefficients, which are not known with sufficient accuracy. These coupling coefficients depend on the Fourier amplitudes of the periodic profile of the grating, as well as on the spatial overlapping of light fields in waveguide modes and the area of a perturbed dielectric function.

The main purpose of our experimental studies was to demonstrate the possibility of enhancing coherent anti-Stokes Raman scattering (CARS) using the created hollow PBG waveguide. Since a gas filling the waveguide layer between the grating and the mirror serves as a nonlinear medium in our experiments and the pressure of this gas never exceeded the atmospheric pressure, the coherence length for CARS-type four-wave mixing processes should considerably exceed the waveguide length, which was typically on the order of several centimeters in our experiments. The efficiency of FWM processes can be increased under these conditions due to field-enhancement effects, which are char-
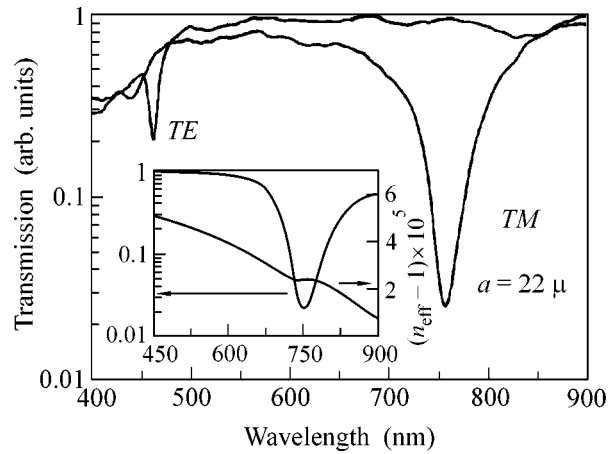

Fig. 1. Transmission spectra measured for *TM* and *TE* modes of a planar corrugated hollow waveguide consisting of a 2400-grooves/mm aluminum-coated grating and an aluminum mirror with $2a = 44$ μm. The inset shows the results of calculations for the transmission spectrum and the spectral dependence of the effective refractive index for the TM_2 mode of this waveguide.
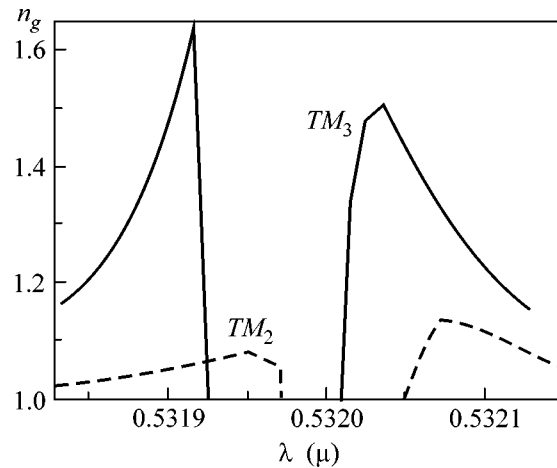

Fig. 2. The spectral dependence of the group index $n_g = c/v_g$ (c is the speed of light in vacuum and v_g is the group velocity in the waveguide structure) for the TM_2 (solid line) and TM_3 (dashed line) modes of a hollow planar corrugated waveguide with a 1200-grooves/mm diffraction grating and $2a = 22$ μm.

acteristic of PBG structures and which are related to a decrease in group velocities of light pulses around photonic band gaps in such structures. This effect is illustrated in Fig. 2, which displays the wavelength dependence of the group index for TM_2 and TM_3 modes in a hollow planar PBG waveguide with a mirror–grating gap $2a = 22$ μm within the spectral range corresponding to the second harmonic of a Nd:YAG-laser radiation. The solid and dashed lines in Fig. 2 represent the group indices for the TM_2 and TM_3 modes, respectively. The
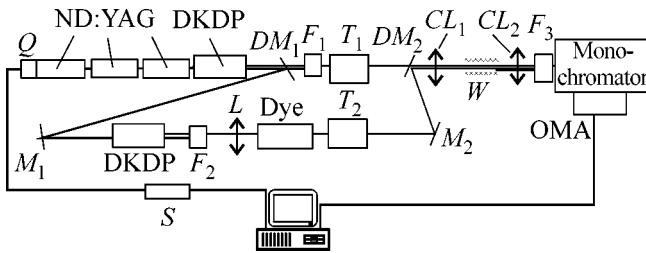
Fig. 3. Diagram of the experimental setup for studying coherent anti-Stokes Raman scattering in a hollow planar corrugated waveguide: M_1, M_2, rotating mirrors; DM_1, DM_2, dichroic mirrors; F_1–F_3, sets of optical filters; OMA, optical multichannel analyzer; S, synchronization unit; CL_1, CL_2, cylindrical lenses; L, spherical lens; T_1, T_2, telescopes; and W, waveguide.
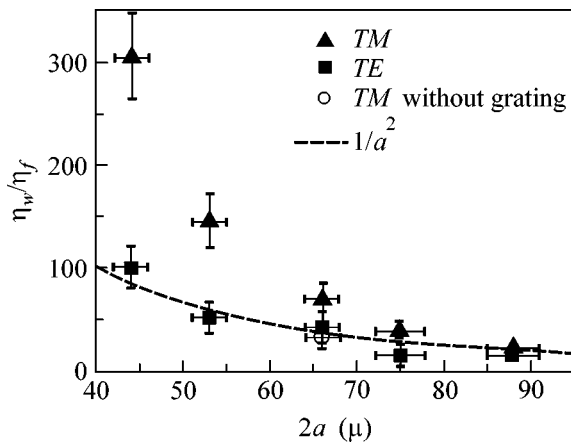

Fig. 4. The ratio of the efficiency w of the CARS process in a hollow planar PBG waveguide to the efficiency η_f of the CARS process with the same energies of pump pulses, but for cylindrically focused pump beams in the absence of a waveguide, as a function of the distance $2a$ between the waveguide walls. The CARS signal is related to Raman-active transitions of molecular nitrogen in atmospheric-pressure air. Triangles show the CARS enhancement ratio for *TM* waveguide modes, while the squares correspond to *TE* modes. The circle shows the CARS enhancement ratio for *TM*-polarized radiation in a waveguide with a second mirror instead of the diffraction grating (an unperturbed waveguide). The dashed line represents the scaling law $1/a^2$, which describes the increase in the efficiency of the CARS process in a planar waveguide with no corrugation relative to the efficiency of the same CARS process in cylindrically focused beams due to purely geometrical factors.

group velocity of light pulses, as can be seen from the results presented in Fig. 2, decreases considerably in this spectral range due to the PBG effect. This increases the mean density flux of electromagnetic radiation in the waveguide, leading to the enhancement of nonlinear-optical processes.

A diagram of the experimental setup employed to investigate the CARS process in a gas medium filling a hollow planar PBG waveguide is shown in Fig. 3. We

studied a two-color CARS process leading to the generation of a signal at the frequency $\omega_{CARS} = 2\omega_1 - \omega_2$, where ω_{CARS} is the frequency of the CARS signal and ω_1 and ω_2 are the frequencies of the pump waves. A Q-switched Nd:YAG laser, generating 15-ns pulses of 1.064-μm radiation, was employed as a master oscillator. The laser pulses produced by this oscillator were amplified up to about 30 mJ in two Nd:YAG amplification stages. The fundamental radiation was then converted into the second harmonic using a DKDP crystal. The second harmonic produced in this crystal served as one of the pump beams in the CARS process (the frequency ω_1). Fundamental radiation that remained frequency-unconverted at the output of the DKDP crystal was separated from the second harmonic with a dichroic mirror DM_1 and was employed to generate the second harmonic in the second DKDP crystal. This second-harmonic beam was then used to pump a sulforhodamine 101 dye laser. Dye-laser radiation served as the second pump beam in the CARS process (the frequency ω_2). The pump beams with the frequencies ω_1 and ω_2 were brought into spatial coincidence with a dichroic mirror DM_2 and were coupled into a hollow planar corrugated waveguide by a cylindrical lens CL_1 with a focal length of 9 cm. The energy of the second-harmonic pulse was 8 mJ, while the energy of dye-laser radiation was equal to 0.8 mJ. Aluminum-coated mirrors and 1200- and 2400-grooves/mm aluminum-coated diffraction gratings were used to create a hollow waveguide. The length of the waveguides used in our experiments was equal to 5 cm. The distance between the waveguide walls was varied from 22 up to 88 μm.

The frequency ω_2 of dye-laser radiation was chosen in such a way as to satisfy the condition of Raman resonance $\omega_1 - \omega_2 = \Omega$ with a Raman-active transition of molecular nitrogen with $\Omega = 2331$ cm^{-1}. This condition was met with the wavelength of dye-laser radiation equal to 0.607 μm. The wavelength of the CARS signal related to molecular nitrogen in the atmospheric-pressure air filling the hollow PBG waveguide was then equal to 0.473 μm. This signal was collimated with a cylindrical lens CL_2 and separated from the pump beams with a set of optical filters. Then, we let the CARS signal pass through a monochromator and detected the signal at the output of the monochromator with the use of an optical multichannel analyzer.

To characterize the enhancement of the CARS process involving Raman-active transitions of molecular nitrogen in atmospheric-pressure air in a hollow planar PBG waveguide, we compared the efficiency η_w of this process in the waveguide with the efficiency η_f of the same process with the same energies of pump pulses, but in the regime of cylindrical focusing of pump beams in the absence of a waveguide. Figure 4 presents the CARS enhancement ratio η_w/η_f measured as a function of the distance $2a$ between the mirror and the 1200-grooves/mm grating, forming a planar PBG waveguide.

Triangles show this ratio for *TM* waveguide modes, while the squares correspond to *TE* modes. The circle shows the CARS enhancement ratio for *TM*-polarized radiation in a waveguide with a second mirror instead of the diffraction grating (an unperturbed waveguide). A planar waveguide obviously provides an increase in the efficiency of any FWM process relative to the efficiency of the same FWM process in cylindrically focused beams due to geometrical factors. In contrast to FWM processes in gas-filled hollow fibers, when this geometrical enhancement ratio scales as $1/a^4$, where d is the fiber inner diameter [19, 20], a planar waveguide provides FWM enhancement scaling as $1/a^2$. This scaling law of FWM enhancement due to purely geometric factors is shown by the dashed line in Fig. 4.

The enhancement of the CARS process in the case of *TE* modes in our experiments virtually coincided, as can be seen from the data presented in Fig. 4, with the enhancement attainable in the waveguide regime due to purely geometric factors. A much higher CARS enhancement ratio, as is seen from Fig. 4, can be achieved for *TM* modes of a planar PBG waveguide, when the maximum enhancement ratio relative to the case of cylindrically focused beams may be as high as 300. These higher values of CARS enhancement ratios attainable for *TM* modes are due to the fact that the frequencies ω_1 and ω_2 of the second harmonic and dye-laser radiation fall within the range of strong coupling between the lowest order bulk mode TM_2 and one of the higher order *TM* modes. The electromagnetic energy density in the waveguide increases under these conditions, which leads to the enhancement of nonlinear-optical processes.

Since the CARS signal intensity is proportional to the product of group indices of pump fields, the decrease in the group velocity of the second-harmonic field by a factor of 1.2–1.4 and the lowering of the group velocity of dye-laser radiation by a factor of 1.1–1.2 result in the enhancement of the CARS process by a factor of 1.6–2.4. Although these estimates are, of course, very rough since the group velocities of pump waves are known only approximately, they indicate that the improvement in CARS efficiency achieved in our PBG-waveguide experiments in the case of *TM* modes may be, at least partially, attributed to the decrease in the group velocities of pump fields. Quantitative discrepancies between the experimental values of the CARS enhancement ratio and the above estimates for this ratio may be indicative of other physical factors leading to FWM enhancement under conditions of our experiments. One of these factors may be related to local field enhancement in plasmon *TM* modes, which increases the efficiency of nonlinear-optical wave mixing, leading to energy transfer to the waveguide modes that provide the dominant contribution to the FWM process. Another group of factors includes effects changing the material component of dispersion, e.g.,

the excitation and ionization of the gas medium filling the waveguide.

The results of experimental and theoretical studies presented in this paper demonstrate the possibility of a substantial enhancement of four-wave mixing processes in a gas medium in a hollow planar corrugated waveguide due to field enhancement effects related to the decrease in the group velocity of one or several pump fields around the photonic band gap. The enhancement of the CARS process achieved in our experiments, performed with a planar waveguide structure consisting of a metal mirror and a diffraction grating, can be considerably increased by optimizing the parameters and the geometry of the waveguide for a specific set of waveguide modes involved in a nonlinear-optical process. The method of enhancement of nonlinear-optical processes demonstrated in this paper opens new possibilities for improving the sensitivity of nonlinear-optical gas-phase analysis, promoting ultrashort-pulse formation with the use of self-phase modulation and high-order stimulated Raman scattering, and increasing the efficiency of high-order harmonic generation and wave mixing in gas-filled hollow waveguides.

REFERENCES

1. A. Yariv and P. Yeh, *Optical Waves in Crystals* (Wiley, New York, 1987).

2. *Nanoscale Linear and Nonlinear Optics*, Ed. by M. Bertolotti, C. M. Bowden, and C. Sibilia (American Inst. of Physics, New York, 2001).

3. J. Joannopoulos, R. Meade, and J. Winn, *Photonic Crystals* (Princeton Univ. Press, Princeton, 1995).

4. T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).

5. A. M. Zheltikov, in *Nanoscale Linear and Nonlinear Optics*, Ed. by M. Bertolotti, C. M. Bowden, and C. Sibilia (American Inst. of Physics, New York, 2001), p. 259.

6. J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, Opt. Lett. **21**, 1547 (1996).

7. J. C. Knight, J. Broeng, T. A. Birks, and P. St. J. Russell, Science **282**, 1476 (1998).

8. A. M. Zheltikov, Usp. Fiz. Nauk **170**, 1203 (2000).

9. N. Bloembergen and A. J. Sievers, Appl. Phys. Lett. **17**, 483 (1970).

10. A. Yariv and P. Yeh, J. Opt. Soc. Am. **67**, 438 (1977).

11. M. Centini, C. Sibilia, M. Scalora, *et al.*, Phys. Rev. E **60**, 4891 (1999).

12. A. M. Zheltikov, A. V. Tarasishin, and S. A. Magnitskii, Zh. Éksp. Teor. Fiz. **118**, 340 (2000) [JETP **91**, 298 (2000)].

13. J. P. van der Ziel, M. Ilegems, P. W. Foy, and R. M. Miku-lyak, Appl. Phys. Lett. **29**, 775 (1976).

14. A. B. Fedotov, A. N. Naumov, D. A. Sidorov-Biryukov, *et al.*, Laser Phys. **11**, 1009 (2001).

15. A. B. Fedotov, A. N. Naumov, D. A. Sidorov-Biryukov, *et al.*, J. Opt. Soc. Am. B (in press).

16. M. Nisoli, S. De Silvestri, O. Svelto, *et al.*, Opt. Lett. **22**, 522 (1997).

17. M. Wittmann, A. Nazarkin, and G. Korn, Opt. Lett. **26**, 298 (2001).

18. A. Rundquist, C. G. Durfee III, Z. Chang, *et al.*, Science **280** (5368), 1412 (1998).

19. R. B. Miles, G. Laufer, and G. C. Bjorklund, Appl. Phys. Lett. **30**, 417 (1977).

20. A. B. Fedotov, F. Giammanco, A. N. Naumov, *et al.*, Appl. Phys. B **72**, 575 (2001).

Translated by A.M. Zheltikov

Observation of the Coherent State in a System of Nuclear Spin-Wave Pairs Excited by Microwave Noise Pumping

A. V. Andrienko

Institute of Molecular Physics, Russian Research Centre, Kurchatov Institute,
pl. Akademika Kurchatova 1, Moscow, 123182 Russia
e-mail: andrienko@imp.kiae.ru
Received December 17, 2001; in final form, December 26, 2001

The parametric excitation of nuclear magnons by a microwave noise field was observed in an antiferromagnet. Two critical pumping amplitudes were found to exist. The first one corresponds to the onset of nonlinear microwave absorption. Above the second amplitude, strong phase correlations appear in a system of excited magnon pairs to form a nonequilibrium Bose condensate, which produces intense coherent electromagnetic radiation from the sample and gives rise to the coherent response of parametric magnons to the modulation of their spectrum (modulation response). It was found that, for the noise pumping, the contribution from the processes of elastic magnon relaxation to the threshold pumping amplitudes becomes nonadditive. © *2002 MAIK "Nauka/Interperiodica".*

PACS numbers: 75.30.Ds; 75.40.Gb

The spin and magnetoelastic waves in magnets are highly suitable objects for studying the physics of nonlinear wave processes. To date, the parametric resonance of these waves in an alternating magnetic field $\mathbf{h}\cos\omega_p t$ parallel to the static field $\mathbf{H}$ has been much studied. If the magnetic microwave field exceeds its threshold value h_c, the system becomes parametrically unstable against the decay of the pump quantum into a pair of magnons (or quasiphonons) with half frequencies and equal (but oppositely directed) wave vectors ($\omega_p = \omega_\mathbf{k} + \omega_\mathbf{-k}$). The critical amplitude h_c of the microwave field, ordinarily referred to as the pumping or parametric excitation threshold, is given by the expression $h_c = \gamma/V$, where γ is the linear relaxation rate of the excited waves and V is the coefficient of their coupling to the pumping field. Immediately above the excitation threshold (i.e., at $h > h_c$), the dynamic ordering characterized only by two parameters sets in the system: the number of parametric pairs of magnons (or quasiphonons) and their phase with respect to the pump field (see [1, 2]). Evidently, this nonequilibrium Bose condensate (NBC) of a macroscopically large number of excited quasiparticle pairs is a forced oscillation of the medium at a frequency of the external field.

The situation becomes much more complicated if the frequency of the external alternating field lies within the interval ($\omega_p - \Delta\omega/2$, $\omega_p + \Delta\omega/2$) and the pump spectral width exceeds the parameter γ of the excited waves. In this case, the occurrence of a non-equilibrium phase transition with the formation of NBC is by no means evident, because the noise field does not have a certain phase and cannot establish the coherent state in the excited system.

The noise-pumped parametric resonance of spin waves was considered by Mikhaǐlov and Uporov in [3] and by Cherepanov in [4]. The formulas obtained by these authors for the average noise pumping threshold P_n, above which the nonlinear (time-averaged) absorption appears in the system, coincided at $\Delta\omega \gg \gamma$ to a factor of two. For the arbitrary values $\Delta\omega \ll \omega_p$, the threshold formula is [5]

$$(P_n/P_c) - 1 = \Delta\omega/\gamma, \tag{1}$$

where $P_c \propto h_c^2$ is the threshold power of monochromatic pumping and the critical power P_n is the integrated noise-pump power (and not the pump spectral density).

However, the above-threshold behavior of a system of nonequilibrium magnons was treated in these works in a basically different way. In [3], it was argued that the microwave absorption is restricted mainly by the nonlinear dissipation, while the phase correlations are irrelevant. By contrast, it was stated in [4] that strong phase correlations must occur in the excited wave pairs (phase restriction mechanism), as in the case of monochromatic pumping. This question still remains to be clarified. The paper of Andrienko and Safonov [5] on the noise pumping of magnetoelastic waves in antiferromagnetic $FeBO_3$ is the only work where these phase correlations were observed experimentally. In [5], two instability thresholds were observed. The first one corresponded to the onset of nonlinear microwave absorption and the second corresponded to the appearance of NBC. However, the strong excitation anisotropy (quasiphonons were excited along the crystal C_3 axis) and
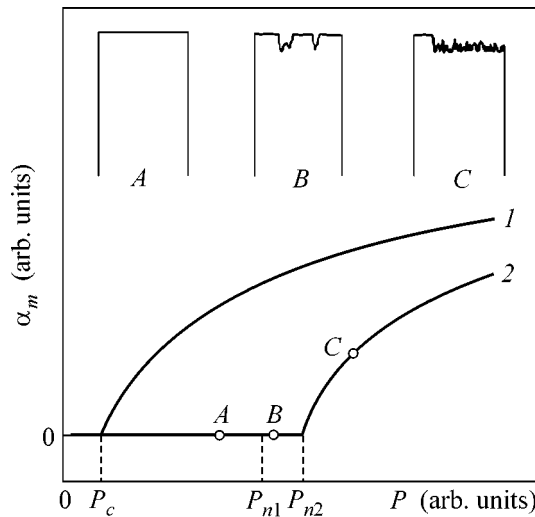
Fig. 1. Power dependence of the modulation response signal for (*1*) monochromatic and (*2*) noise pumping (recorded on a recorder), and oscillograms of the microwave noise-pump pulses passed through the cavity for points *A*, *B*, and *C*.

the possible influence of the sample boundaries on the NBC formation (the thickness of a plane-parallel platelet was much smaller than the mean free path of the magnetoelastic wave) did not allow us to answer the question about the possibility of NBC formation in an infinite sample.

One more possible incoherent action of external fields on the spin-wave system was considered by Zautkin *et al.* in [6]. In that work, the parametric excitation of electronic magnons in yttrium iron garnet by a monochromatic microwave field under conditions of noise modulation of the spin-wave spectrum was studied and it was stated that this case corresponds to the nonmonochromatic pumping with the fluctuating phase. The threshold field corresponding to the onset of nonlinear absorption and the nonlinear magnetic susceptibility of the sample were measured. Based on the qualitative agreement between the theory and the experiment, a conclusion was drawn that the phase correlations are essential. However, strong phase correlations in the wave system were not justified directly.

In this work, the noise microwave pumping of nuclear spin waves (NSWs) and the behavior of a non-equilibrium NSW system excited by the noise pumping are studied experimentally. Because the mean free path of excited waves was much shorter than the crystal sizes, the boundary effect on the formation of a steady-state parametric waves was excluded. The main goal of this study is the search for the states with strong phase correlations (nonequilibrium Bose condensate) in this system.

EXPERIMENTAL

Nuclear magnons in an easy-plane $CsMnF_3$ antiferromagnet (Néel temperature 53.5 K) were excited by a noise field with frequency $\omega_p/2\pi = 1080$–1200 MHz and spectral width $\Delta\omega/2\pi \leq 1.5$ MHz. The noise field was produced by a microwave oscillator whose frequency was modulated by a white-noise generator in the rf region, with the generated microwave power remaining constant. The pump spectrum was recorded by a spectrum analyzer and was close to Gaussian. The pump spectral width was measured at the half-maximum with an accuracy of 10%.

A single crystal of size $\sim 3 \times 3 \times 5$ mm was placed in a helium-filled helix microwave cavity with a quality factor $Q \sim 500$. Experiments were carried out at temperatures $T = 1.9$–4.2 K and fields $H = 500$–2000 Oe with parallel microwave pumping ($\mathbf{h} \parallel \mathbf{H}$). The relaxation rate of the excited magnons was determined from the monochromatic pumping threshold h_c with an accuracy of 25% [7] and was found to be equal (depending on T and H) to $\gamma/2\pi = (6$–$20)$ kHz. The accuracy of measuring γ in relative units was 5%. The magnon mean free path was $L \leq 1$ mm, i.e., several times smaller than the characteristic sample size.

The microwave pumping was performed in the cw mode and the pulsed mode with a repetition rate of 10–100 Hz and a pulse duration of 300–2000 μs. In the pulsed mode, the threshold pump power was determined from the onset of a nonlinear distortion (kink) in the microwave pulse passed through the cavity (Fig. 1). In the continuous-wave mode, it was determined from the onset of a modulation response in the spectrum of the microwave signal passed through the cavity with the sample (see below for detail). The relative accuracy of measuring the threshold power was 10%.

The nonequilibrium Bose condensate was detected by two methods based on the observation of collective effects associated with the appearance of these states.

The first method consists in the observation of modulation response α_m. This method was developed and widely used in studying the properties of a monochromatically pumped nonequilibrium magnon Bose condensate [2] and later successively used for the observation of a noise-pumped quasiphonon NBC [5]. A weak rf field $H_m\cos(\omega_m t)$ modulating the spin-wave spectrum was applied to the sample. This gave rise to the oscillations of the NBC amplitude and phase about their equilibrium values (the so-called collective oscillations) and led to the modulation of the absorbed microwave power with amplitude ΔP and frequency ω_m. The occurrence of this amplitude modulation ($\Delta P = \alpha_m H_m$) gives evidence for the presence of the NBC in the sample.

To detect α_m, the microwave signal passed through the cavity was detected, fed to a spectrum analyzer or selective microvoltmeter tuned to the frequency ω_m, and then led to the *Y* axis of an *X*–*Y* recorder.

The second method consisted in the observation of the electromagnetic radiation from a system of excited nuclear magnons. This effect was earlier observed in a system of magnetoelastic waves [8] and interpreted as a characteristic electromagnetic radiation appearing after switching off the microwave pumping of a monochromatically produced nonequilibrium Bose condensate. The frequency of this field is close to the pumping frequency; its intensity changes nonmonotonically with time and is comparable with the microwave power absorbed by the sample. In this work, the same radiation was observed from a system of nuclear spin waves.

RESULTS AND DISCUSSION

Investigations showed that two thresholds should be distinguished in the noise pumping of nuclear magnons, $P_{n1} < P_{n2}$. The first threshold corresponds to the onset of nonlinear microwave absorption and the second corresponds to the appearance of Bose condensate.

The records of the modulation response amplitude are shown in Fig. 1 as functions of the integrated pump power. The records were made in the cw generation mode. In the same figure, the oscillograms of the microwave pulses passed through the cavity and recorded for three noise-field amplitudes are shown. The nonlinear absorption at point A is absent. The absorption "splashes" above the threshold P_{n1} are seen in oscillogram B; i.e., nonequilibrium magnons exist (on the average) in the sample, but $\alpha_m = 0$. As the pump power increases, the nonlinear absorption increases (the splashes occur more frequently). Above the P_{n2} threshold, the pulse has shape C and the modulation response ($\alpha_m \neq 0$) appears, indicating the presence (on the average) of phase correlations in the system, i.e., the formation of NBC. The ratio of thresholds P_{n2}/P_{n1} increased with $\Delta\omega$ but did not exceed 1.25. Notice that this situation is cardinally different from the quasiphonon noise pumping in [5], where this ratio was as high as 4.

The occurrence of NBC was also detected from the electromagnetic radiation of the sample immediately after switching off the pump pulse (Fig. 2). It is well known that the signal power passed through the cavity is proportional to the number of photons in the cavity. In the absence of NBC, the microwave signal decays monotonically immediately after switching off the oscillator, with a characteristic photon decay time of ~0.1 μs determined by the cavity Q factor. If the pump power exceeds its threshold value, i.e., if the NBC condensate is present, the sample radiates behind the pulse trailing edge after switching off the pump field. The microwave power passed through the cavity first decreases drastically (the trailing edge falls off more steeply than at $P_{in} < P_c$). Then the signal starts to increase, and its amplitude reaches a maximum within $\tau \sim 0.5$ μs of the pulse. After this, the radiation intensity decreases exponentially with a characteristic time of ~1 μs. The maximal radiation power is comparable
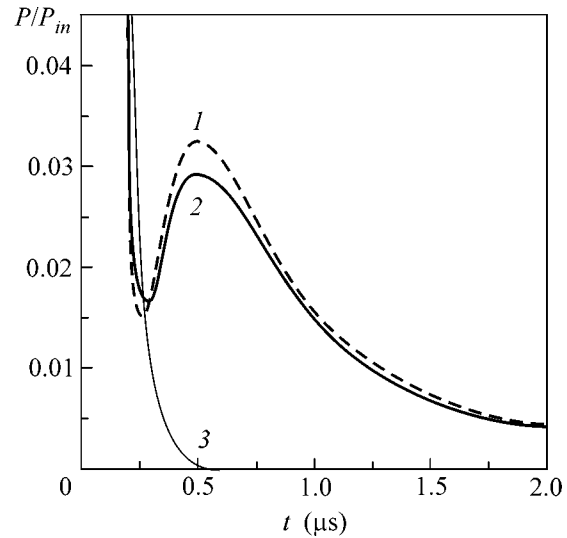

Fig. 2. Oscillograms of the trailing edges of microwave pulses and the signal of electromagnetic radiation from the sample for the input power $P_{in} \sim 1000 P_c$: (1) monochromatic and (2) noise pumping. Curve 3 demonstrates the pulse trailing edge for $P_{in} < P_c$; $t = 0$ corresponds to the moment at which the pump oscillator was switched off.

with the power absorbed by the sample from the pump field.

Similar radiation was studied earlier in [7] for the parametric excitation of the magnetoelastic waves in $FeBO_3$. It is caused by the coalescence of two quasiparticles (in our case, magnons) of a parametric pair to create a photon with a frequency close to $\omega_p (m + m \longrightarrow ph)$, i.e., by the reverse process of pumping. Inasmuch as all parametric pairs in the NBC have the same phase, this radiation is coherent and has a large amplitude, contrary to the radiation of individual magnons at frequency ω_k.

In this work, the electromagnetic radiation of parametric NSW pairs was observed for the first time. For a fixed power of the microwave oscillator such that $P \gg P_n$, the amplitudes of this radiation and the signal decay times were virtually the same for the noise and coherent pumpings. This suggests that the NBC of nuclear magnons produced by the noise pumping has approximately the same spectral width and the same amplitude as in the case of monochromatic pumping.

The NSW parametric excitation threshold is presented in Fig. 3 as a function of the spectral width of microwave pumping. Since the difference between the thresholds P_{n1} and P_{n2} did not exceed the experimental error, only the threshold P_{n1} was tracked in the measurements, and it is denoted below by P_n. In compliance with the theoretical Eq. (1), $(P_n/P_c) - 1$ was proportional to $\Delta\omega$. However, the calculated value of $(P_n/P_c) - 1$ at $T = 4.2$ K was greater than its measured value by a factor of ~3 and at $T = 2.05$ K even by a factor of ~7.
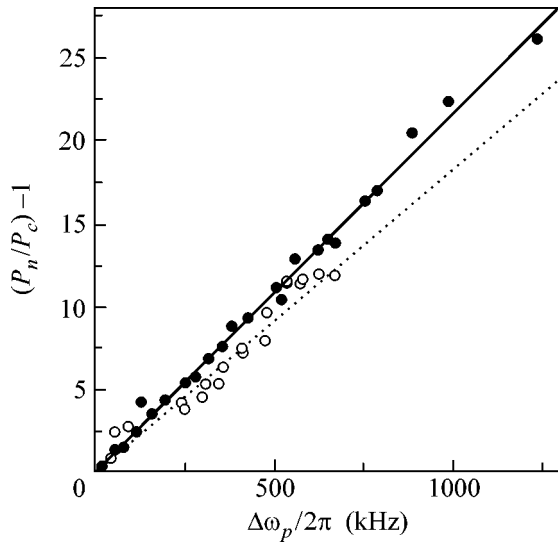
Fig. 3. Relative increase in the pump power vs. spectral width of noise pumping: (●) $T = 2.05$ K, $\omega_p/2\pi = 1088$ MHz, $H = 1750$ Oe, and $\gamma/2\pi = 7.1$ kHz and (○) $T = 4.2$ K, $\omega_p/2\pi = 1093.3$ MHz, $H = 1680$ Oe, and $\gamma/2\pi = 16.8$ kHz. The solid straight line is calculated by Eq. (1) for $T = 2.05$ K with a coefficient of 0.15, and the dotted line is calculated by Eq. (1) for $T = 4.2$ K with an adjustable multiplier of 0.31.
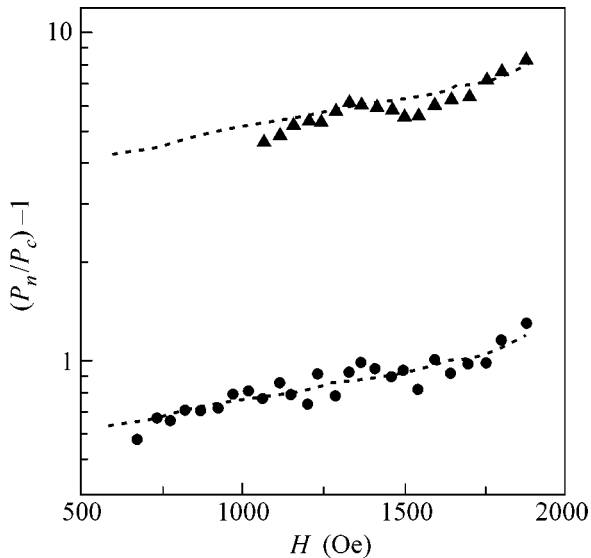

Fig. 4. Relative increase in the noise-pump threshold power vs. static magnetic field for the fixed pump spectral widths: $\Delta\omega/2\pi = $ (●) 40 and (▲) 270 kHz; $T = 2.08$ K and $\omega_p/2\pi = 1094.4$ MHz. The curves are calculated by Eq. (1) with an adjustable multiplier of 0.175.

Therefore, the theory does not account for the temperature dependence of the noise pumping threshold.

The results of measuring the noise pumping threshold as a function of static magnetic field are shown in Fig. 4 for two fixed values of $\Delta\omega$. The curves were cal-

culated in the following manner. The $\gamma(H)$ values were derived from the monochromatic pumping threshold $h_c(H)$ and substituted into Eq. (1) for calculating the H dependence of $(P_n/P_c) - 1$. A good qualitative agreement between the theory and the experiment is seen again, though with the above-mentioned adjustable parameter. It is essential that the P_c value changes 40 times in the interval of fields used. The relaxation rate of the excited nuclear magnons and the coefficient of their coupling to the microwave field also strongly vary. However, the adjustable multiplier remains constant; i.e., the theory properly reproduces the functional dependences of the noise pumping threshold P_n on the parameters of excited NSWs.

A comparison of these results with the data reported in [5], where the quasiphonon noise excitation in $FeBO_3$ was studied, shows that the NSW pumping is much better described by the theory. This is likely due to the fact that the boundary effect on the system of excited magnons is weak; i.e., the theoretical model for the NSWs is close to the experimental situation. A substantial discrepancy between the theory and experiment is observed only at low temperatures.

An analysis of the NSW relaxation mechanisms showed [8] that the NSW relaxation in $CsMnF_3$ at $T < 2.2$ K is due by $\geq 95\%$ to the NSW elastic scattering from the fluctuations of nuclear magnetization, while the contribution of elastic processes at 4.2 K comprises ~50%. In my opinion, the dominant role of the elastic relaxation processes is precisely the reason why the NSW noise pumping threshold is markedly lower than its calculated value.

The assumption about the special role of the two-magnon processes and their nonadditive contribution to the threshold formula was done by Zakharov and L'vov in [9]. The idea was that the wave energy does not change in the elastic scattering (only the wave propagation direction changes), so that the secondary magnons continue to participate in the pumping process and, hence, reduce the parametric excitation threshold, which should be given in this case by the expression

$$h_c = (\gamma_{nel}(\gamma_{nel} + \gamma_{el}))^{1/2}/V, \qquad (2)$$

where γ_{el} and γ_{nel} are the relaxation rates of the elastic and inelastic processes, respectively. However, the experiments with monochromatic pumping [7, 10] did not corroborate this assumption. The experimentally observed contribution of all γ's to the threshold amplitude was found to be additive, $h_c = (\gamma_{nel} + \gamma_{el})/V$. The point is likely that the key role is played by the wave phase, which changes after the magnon quasi-elastic scattering. Besides, the magnon energy and the absolute value of the magnon wave vector slightly change upon the quasi-elastic scattering, resulting in the so-called spectral drift. As a result, the secondary magnons are not coupled to the pump field. As for the noise field, whose frequency and phase fluctuate rapidly, neither

the phase change of the secondary magnons nor the weak spectral drift of NSWs (within $\Delta\omega$) influence their coupling to the exciting microwave field. For this reason, the secondary magnons continue to participate in the pumping process and reduce its threshold (these secondary magnons can be regarded as an additional noise field). Therefore, the contribution of two-magnon processes to the threshold amplitude should be nonadditive for the incoherent pumping. Our experiments showed that the adjustable parameter becomes more and more different from unity with an increase in relative contribution of elastic processes. Unfortunately, the theoretical calculations of the noise pumping threshold with regard to the special role of quasi-elastic processes have not been carried out as yet. Nevertheless, it is evident that the qualitative model suggested above agrees well with the experiment.

One may assume that it is the dominant contribution of elastic processes to the NSW relaxation which leads to a small difference between the P_{n1} and P_{n2} thresholds. Evidently, the NBC formation threshold P_{n2} depends on the total number of magnons in the pumping spectral range $\Delta\omega$. At $P > P_{n1}$, the nonequilibrium magnons are created in the sample by pumping. The secondary magnons formed in the two-magnon scattering processes remain in the same region of the magnon spectrum. Consequently, they are accumulated; i.e., the occupation numbers of magnons with frequencies $\omega_k \cong \omega_p/2$ rapidly increase. As a result, the critical number of magnons necessary for the formation of NBC is achieved earlier, i.e., at lower supercriticality (P/P_{n1}) than would be the case for the three-particle relaxation processes, which reduce the magnon occupation numbers near the frequency $\omega_p/2$.

In summary, the formation of a coherent state (nonequilibrium Bose condensate) of the nuclear spin-wave pairs upon their parametric excitation by a noise microwave field has been established experimentally. The existing theory satisfactorily accounts for the pumping threshold, except for the low-temperature region. The observed discrepancy between theory and experiment is explained on the basis of an assumption about the special role of the elastic magnon relaxation processes.

REFERENCES

1. V. S. L'vov, *Nonlinear Spin Waves* (Nauka, Moscow, 1987).
2. A. V. Andrienko, V. L. Safonov, and A. Yu. Yakubovskiĭ, Zh. Éksp. Teor. Fiz. **93**, 907 (1987) [Sov. Phys. JETP **66**, 511 (1987)].
3. A. S. Mikhaĭlov and I. B. Uporov, Zh. Éksp. Teor. Fiz. **77**, 2383 (1979) [Sov. Phys. JETP **50**, 1149 (1979)].
4. V. B. Cherepanov, Fiz. Tverd. Tela (Leningrad) **22**, 43 (1980) [Sov. Phys. Solid State **22**, 25 (1980)].
5. A. V. Andrienko and V. L. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 446 (1994) [JETP Lett. **60**, 464 (1994)].
6. V. V. Zautkin, B. I. Orel, and V. B. Cherepanov, Zh. Éksp. Teor. Fiz. **85**, 708 (1983) [Sov. Phys. JETP **58**, 414 (1983)].
7. A. V. Andrienko and V. L. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 787 (1994) [JETP Lett. **60**, 800 (1994)].
8. A. V. Andrienko, V. I. Ozhogin, V. L. Safonov, and A. Yu. Yakubovskiĭ, Usp. Fiz. Nauk **161** (10), 1 (1991) [Sov. Phys. Usp. **34**, 843 (1991)].
9. V. E. Zakharov and V. S. L'vov, Fiz. Tverd. Tela (Leningrad) **14**, 2913 (1972) [Sov. Phys. Solid State **14**, 2513 (1972)].
10. S. A. Govorkov and V. A. Tulin, Zh. Éksp. Teor. Fiz. **91**, 2332 (1986) [Sov. Phys. JETP **64**, 1384 (1986)].

Translated by V. Sakun

Droplet Spot as a New Object in Physics of Vacuum Discharge

A. V. Batrakov*, B. J. Jüttner**, S. A. Popov*, D. I. Proskurovski ĭ*, and N. I. Vogel***

* High Current Electronics Institute, Siberian Division, Russian Academy of Sciences,
Akademicheskiĭ pr. 4, Tomsk, 634055 Russia
e-mail: batrakov@Lve.hcei.tsc.ru
** Humboldt University Berlin, Institute of Physics, D-10117 Berlin, Germany
*** University of Technology Chemnitz, Department of Physics, 09107 Chemnitz, Germany
Received December 11, 2001

It is established experimentally that the burning of a low-current (several and tens of amperes) pulsed (microseconds) vacuum discharge is accompanied by the formation of plasma microbunches around some of the droplets leaving the cathode spot. The parameters of these bunches (electron concentration $n_e \sim 10^{26}$ m^{-3} and equilibrium temperature $T_e \sim 1$ eV) are close to the parameters of cathode-spot plasma. The data obtained suggest that the initial temperature of droplets and the thermionic emission from them play a key role in the formation of such plasma microbunches. By analogy with the well-known cathode and anode spots in vacuum discharges, these droplet plasma formations are classified as "droplet spots." This work reports the first results on studying the formation dynamics and the characteristics of the droplet spots. It is noted that the concept of droplet spots will require a certain refinement of the plasma formation mechanism in vacuum discharges. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 52.80.Vp; 52.70.Kz

It is well known [1, 2] that vacuum arcing is accompanied by the emission of droplets as a result of ejection of liquid metal from the cathode spot crater, i.e., emission center (ecton [3]) providing current transfer from the cathode to the discharge gap. The regularities of droplet generation are a source of information about the processes occurring in the cathode spot and, thereby, maintain a certain interest in the study of this phenomenon. The detection of droplet glow was and still remains one of the methods of studying the droplet emission [4–6]. However, up to now, neither the droplet radiation spectrum nor the size of the luminous area have been analyzed. It was customarily assumed that the droplet glow is caused only by the thermal radiation from the droplet surface, because the temperature of all droplets leaving the cathode spot is rather high. We established experimentally that the glow of some of the objects leaving the cathode spot is caused not by the droplet thermal radiation but by the radiation of dense plasma surrounding the droplets.

Experiments were conducted under a high oil-free vacuum of 10^{-7} torr. A pulsed arc discharge with a duration of up to 10 μs and a current of 2–20 A was initiated between the cathode and anode separated by 1–5 mm. In the experiments, two types of liquid-metal gallium–indium cathodes were used: capillary cathodes corresponding to the case of a "deep" liquid whose flow was practically not restrained by viscosity, and point cathodes in the form of a tungsten needle covered with a thin liquid-metal film. Solid (copper, tungsten, tantalum, and graphite) point cathodes were also used. The discharge current was detected by a one-gigahertz LeCroy oscilloscope. Plasma glow in the cathode discharge was detected by an Imacon-468 ultrahigh-speed camera (up to 100 M frame/s). The sensitivity range of this camera lies in the spectral range 200–700 nm. For imaging, a QUESTAR QM-100 long-focus microscope with the spectral sensitivity range 200 nm–3.5 μm was used at the camera input. The optical resolution of the apparatus was no worse than 3 μm. Plasma glow images were obtained either without an optical filter or with a 350-nm interference (10-nm width) filter.

In our experiments, we repeatedly observed the process of droplet ejection from the cathode surface in burning of the cathode spot of a low-current pulsed vacuum arc. A typical example of this observation is shown in Fig. 1.

The appearance of bright objects leaving the cathode was a great surprise in these experiments. The luminosity of these objects was comparable with the cathode spot luminosity or even exceeded it (Fig. 2). This process is very rarely observed in the arcing on a deep fluid. Its probability can be estimated at one event per 300 μs of arcing with a current of a few amperes. For the case of a thin liquid film on the refractory substrate, the probability of this process was an order of magnitude higher under the same arcing conditions,
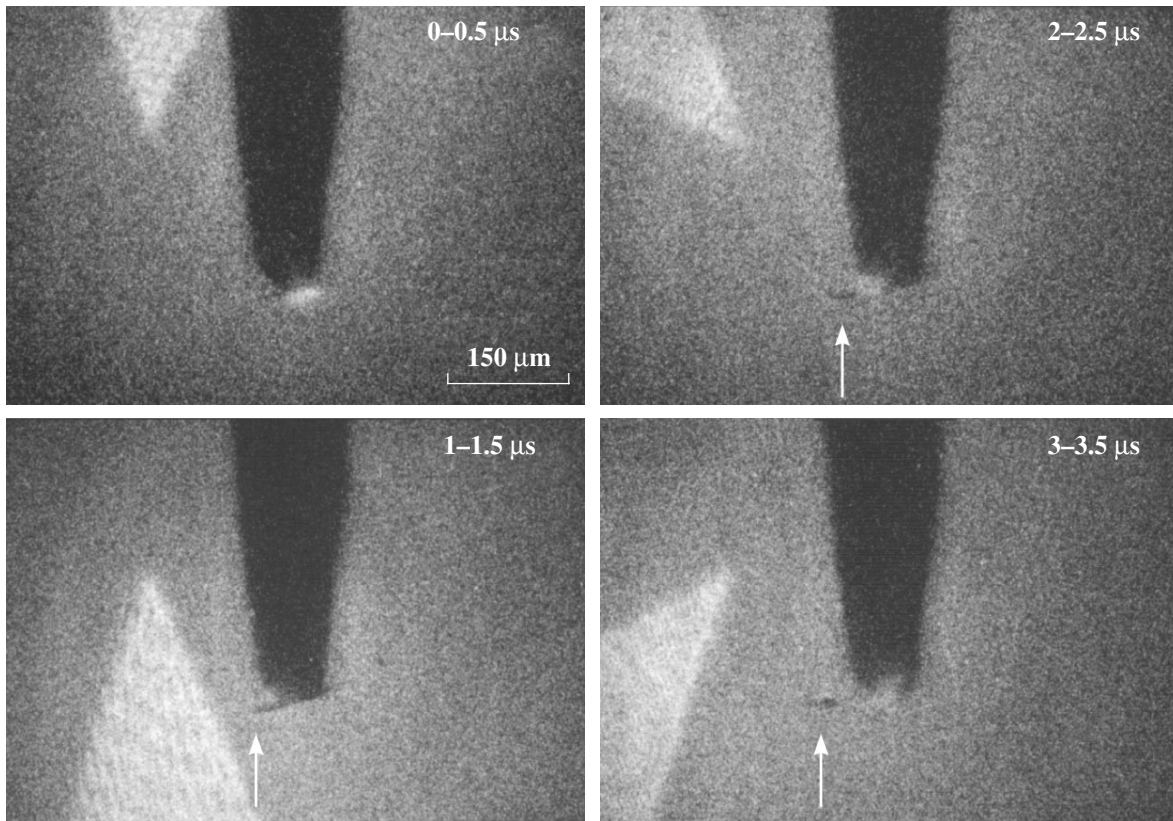
Fig. 1. The process of droplet separation from a liquid-metal point cathode. Light wedges in the images are artifacts caused by the illumination of the gap by a helium–neon laser. The exposure times are indicated in the frames. Time is read from the onset of discharge.

allowing some of the characteristic features of this phenomenon to be clarified.

The luminous objects move rectilinearly; they are not deflected by an external magnetic field of 0.15 T. The values of 10–100 m/s observed for their velocities are typical of the droplet velocities observed earlier in [5–8]. However, the sizes of these objects are 10–100 µm, which is an order of magnitude larger than the maximal sizes of droplets emitted by the cathode spot at vacuum discharge currents close to the threshold arc current [7]. Moreover, the boundaries of these objects are not sharply outlined. These facts suggest that the luminous objects are droplets surrounded by dense plasma. In this case, the observed glow is due to the visible radiation of the plasma surrounding droplets.

The surprising thing is that the luminous objects arise at a considerable distance from the cathode spot. An example of such a scenario is demonstrated in Fig. 3. The frames were obtained for arcing with a current of 5–10 A at the liquid-metal point cathode. It is clearly seen that the lowest of the luminous objects in the third frame (3–4-µs exposure) is not seen in the preceding frame. Taking into account that the time interval between the end of the preceding frame and the beginning of the subsequent frame is zero, one can conclude

that the droplet was "ignited" not in the cathode spot but at a certain distance from it (~70 µm).

The probability of "burning" droplets to appear was found to strongly depend on the cathode material and increase with its melting point. For instance, the appearance of burning droplets was not a rare event for the copper cathode and was observed in each arcing event with a duration of ~10 µs and a current of a few amperes. As for the refractory metals, many droplets were burning, as judged from the character of their glow (Fig. 4a).

The luminosity of burning droplets in the UV region is high and comparable to the luminosity of the cathode spot (Figs. 4b–4d), also indicating that the glow of these objects has a nonthermal nature.

The characteristic feature of arcing shown in Fig. 4a is that the generation of burning droplets has a cyclic character (droplets are generated in layers). This observation is a good illustration of the well-known fact that the cathode processes in vacuum discharge have a cyclic character [1–3]. In some cases, the cyclic character of the processes occurring in the cathode spot manifests itself in the twinkling of the burning droplets (Fig. 4c).[1] Figure 4d illustrates the case where one of

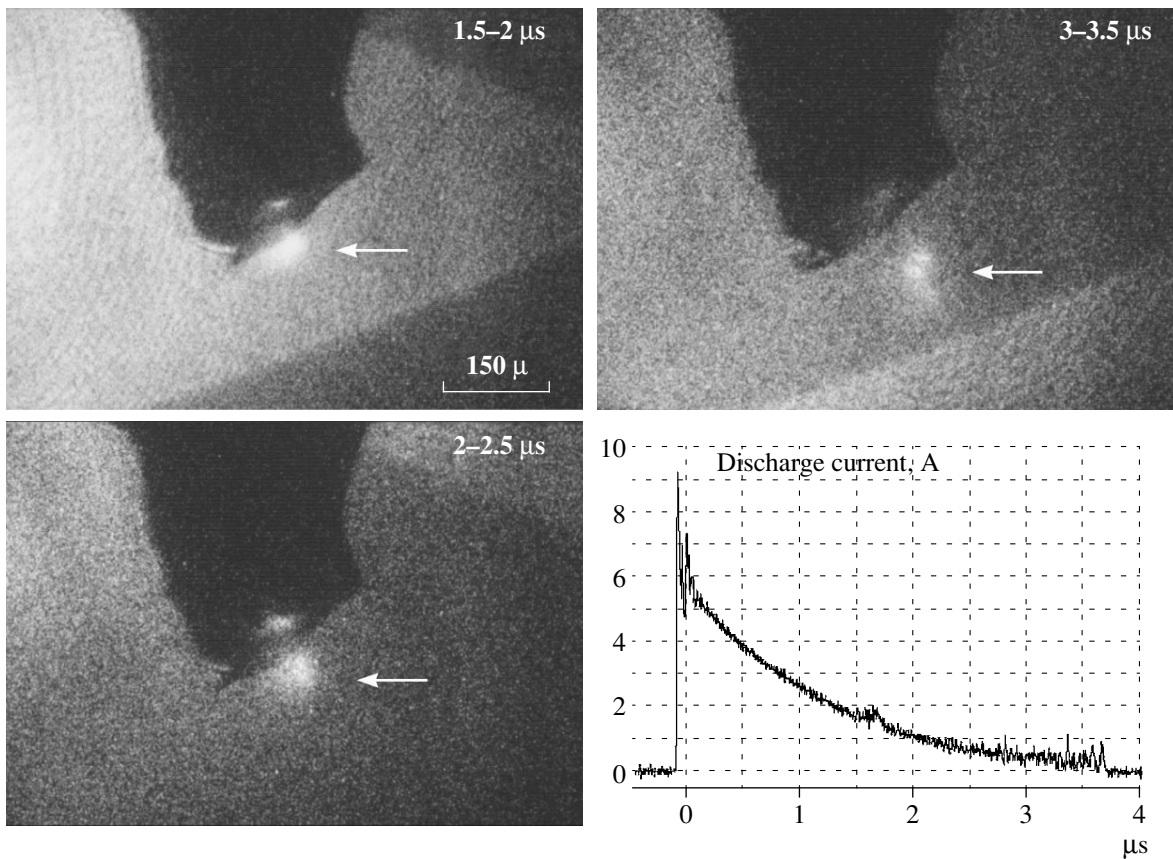_____

[1] In both cases, the cycle period is ~1 µs.

Fig. 2. The formation of a luminous object in arcing on a gallium–indium cathode and the current oscillogram corresponding to this process.

the most distant droplets is ignited. The glow features mentioned above suggest that the bright luminous objects observed in the experiment are dense plasma bunches surrounding the droplets.

With the aim of measuring the parameters of the plasma of burning droplets, additional experiments were carried out using picosecond laser interferometry in conjunction with absorption shadowgraphy. Plasma laser interferometry can be used for measuring the plasma refractive index and, hence, the concentration n_e of the plasma electronic component [9]. Assuming that the plasma bunch has an axially symmetric shape, one can apply the inverse Abelian transform to reproduce the spatial distribution n_e. The method of laser interferometry was implemented using the apparatus and technique described in detail in [10]. The probe radiation wavelength was 532 nm, and the pulse duration was 100 ps. The sensitivity of the apparatus used for measuring the plasma parameters in our experiments allowed the detection of concentrations within 10^{25}–10^{27} cm^{-3} for the plasma electronic component in the plasma bunches with characteristic sizes of 10 μm. Because of the smallness of the objects studied and a high degree of plasma nonideality, only the order of magnitude can be guaranteed for the numerical values of the measured concentrations.

The interference pattern of the cathode spot at the graphite cathode is shown in Fig. 5a. Two irregularities are encircled with the white circle in this interferogram. The first irregularity corresponds to line *1*. The next line deviates only slightly from the rectilinearity, indicating that there is no plasma with the detectable density. Line *3* again shows deflection corresponding to the phase-shifted probe beam. Line *4* is similar to line *2*. The spatial distribution of the plasma electronic component, as obtained by computer processing of the interferogram presented in Fig. 5a, is shown in Fig. 5b. Evidently, the upper plasma bunch corresponds to the cathode spot of vacuum discharge. As to the lower bunch, it represents an independent plasma formation whose n_e value is equal to ~10^{26} m^{-3} and only slightly differs from the n_e values in the cathode spot.

The plasma bunches were observed both in the immediate proximity of the cathode and away from it (up to 100 μm). Such bunches were also detected at the liquid-metal point cathodes. The time delay between the onset of discharge and the appearance of bunches was 0.5–1 μs. The estimated maximal velocity of bunches did not exceed 100 m/s.

Similar bunches were also observed in the absorption shadowgraphs obtained in one experimental run with laser interferometry. An example of an absorption
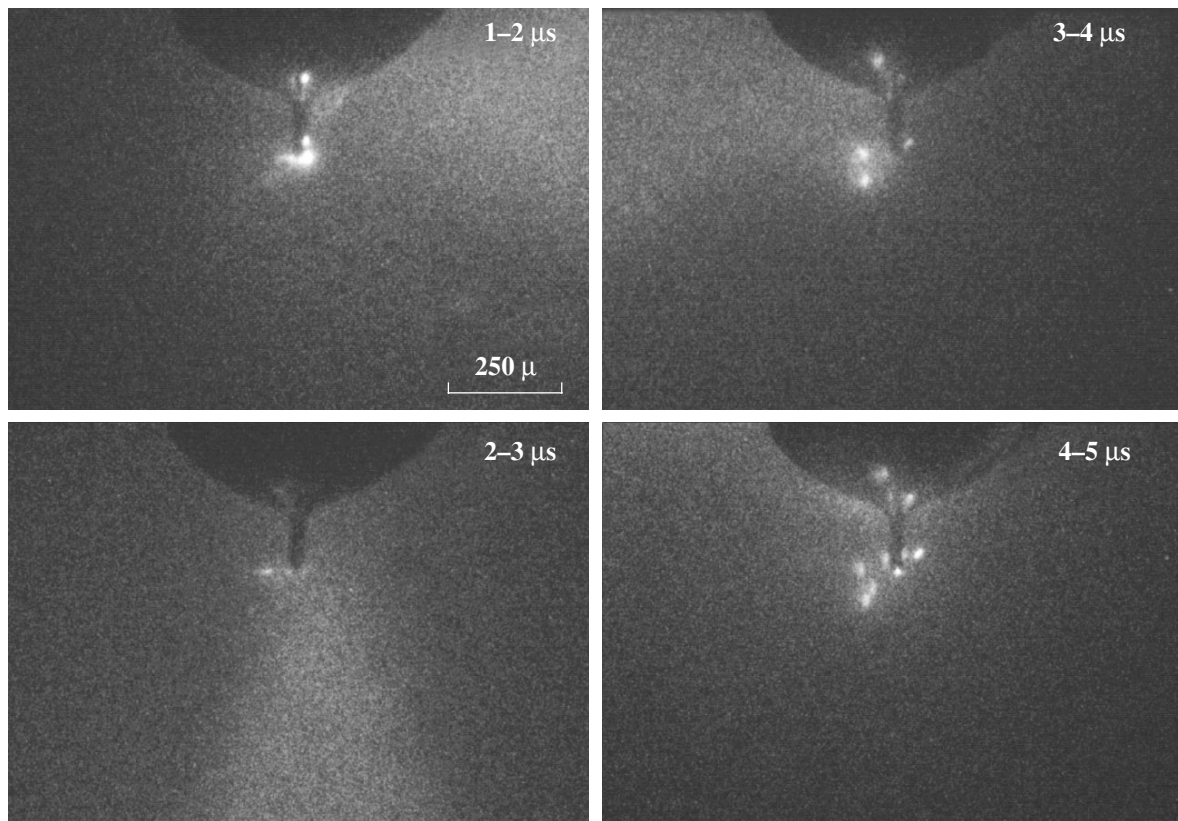
1–2 µs

250 µ

3–4 µs

2–3 µs

4–5 µs

Fig. 3. The formation of plasma bunches near a gallium–indium film cathode.

4–5 µs

150 µ

(a)

7.4–7.8 µs

250 µ

(b)

0–6.5 µs
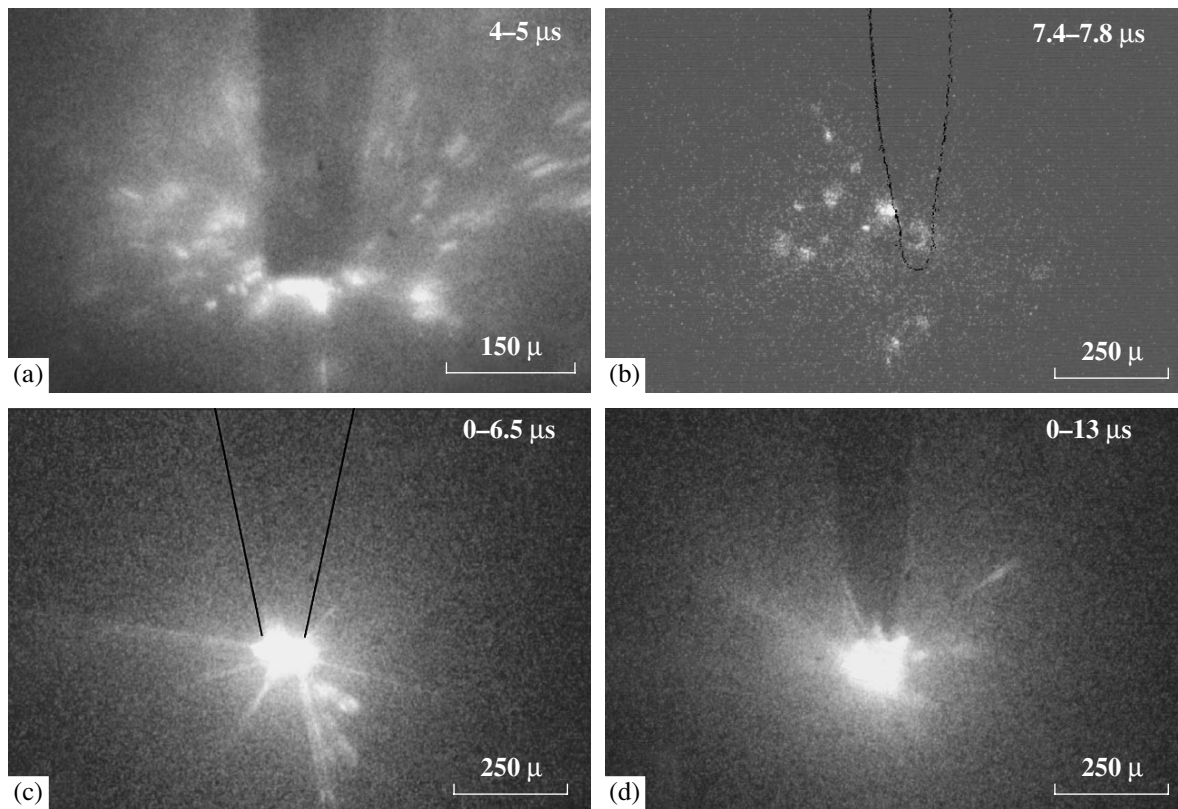
250 µ

(c)

0–13 µs

250 µ

(d)

Fig. 4. Luminous plasma bunches at (a) tantalum and (b–d) tungsten cathodes with arc currents of 2–5 A. Frames (b–d) were obtained using a narrow-band (10 nm) interference filter for a wavelength of 350 nm.
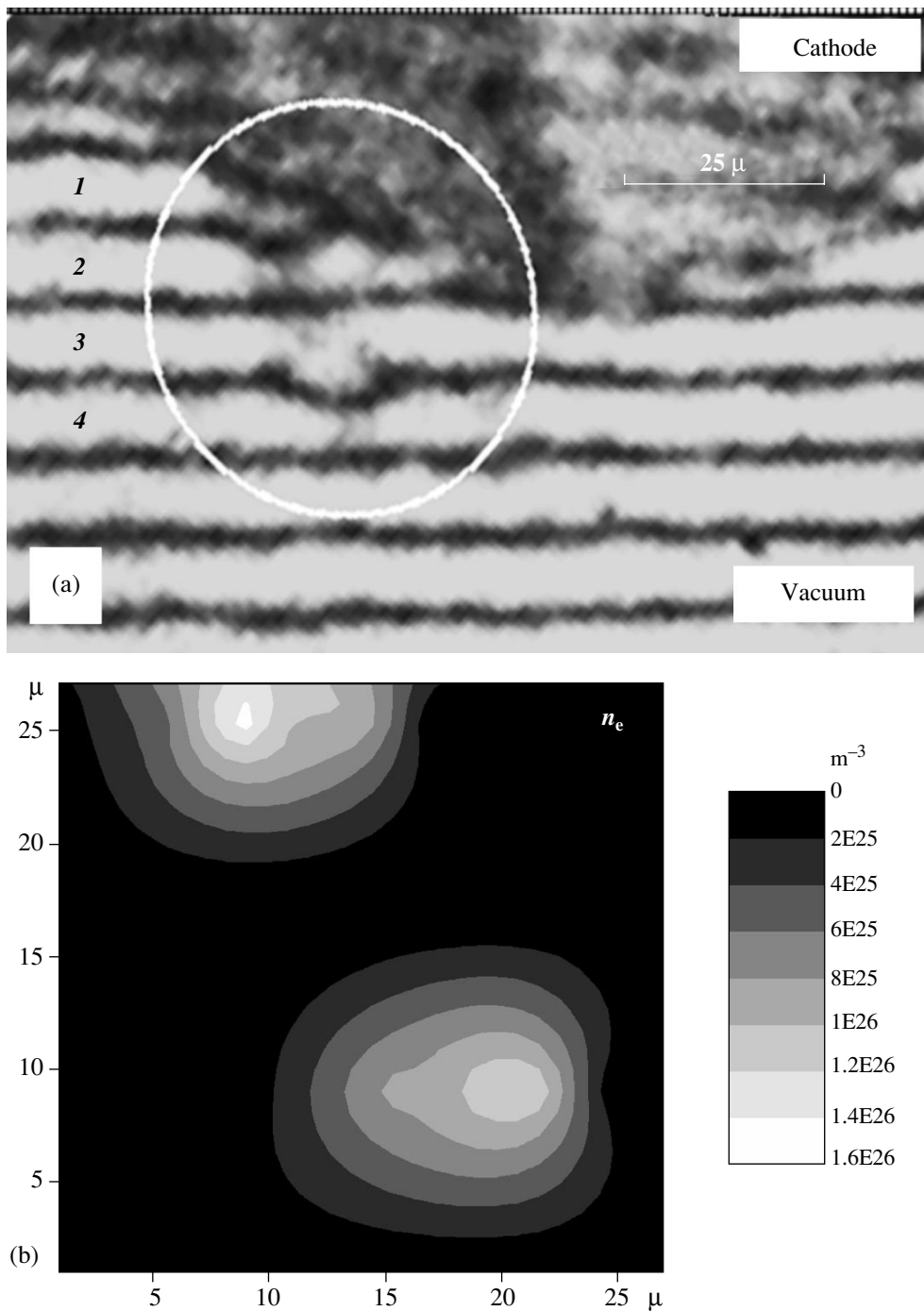
Fig. 5. (a) Interference pattern of the emission center at a graphite cathode with an arc current of 20 A, and (b) spatial distribution of the plasma electronic component for this interferogram.

shadowgram of a plasma bunch emitted from the cathode spot is shown in Fig. 6. One can see that the plasma bunches may contain a dense core with an optical depth of more than 4 at their center, i.e., a practically nontransparent medium. This medium can be interpreted either as a condensed substance of the droplet or as plasma with an above-critical density corresponding to the total internal reflection of the probe radiation.

Dense plasma bunches without cores were also observed in the discharge. However, this does not exclude the presence of several cores with sizes (less than 1 μm) too small to be optically detected. The characteristic sizes, the appearance time, and the estimated maximal velocity of plasma bunches in the absorption experiments were similar to the characteristics of the objects observed in the interferograms. This gave
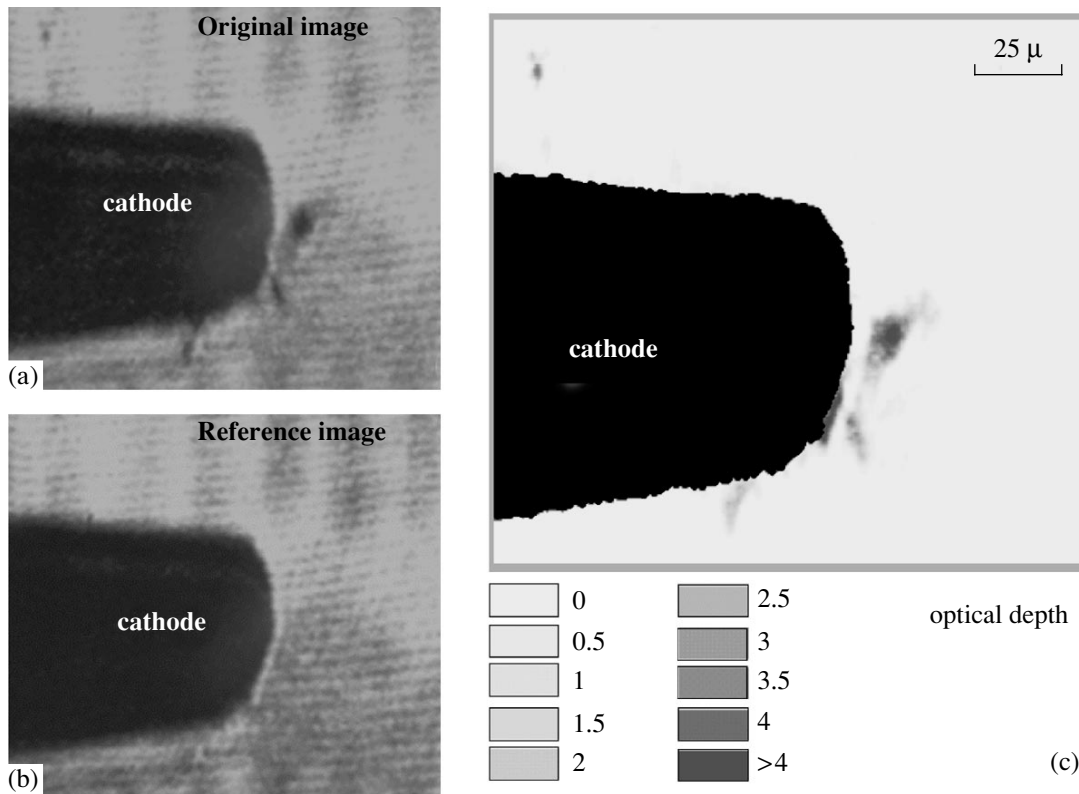
Fig. 6. (a) Absorption image of a gallium–indium cathode with an arc current of 20 A, (b) image of the same cathode without discharge, and (c) their difference.

grounds to assume that in both cases we observed the same objects. In such a situation, knowing the measured absorption coefficients and the n_e values obtained from the interferograms, one can reproduce plasma temperature for both the cathode spot and the dense plasma bunches observed away from it [11]. It was found in all our observations that T_e was on the order of 1 eV both in the cathode spots and in the bunches. This value of T_e is appreciably lower than the value estimated from the charge state using the Saha equation [12–14] and than the value measured by the probe method [15]. At the same time, our results agree well with the data of laser Thomson scattering by the electrons of the cathode-spot plasma (1.4 ± 0.2 eV [16]) and with the data of spectroscopic measurements (0.8–2 eV [17]). This is likely caused by the fact that the optical methods of plasma diagnostics are most sensitive at high plasma concentrations. It is conceivable that plasma Joule heating to a temperature of ~4 eV occurs in the region of lower concentrations far away from the surface.

Therefore, the results of three experimental runs unequivocally indicate that the dense plasma microbunches are formed in the cathode spot of vacuum arc discharge. The motion of these bunches and the presence of a dense condensed core inside them indicate that these bunches appear as a result of intense

plasma formation in the course of evaporation of the droplets emitted by the cathode spot. The values of n_e and T_e for the plasma surrounding the liquid-metal droplets are identical to the parameters of cathode-spot plasma.

The process of formation of dense plasma around droplets immersed in a less dense plasma is considered in [18]. If the droplet is relatively "cold" and its temperature is not sufficiently high for the thermionic emission to occur, it behaves as a "floating" probe in the cathode-spot plasma and rapidly acquires a "floating" negative potential relative to the plasma potential. Under these conditions, the droplet heating by the plasma electrons belonging to the "tail" of Maxwellian distribution is inefficient because of the limited ion current into the droplet. However, if the droplet is "hot" from the outset, it becomes an "emitting" probe because of the thermionic emission from it, and its potential with respect to plasma proves to be positive. This enhances the plasma electron flow into the droplet and, as a result, drastically raises its temperature. The calculations performed in [18] show that the heat carried away through atom evaporation and radiation does not compensate for the heat flow from plasma into the droplet; as a result, the droplet temperature increases up to a value on the order of 5000–7000 K, until the heat

flow into the droplet is counterbalanced by the heat carried away by the electrons emitted from the droplet.

The key role of electron emission from the droplet surface in the mechanism of droplet "ignition" is evident from the fact that the probability of this phenomenon increases substantially with increasing melting point of the cathode material. The fact that the parameters of cathode-spot plasma are identical to the parameters of plasma bunches around the droplets and the intense electron emission from the droplets in the course of plasma formation give grounds to classify these droplet plasma bunches as "droplet spots," by analogy with the cathode spots formed in vacuum discharge at cathodes and anode spots at anodes.

One may assume that the droplet spot shares a number of traits with the cathode spot of a unipolar arc, because in both cases the energy of the plasma surrounding the electrode serves as an energy source feeding this spot. On the other hand, the droplet spot shares some traits with the cathode spot of a cascade arc, because a portion of current is closed to the droplet through the electron emission from plasma into the droplet and the thermionic emission from the droplet. The concept of droplet spots calls for a more detailed investigation into the conditions of their formation and functioning, as well as for a certain refinement of the mechanisms of plasma formation in vacuum discharges. One can expect that, in arcing on the refractory metals, a considerable portion of microdroplets "burns up" in the cathode region. One can also expect that, when moving over the cathode surface, the droplet spots can influence the process of initiation of new emission centers (ectons) and, thereby, the motion of cathode spots on the whole.

REFERENCES

1. E. A. Litvinov, G. A. Mesyats, and D. I. Proskurovskiĭ, Usp. Fiz. Nauk **139**, 265 (1983) [Sov. Phys. Usp. **26**, 138 (1983)].

2. B. Jüttner, Beitr. Plasmaphys. **19**, 25 (1979).

3. G. A. Mesyats, *Ectons in Vacuum Discharge: Breakdown, Spark, Arc* (Nauka, Moscow, 2000).

4. J. E. Daalder, J. Phys. D **9**, 2379 (1976).

5. A. G. Goloveĭko, Izv. Vyssh. Uchebn. Zaved., Énerg. **6**, 83 (1966).

6. T. Schulke and A. Anders, Plasma Sources Sci. Technol. **8**, 567 (1999).

7. T. Utsumi and J. H. English, J. Appl. Phys. **46**, 126 (1975).

8. S. A. Popov, D. I. Proskurovskiĭ, and A. V. Batrakov, IEEE Trans. Plasma Sci. **27**, 851 (1999).

9. L. N. Pyatnitskiĭ, *Laser Plasma Diagnostics* (Atomizdat, Moscow, 1976).

10. N. Vogel, J. Heinzinger, and F. Cichos, IEEE Trans. Plasma Sci. **23**, 926 (1995).

11. N. Vogel and V. Skvortsov, IEEE Trans. Plasma Sci. **25**, 553 (1997).

12. E. Hantzsche, Contrib. Plasma Phys. **19**, 59 (1979).

13. A. Anders, Phys. Rev. E **55**, 969 (1997).

14. R. B. Baksht, A. P. Kudinov, and E. A. Litvinov, Zh. Tekh. Fiz. **43**, 146 (1973) [Sov. Phys. Tech. Phys. **18**, 94 (1973)].

15. V. F. Puchkarev, J. Phys. D **24**, 685 (1991).

16. R. B. Baksht, B. A. Kablambaev, G. T. Razdobarin, and N. A. Ratakhin, Zh. Tekh. Fiz. **49**, 1245 (1979) [Sov. Phys. Tech. Phys. **24**, 689 (1979)].

17. G. A. Lyubimov and V. I. Rakhovskiĭ, Usp. Fiz. Nauk **125**, 665 (1978) [Sov. Phys. Usp. **21**, 693 (1978)].

18. A. V. Kozyrev and A. N. Shishkov (in press).

Translated by V. Sakun

High Intensity of Interband Transitions in Double-Barrier Structures with a High-Frequency Electric Field

E. I. Golant* and A. B. Pashkovskiǐ

State Research and Production Enterprise "Istok," Fryazino, Moscow region, 141120 Russia

*e-mail: eugene@limb.asn.ru

Received December 7, 2001

A solution of a two-band analogue of the time-dependent Schrödinger equation describing resonance coherent electron interaction with a high-frequency field was found for a symmetric double-barrier structure, and an analytic expression was obtained for the small-signal conductivity proportional to the electron transition intensity. It was found that the high-frequency conductivity of double-barrier structures in the case of interband transitions could be significantly higher than in the case of intersubband transitions. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 73.40.-c; 73.21.Fg

Recently, considerable attention has been given to semiconductor heterostructures with coherent (collisionless) electron transport [1, 2]. The mean time of electron escape from an active site in these structures, which commonly represents one or several quantum wells, is significantly shorter than the characteristic time of any of the processes breaking the coherence of the electron wave function. Studying intersubband electron transitions in such structures allowed us to predict a number of new physical effects and to direct ways of their applications to the effective generation of electromagnetic oscillations in the terahertz range [1–3].

However, these studies were related solely to electron transitions inside the conduction band of materials with quadratic dispersion laws. At the same time, the significant progress that has been made recently in the molecular beam epitaxy of semiconductor heterostructures allowed interband electron transitions in quantum wells formed by band discontinuities at heteroboundaries between InAs and GaInSb to be found and used for the generation of oscillations in the IR range [4, 5]. It is natural to suppose that creating conditions for coherent electron transport in such structures may be of both fundamental (from the viewpoint of studying band structure features) and practical interest. Moreover, providing the fulfillment of these conditions for interband transitions may be much easier than for intraband transitions because of less intense phonon scattering and suppressed Auger recombination in structures with interband transitions [4, 5].

The simplest way of describing interband transitions is to use the two-band model [6, 7], which is obtained when the so-called $k\hat{p}$ method of the perturbation theory is used for describing the band structure of a semiconductor. In this work, a two-band model is used in which only the states of the light-hole subband are taken into account in the valence band when interaction between the states of the conduction band and the valence band is calculated. Interaction with other (distant) bands is approximately taken into account by adjusting the free parameter. This model is used for studying the intensity of electron transitions between the levels of a symmetric double-barrier structure, one of whose levels can lie in the conduction band and another, in the valence band of the semiconductor material. The potential barriers of the structure are assumed to be sufficiently thin to provide coherence of electron transport.

The equation for the envelop of the Bloch function of an electron within the two-band model takes the form of the Schrödinger equation for the two-component wave function,

$$i\hbar\frac{\partial}{\partial t}\begin{pmatrix}\psi_e\\\psi_l\end{pmatrix} = \hat{H}_{0R}\begin{pmatrix}\psi_e\\\psi_l\end{pmatrix} + \hat{H}(t)\begin{pmatrix}\psi_e\\\psi_l\end{pmatrix}, \qquad (1)$$

where ψ_e and ψ_l are the corresponding envelops of the wave functions of the states in the conduction band and in the light-hole band, and the Hamiltonian takes the form

$$\hat{H}_{0R} = \begin{pmatrix} E_C + \hat{k}_-\tilde{\zeta}\hat{k}_+ & \hat{k}_-P \\ P\hat{k}_+ & E_V \end{pmatrix},$$

$$\hat{H}(t) = \begin{pmatrix} 2U(x)\cos\omega t & 0 \\ 0 & 2U(x)\cos\omega t \end{pmatrix}. \tag{2}$$
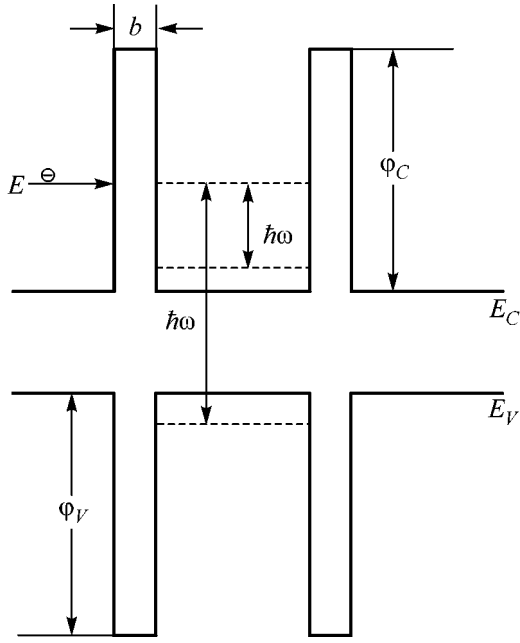
Schematic band diagram of the structure under consideration.

Here, $\hat{k}_+ = \hat{k}_x + i\hat{k}_y$; $\hat{k}_- = \hat{k}_x - i\hat{k}_y$; $\hat{k}_{x,y} = -i\nabla_{x,y}$; P is the matrix element of the momentum operator for the conduction band and the light-hole band; $2U(x)$ is the variable potential; and $\zeta = \hbar^2\tilde{\zeta}/2m_0$, where $\tilde{\zeta}$ is the parameter that takes into account the contribution of the distant bands. It is assumed that the x axis is directed perpendicular to the plane of the potential barriers and the y axis lies in the barrier plane (see figure). In the subsequent discussion, we will consider only those solutions of the given system of equations that represent plane waves in the direction of the y coordinate. Then, the operators can be represented as follows: $\hat{k}_- = \hat{k}_x - ik_y$, $\hat{k}_+ = \hat{k}_x + ik_y$, and $\hat{k}_x = -i\nabla_x$. We will call this model reduced, because the term $\hat{k}_+\tilde{\zeta}_1\hat{k}_-$ is omitted in the $\hat{H}_{22}$ operator. The presence of this term leads to the appearance of nonphysical localized states [8] associated with the incompleteness of the basis set.

In the absence of the variable field ($U(x) = 0$), ψ_l depends linearly on ψ_e, the system in Eq. (1) is degenerate and can be reduced to a scalar time-independent Schrödinger equation,

$$\hat{k}_- D\hat{k}_+\psi_e = (E - E_C)\psi_e,$$

$$D = \frac{\hbar^2}{2m_0}\left(\zeta + \frac{2m_0 P^2}{\hbar^2(E - E_V)}\right) \tag{3}$$

with the dispersion law

$$k^2 = k_x^2 + k_y^2 = \frac{2m_0(E - E_V)(E - E_C)}{\hbar^2}\frac{}{E_P + \zeta(E - E_V)}$$

$$= \frac{2m}{\hbar^2}(E - E_C), \tag{4}$$

$$E_P = \frac{2m_0 P^2}{\hbar^2}, \quad m = m_0\frac{(E - E_V)}{E_P + \zeta(E - E_V)};$$

here, E is the electron energy, and E_C and E_V are the energies of the conduction band bottom and the valence band top, respectively. It should be noted that the two-band effective mass m introduced here can differ from the effective mass of current carriers observed actually.

It will suffice to calculate the high-frequency conductivity of the double-barrier structure under consideration (and the transition intensity proportional to it) within the small-signal (linear with respect to U) approximation. In this case, the first-order correction in the steady-state regime takes the form [1]

$$\psi = \psi_0 e^{-i\omega_0 t} + \psi_+ e^{-i(\omega_0 + \omega)t} + \psi_- e^{-i(\omega_0 - \omega)t}, \tag{5}$$

and the equation for $\psi_\pm$ in the first order of the perturbation theory can be obtained from Eqs. (1) and (2),

$$\hat{H}_{0R}\begin{pmatrix}\psi_e^\pm \\ \psi_l^\pm\end{pmatrix} + \begin{pmatrix}U(x) & 0 \\ 0 & U(x)\end{pmatrix}\begin{pmatrix}\psi_e \\ \psi_l\end{pmatrix}$$

$$= (E \pm \hbar\omega)\begin{pmatrix}\psi_e^\pm \\ \psi_l^\pm\end{pmatrix}. \tag{6}$$

This system, as well as the system of equations without the variable field, is degenerate; therefore, $\psi_l^\pm$ can be written in the form

$$\psi_l^\pm = \frac{P}{E - E_V \pm \hbar\omega}\hat{k}_+\psi_e^\pm + \frac{U(x)\psi_l}{E - E_V \pm \hbar\omega}. \tag{7}$$

Substituting Eq. (7) into the first equation of the system in Eq. (6), we obtain

$$\hat{k}_- D_\pm\hat{k}_+\psi_e^\pm - (E - E_C \pm \hbar\omega)\psi_e^\pm$$

$$+ \frac{\hat{k}_- P}{E - E_V \pm \hbar\omega}\left\{\frac{U(x)P}{E - E_V}\hat{k}_+\psi_e\right\} + U(x)\psi_e = 0, \tag{8}$$

where

$$D_\pm = \frac{\hbar^2}{2m_0}\left(\zeta + \frac{2m_0 P^2}{\hbar^2(E - E_V \pm \hbar\omega)}\right). \tag{9}$$

In order to find wave functions, it is necessary to specify the forms of the barriers and the perturbing field.

Double-barrier structure. Even in the single-band model, electron transitions in double- and triple-barrier nanostructures are amenable to the complete analytical treatment only within the thin (δ-like) barrier approximation [1–3]. In this case, a barrier with height φ and thickness b is replaced by a δ-like barrier with the power $\alpha = \varphi b$, and the wave-function matching condition at the barrier takes the form [1, 9]

$$\psi'(+0) - \psi'(-0) = \frac{2m^*\alpha}{\hbar^2}\psi(0), \qquad (10)$$

where m^* is the electron effective mass. The analogous approximation can also be used in the two-band model.

Consider the first barrier of the heterostructure shown in the figure. Let φ_C be the barrier height in the conduction band, and φ_V, the barrier height in the valence band. It can be shown that for thin rectangular barriers

$$\psi'(+0) - \psi'(-0)$$
$$\approx \left[\frac{2m_0}{\hbar^2}\frac{(E-E_V)}{E_P+\zeta(E-E_V+\varphi_V)}\varphi_C b\right]\psi(0) = g\psi(0),$$
$$E > E_c, \qquad (11)$$

$$\psi(+0) - \psi(-0) \approx \frac{\varphi_V b}{E-E_V}\psi'(0), \quad E < E_V,$$

where the expression in the square brackets designated in Eq. (11) as g is an analogue of the product of barrier power by the effective electron mass $2m^*\alpha/\hbar^2$ in Eq. (10). As in the case of a single-band δ-like barrier, it is assumed that $g \longrightarrow$ const at $b \longrightarrow 0$. In the two-band model, this asymptotic behavior may correspond to a δ-like barrier in either the conduction band or the valence band.

Furthermore, for the case $k_y = 0$, consider a flow of electrons with energy $\varepsilon = E - E_C$ incident from the left on a symmetric double-barrier structure with the quantum-well width a shown in the figure. A uniform high-frequency electric field $U(x) = -qFx$ with the strength F varying with time according to the equation $2F\cos\omega t = F(e^{i\omega t} + e^{-i\omega t})$ is applied to the structure.

The unperturbed electron component, the electron wave function ψ_0 (for simplicity, index e is omitted from here on) normalized to the unit amplitude of the wave incident from the left, takes the form

$$\psi_0(x) = \begin{cases} \exp(ikx) + D_0\exp(-ikx), & x < 0 \\ A_0\sin(kx) + B_0\cos(kx), & 0 < x < a \\ C_0\exp(ik(x-a)), & x > a. \end{cases} \quad (12)$$

In the small-signal approximation, the first-order correction ψ_1 to the ground-state wave function equals [1] $\psi_1 = \psi_+(x)e^{-i(\omega_0+\omega)t} + \psi_-(x)e^{-i(\omega_0-\omega)t}$ ($\omega_0 = \varepsilon/\hbar$). In this case,

$$\psi_\pm(x) = \begin{cases} D_\pm\exp(-ik_\pm x), & x < 0 \\ A_\pm\sin(k_\pm x) + B_\pm\cos(k_\pm x) + \chi_\pm(x), \\ 0 < x < a \\ C_\pm\exp(ik_\pm(x-a)) + P_\pm\exp(ik(x-a)), \\ x > a, \end{cases} \quad (13)$$

where

$$k_\pm^2 = \frac{2m_0}{\hbar^2}\frac{(E-E_V\pm\hbar\omega)(E-E_C\pm\hbar\omega)}{E_P+\zeta(E-E_V\pm\hbar\omega)} - k_y^2,$$
$$P_\pm = \mp\frac{qFa}{\hbar\omega}\psi_0(a), \qquad (14)$$

and $\chi_\pm$ is a particular solution of Eq. (8), which is sought inside the structure in the form

$$\chi_\pm = \beta_\pm x\psi_e + \gamma_\pm\psi_e'. \qquad (15)$$

After substituting Eq. (15) in Eq. (8), we obtain

$$\beta_\pm = \mp\frac{qF}{\hbar\omega},$$

$$\gamma_\pm = \frac{qF}{2m\omega^2}$$
$$\times \frac{E_p(2E-2E_V\pm\hbar\omega)+2\zeta(E-E_V)(E-E_V\pm\hbar\omega)}{E_p(E-E_V+\varepsilon\pm\hbar\omega)+\zeta(E-E_V)(E-E_V\pm\hbar\omega)}, \qquad (16)$$

$$m_\pm = m_0\frac{(E-E_V\pm\hbar\omega)}{E_p+\zeta(E-E_V\pm\hbar\omega)}.$$

Here q is the electron charge, and $m_\pm$ is determined similarly to Eq. (4).

The system of equations for determining coefficients $A_\pm$, $B_\pm$, $C_\pm$, and $D_\pm$ is written in the matrix form as

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ ik_\pm - g & k_\pm & 0 & 0 \\ 0 & \sin k_\pm a & \cos k_\pm a & -1 \\ 0 & -k_\pm\cos k_\pm a & k_\pm\sin k_\pm a & ik_\pm - g \end{pmatrix}$$
$$\times \begin{pmatrix} D_\pm \\ A_\pm \\ B_\pm \\ C_\pm \end{pmatrix} = \begin{pmatrix} \chi_\pm(0) \\ -\chi_\pm'(a) \\ P_\pm - \chi_\pm(0) \\ (g-ik)P_\pm - \chi_\pm'(a) \end{pmatrix}. \qquad (17)$$

Let the energy of incident electrons correspond to the energy level N of the quantum well (the numbering of levels starts from unity), and let the frequency of the electric field correspond to transitions to the level with the number L. Here, in accordance with the parity selection rules, $N - L$ is an even number. The condition for the resonance electron tunneling, which coincides with the condition for the resonance interaction of electrons with the HF field (the determinant of system (17) is minimum), takes the form [9]

$$\tan(k_\pm a) = -2k_\pm/g. \tag{18}$$

For sufficiently powerful barriers

$$g \gg k, k_\pm, \tag{19}$$

the system in Eq. (17) can be solved analytically,

$$C_\pm \approx D_\pm \approx B_\pm \approx 2\gamma_\pm \frac{g^2}{ik_\pm}, \quad A_\pm \approx B_\pm \frac{g}{k_\pm}, \tag{20}$$

which gives the following equation for the high-frequency conductivity of the structure for the electron concentration n in the flow incident on the structure:

$$\sigma_e = \pm \frac{q^2 \hbar^2 g^4 n}{\pi L \omega^3 m^3} \left(1 - \frac{1}{2} \right.$$

$$\times \left. \frac{E_p(2\varepsilon \pm \hbar\omega)}{E_p(E - E_V + \varepsilon \pm \hbar\omega) + \zeta(E - E_V)(E - E_V \pm \hbar\omega)} \right)^2$$

$$\times \frac{m}{|m_\pm|} G_n, \tag{21}$$

$$G_n = \left[1 + \frac{\varepsilon}{E - E_V} \frac{E_p}{E_p + \zeta(E - E_V)} \right]^{-1}.$$

Here, G_n is the factor that provides the normalization of the squared two-component wave function to unity. At low energies $\varepsilon \ll E_g$, $E_g = E_C - E_V$, and Eq. (21) for transitions inside one band is transformed to the equation obtained in [1] for a double-barrier structure with a parabolic dispersion law.

At $\zeta = 0$ (the effective electron mass is equal to the light-hole mass), Eq. (21) is reduced to

$$\sigma_e = \pm \frac{q^2 \hbar^2 g^4 n}{\pi L \omega^3 m^3} \left(1 - \frac{2\varepsilon \pm \hbar\omega}{2(E - E_V + \varepsilon \pm \hbar\omega)} \right)^2$$

$$\times \frac{m}{|m_\pm|} \left[1 + \frac{\varepsilon}{E - E_V} \right]^{-1} = \pm \frac{q^2 \hbar^2 g^4 n}{\pi L \omega^3 m^3} R. \tag{22}$$

For transitions within one band and not very high electron energies, the R factor only slightly differs from unity (at $\varepsilon \le E_g$, $0.3 \le R \le 1$). However, the transitions between levels located within different bands can give a rather high conductivity σ_p of the structure. In particular, in a sufficiently wide quantum well in which levels are closely spaced between each other at $\hbar\omega \approx E_g$ and $\varepsilon \ll E_g$,

$$\frac{\sigma_e}{\sigma_p} \approx \left(\frac{\varphi_C}{\varphi_V} \right) \frac{L^5}{\pi^2 (N^2 - L^2)^3}. \tag{23}$$

Thus, as distinct from the intraband resonance transitions where the conductivity increases singularly only if both the electron energy and the field quantum energy tend to zero, the conductivity of a quantum well confined by thin barriers with coherent electron transport in the case of interband transitions increases singularly when the field quantum energy tends to the forbidden band width. The effect found in this work can be used for designing quantum cascade lasers.

REFERENCES

1. A. B. Pashkovskiĭ, Zh. Éksp. Teor. Fiz. **109**, 1779 (1996) [JETP **82**, 959 (1996)].

2. E. I. Golant and A. B. Pashkovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 559 (1996) [JETP Lett. **63**, 590 (1996)].

3. E. I. Golant and A. B. Pashkovskiĭ, Zh. Éksp. Teor. Fiz. **112**, 237 (1997) [JETP **85**, 130 (1997)].

4. Rui Q. Yang, B. H. Yang, D. Zhaang, *et al.*, Appl. Phys. Lett. **71**, 2409 (1997).

5. J. L. Bradshaw, Rui Q. Yang, J. D. Bruno, *et al.*, Appl. Phys. Lett. **75**, 2362 (1999).

6. E. O. Kane, in *Semiconductors and Semimetals*, Ed. by R. K. Willardson and A. C. Beer (Academic, New York, 1966), Vol. 1, Chap. 3.

7. C. Sirtory, F. Capasso, and J. Faist, Phys. Rev. B **50**, 8663 (1994).

8. M. F. H. Schuurmans and G. W't Hooft, Phys. Rev. B **31**, 8041 (1985).

9. V. M. Galitskiĭ, B. M. Karnakov, and V. I. Kogan, *Problems in Quantum Mechanics* (Nauka, Moscow, 1981), p. 173.

Translated by A. Bagatur'yants

Ultrafast Structural Transformations in Graphite[1]

S. I. Ashitkov[1], M. B. Agranat[1]*, P. S. Kondratenko[1], S. I. Anisimov[1], V. E. Fortov[1],
V. V. Temnov[2], K. Sokolowski-Tinten[2], P. Zhou[2], and D. von der Linde[2]

[1] *Institute of High Energy Density, Joint Institute for High Temperatures, Russian Academy of Sciences,*
Moscow, 127412 Russia
* e-mail: utpr@iht.mpei.ac.ru

[2] *Institute for Laser and Plasma Physics, University of Essen, 45117 Essen, Germany*

Received December 11, 2001

Femtosecond laser-induced structural transitions in graphite were studied by time-resolved optical anisotropy measurements. The decay of the anisotropic reflectivity seems to indicate a loss of long-range order on a sub-picosecond time scale, which is much faster than the electron–phonon relaxation time. This observation confirms the nonthermal nature of the structural transition. © *2002 MAIK "Nauka/Interperiodica".*

PACS numbers: 64.70.Kb; 61.80.Ba; 61.50.Ks

Ultrafast light-induced phase transitions have been studied in a number of covalently bonded materials and using many different experimental techniques, including time-resolved measurements of the reflectivity [1–4] and of the reflected second harmonic [2, 5, 6], as well as by molecular dynamics simulations [7]. More recently, the direct observation of the disordering of a solid during melting has become possible through the use of ultrafast time-resolved X-ray diffraction [8–10]. While reflectivity measurements provide information only about the changes of the electronic properties, which cannot be uniquely related to transient structural changes, time-resolved X-ray diffraction has not yet been applied to the particular case of femtosecond-excited graphite.

In this letter, we present the results of femtosecond time-resolved optical anisotropy measurements of ultrafast order–disorder transitions in single crystalline graphite after excitation by femtosecond laser pulses. The novel time-resolved optical anisotropy technique developed in [11] was successfully applied to study the melting dynamics of superficial layers of zinc and graphite heated by picosecond laser pulses [12]. The idea of the method can be described as follows. When the *p*- or *s*-polarized electromagnetic wave is incident on a surface of an optically isotropic medium, the reflected wave has the same polarization as the incident wave. However, if the medium is optically anisotropic, the reflected wave, in general, also contains a component perpendicular to the polarization of the incident wave. The latter component disappears if the crystal is converted to an isotropic state as a result of a phase transition [11, 12].

Optical anisotropy in uniaxial crystals most strongly manifests itself in the reflection from planes containing the anisotropy axis (**C**); this situation is realized in our experiments. A *p*-polarized wave with intensity I_{in} normally incident on the surface containing the anisotropy axis possesses both *s*- and *p*-polarized components after reflection. Their intensities I_s and I_p can easily be obtained using the Fresnel formula:

$$I_s \equiv R_{ps} I_{in} = I_{in} \left| \frac{2\delta}{(n+1)^2 - \delta^2} \right|^2 \sin^2 2\varphi,$$

$$I_p \equiv R_{pp} I_{in} = I_{in} \left| \frac{n^2 - \delta^2 - 1 + 2\delta \cos 2\varphi}{(n+1)^2 - \delta^2} \right|^2. \quad (1)$$

Here, $n = (n_o + n_e)/2$ and $\delta = (n_e - n_o)/2$ are, respectively, the isotropic and anisotropic parts of the complex refractive index, with n_e and n_o being the complex refractive indices for the extraordinary and ordinary waves, respectively; φ is the angle between the polarization plane and the anisotropy axis. Only a small fraction of the incident intensity is converted into the *s*-component: $R_{ps} \propto 10^{-2} – 10^{-3}$ [11, 12], which is maximal at $\varphi = 45°$. Being the ordinary reflectivity for $\varphi = 90°$ and extraordinary reflectivity for $\varphi = 0°$, the reflectivity of the parallel *p* component remains quite high: $R_{pp} \propto 1$. The large difference in the intensities of the *p* and *s* components immediately gives $|\delta| \ll |n^2 - 1|$ and allows one to neglect in Eq. (1) the small terms proportional to δ^2. Equations (1) for the particular case $\varphi = 45°$ have a simple form:

$$R_{ps} \approx \frac{|2\delta|^2}{|n+1|^4}, \quad R_{pp} \approx \left| \frac{n-1}{n+1} \right|^2,$$

$$\text{for } R_{ps} \ll R_{pp} \propto 1, \quad \varphi = 45°. \quad (2)$$

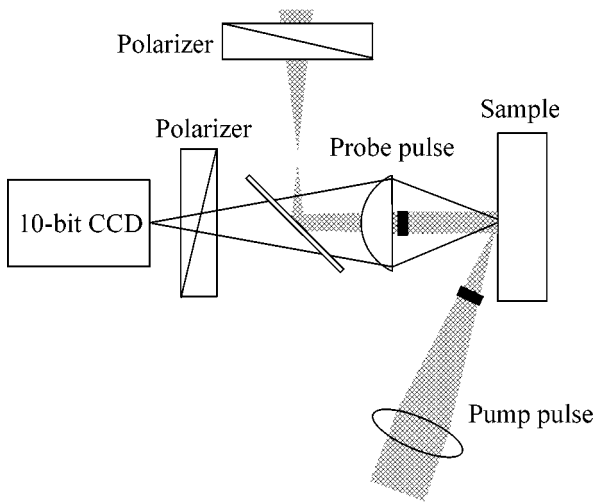[1] This article was submitted by the authors in English.

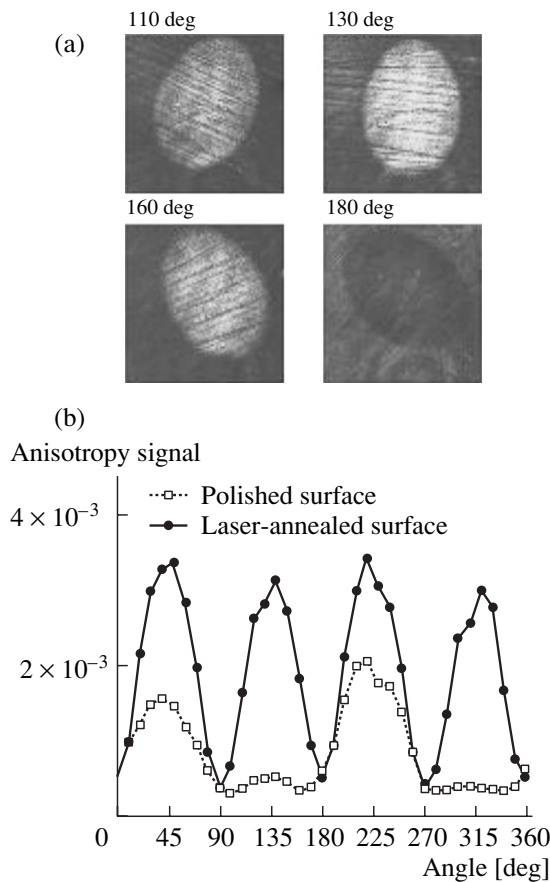**Fig. 1.** Schematic of the experimental setup.

Fig. 2. (a) Anisotropic (*s*-polarized) reflectivity images of the laser-annealed area at different angles $\varphi = 110°$, $130°$, $160°$, and $180°$; the elliptical ablation crater can be easily recognized. (b) The angular dependence of anisotropic reflectivity of the polished and laser-annealed surface.

Note that Eqs. (2) are different from those derived in [11] for metallic surfaces in the impedance approximation $|\delta| \ll |n|$, $|n| \gg 1$. It follows from Eq. (2) that the intensity of the *s* wave vanishes when the material looses order or a crystal becomes isotropic corresponding to $\delta \longrightarrow 0$.

The experimental results were obtained by time-resolved polarization microscopy; the schematic of the setup is presented in Fig. 1. A *p*-polarized pump pulse ($\tau = 100$ fs, $\lambda = 800$ nm) at an angle of incidence of approximately 45° was used to excite the surface of a single-crystal graphite sample. The excited surface area was illuminated by a weak time-delayed *p*-polarized probe pulse (second harmonic, $\tau = 100$ fs, $\lambda = 400$ nm) normally incident on the surface through a high-resolution microscope objective (20X, NA = 0.3). A CCD camera located in the image plane of the microscope objective was used to acquire the transient surface reflectivity images formed by the reflected probe pulse. By adjusting a polarizer placed in front of a CCD camera, the "isotropic" (*p*-polarized) or "anisotropic" (*s*-polarized) reflectivity component can be detected.

Perfect crystalline surfaces parallel to the basal graphite plane (perpendicular to the optical axis) can easily be obtained by removal of the upper layers [4]. In contrast, the preparation of a surface containing the anisotropy axis and having a perfect crystalline structure in a thin surface layer with a thickness on the order of the skin depth represents a serious problem. An effective way to prepare a perfect crystalline surface is based on "laser annealing" [6]: removal of the nonperfect upper layers of the material by laser ablation. In Fig. 2, we compare the angular dependence of the anisotropy signal $R_{ps}(\varphi)$ from the initially polished graphite surface with that from the annealed surface area (annealing pulse: $\tau = 100$ fs, $\lambda = 800$ nm, $F = 0.6$ J/cm²).

The anisotropic reflectivity images are taken at different angles φ between the optical axis and the polarization vector by rotating the sample; the annealed area is clearly seen in the center of each image (Fig. 2a). The visible scratches with a typical depth of 1 μm originate from the polishing procedure and cannot be eliminated by laser annealing, since only a few tens of nanometers are removed by laser ablation. Nevertheless, laser annealing does significantly improve the crystalline structure of the surface, which is shown in Fig. 2b. The anisotropy signal from the initial surface is relatively weak and contains only two maxima. In contrast, the angular dependence measured in the center of the annealed area shows four distinct maxima of equal amplitudes, which is in good agreement with the expected $\sin^2(2\varphi)$ angular dependence [see Eq. (2)]. The maximal value of the anisotropy signal $R_{ps}^{max} \approx 4 \times 10^{-3}$ is consistent with the value 5.5×10^{-3} obtained from Eq. (2) by assuming $n_o = 2.62 - 1.28i$ and $n_e = 2.04 - 0.62i$ [13]. The nonzero minimal value of the
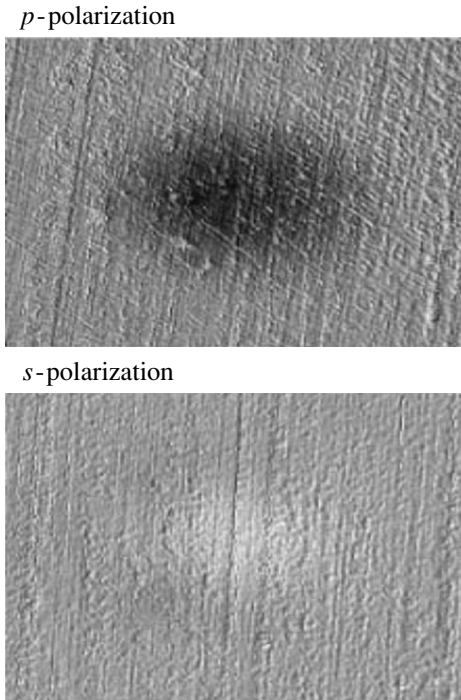
p-polarization

s-polarization

Fig. 3. Time-resolved reflectivity images for *s* and *p* polarizations; laser fluence 0.35 J/cm^2, Δt = 700 fs, and frame size 150 × 120 μ.
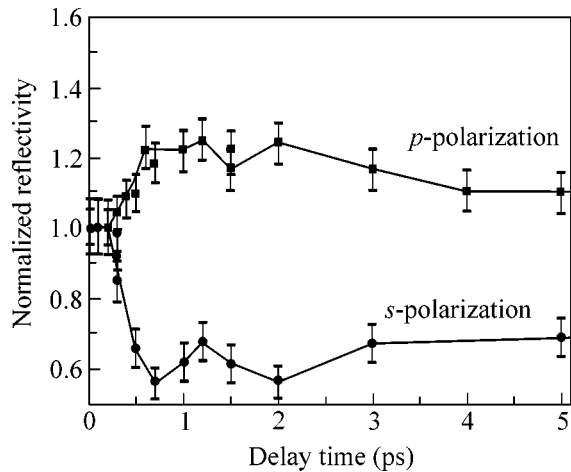

Fig. 4. Temporal evolution of the isotropic (*p*-polarized) and anisotropic (*s*-polarized) reflectivity; laser fluence 0.35 J/cm^2.

tion by Nomarski (differential interference contrast) optical microscopy also indicates an improved surface morphology of the laser-annealed area, as compared to the as-polished surface.

For time-resolved experiments, the pump pulse with maximal fluence 0.35 J/cm^2 was focused inside the laser-annealed surface area. The orientation of the sample was set to maintain the maximal initial anisotropy signal at $\varphi = 45°$. An example of transient anisotropy measurements is shown in Fig. 3, where the normalized isotropic and anisotropic reflectivity images for the delay time $\Delta t = 0.7$ ps are presented. The normalization procedure implies the division of the transient reflectivity image by the initial reflectivity image of unexcited surface to eliminate the influence of inhomogeneous probe illumination and initial surface defects. The small increase in an isotropic (*p*-polarized) and significant decrease of the anisotropic (*s*-polarized) reflectivity can be inferred from the reflectivity images in Fig. 3. The temporal evolution of normalized isotropic and anisotropic reflectivity components is summarized in Fig. 4. Within a few hundreds of femtoseconds after laser excitation, an increase in the isotropic *p* component is observed, similar to the results reported by Downer *et al.* [4] for the experiments performed on the basal plane of graphite. In contrast, the anisotropic *s* component drops very sharply within 300–500 fs. This behavior indicates a very rapid loss of crystalline order, much faster than the electron–phonon relaxation time, which is on the order of 1–2 ps [4]. Thus, the destruction of crystalline order is caused by a nonthermal mechanism due to the strong electronic excitation of the material.

To summarize, we have extended time-resolved optical anisotropy measurements [11, 12] to the femtosecond time domain and applied this technique to detect an ultrafast structural transformation in *fs*-laser excited graphite. Laser annealing using *fs*-laser ablation proved to be a robust technique for the preparation of well-ordered surfaces containing the anisotropy axis. The first time-resolved measurements indicate the disappearance of crystalline order on a subpicosecond time scale and thus emphasize the nonthermal nature of the structural transition. A slight modification of the experimental geometry will allow one to follow the temporal evolution of both ordinary and extraordinary refractive indices. This experiment is expected to provide more insight into the physical mechanisms of the structural transformation in graphite and is currently in progress.

anisotropy signal $R_{ps}^{\min} \approx 0.4 \times 10^{-3}$ is determined from the extinction ratio of the polarizers in the current geometry of the experiment. A detailed surface inspec-

REFERENCES

1. C. V. Shank, R. Yen, and C. Hirlimann, Phys. Rev. Lett. **50**, 454 (1983).

2. K. Sokolowski-Tinten, J. Bialkowski, and D. von der Linde, Phys. Rev. B **51**, 14186 (1995).

3. L. Huang, J. P. Callan, E. N. Glezer, and E. Mazur, Phys. Rev. Lett. **80**, 185 (1998).

4. D. H. Reitze, H. Ahn, and M. C. Downer, Phys. Rev. B **45**, 2677 (1992).

5. C. V. Shank, R. Yen, and C. Hirlimann, Phys. Rev. Lett. **51**, 900 (1983).

6. H. W. K. Tom, G. D. Aumiller, and C. H. Brito-Cruz, Phys. Rev. Lett. **60**, 1438 (1988).

7. P. L. Silvestrelli and V. Parrinello, J. Appl. Phys. **83**, 2478 (1998).

8. C. W. Siders, A. Cavalleri, K. Sokolowski-Tinten, *et al.*, Science **286**, 1340 (1999).

9. A. Rousse, C. Rischel, S. Fourmaux, *et al.*, Nature **410**, 65 (2001).

10. K. Sokolowski-Tinten, C. Blome, C. Dietrich, *et al.*, Phys. Rev. Lett. **87**, 225 701 (2001).

11. M. B. Agranat, S. I. Anisimov, V. E. Fortov, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 2162 (1998) [JETP **86**, 1184 (1998)].

12. M. B. Agranat, S. I. Ashitkov, V. E. Fortov, *et al.*, Appl. Phys. A **69**, 637 (1999).

13. E. D. Palik, *Handbook of Optical Constants of Solids II* (Academic, San Diego, 1991).

Spontaneous Formation of a System of Highly Ordered Germanium Nanoislands upon Epitaxial Deposition on Profiled (111) Silicon Substrates under Electromigration Conditions

I. V. Zakurdaev*, S. Yu. Sadofyev*, and A. O. Pogosov**

* Ryazan State Radioengineering Academy, Ryazan, 391000 Russia
** Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia
Received December 11, 2001

Atomic-force microscopy was used to study the surface topography of SiGe structures grown by epitaxial deposition of Ge on profiled Si(111) substrates under electromigration conditions. Systems of highly ordered germanium nanosized islands with dimensions of 10–20 nm and a density of 6×10^{10} cm^{-2} were obtained. It is shown that the geometrical parameters of self-organizing nanoislands can be controlled by a proper choice of the growth and postgrowth annealing conditions for these structures. © *2002 MAIK "Nauka/Interperiodica".*

PACS numbers: 68.65.-k; 61.46.+w; 68.37.Ps; 81.15.Hi

In the last 5–10 years, the constant growth of researchers' interest in the heteroepitaxial system Ge$_x$Si$_{1-x}$/Si has been observed because of the high technological potential of self-organizing SiGe islands ("quantum dots") for various device applications. In spite of the large number of experimental works devoted to the investigation of the formation and evolution of SiGe islands on substrates of various orientations, the problem of creating structures with quantum dots as the elemental base for a new generation of devices is now far from being completely solved. The main difficulties in the production of high-quality structures with quantum dots are associated with the large dimensions of the self-organizing germanium islands, which prevents the manifestation of quantum properties in such systems, and with the random character of the localization of these islands in the film–substrate heterojunction plane. In this connection, efforts directed at the investigation of the possibilities of additionally affecting the mechanism of self-organization of the surface and the development of the technological elements of the production of such structures are very important. In this paper, we present some preliminary results of the investigation into the process of spontaneous formation of germanium nanoislands upon epitaxial deposition on Si(111) profiled substrates under conditions of electromigration of germanium adatoms.

The samples to be studied were obtained by molecular-beam epitaxy in an experimental ultrahigh-vacuum system whose basic distinction from standard devices of molecular-beam epitaxy is in the possibility of annealing samples by both indirectly heating them using an external source of heat and by directly passing ac or dc current of a given polarity through the crystal. The system is equipped with a standard molecular source of germanium, which represents a crucible made of boron nitride surrounded with a tantalum heater and a system of protective thermal shields. As the source of silicon, an additional silicon crystal located at a distance of 15 mm over the surface of the substrate was used. To rapidly chop the molecular beam, a tantalum bellow-controlled shutter located in the immediate proximity to the source was employed.

As substrates, KÉS-0.1 silicon single crystals oriented 4.4° off the (111) plane, which were cleaved out in the rectangular form with dimensions of $0.3 \times 5 \times 15$ mm along the [$1\bar{2}1$] direction, were used. The temperature of the samples during annealing was controlled with an optical pyrometer or, at low temperatures, with a chromel–alumel thermocouple.

The sample was placed into the vacuum chamber and, after the maximum vacuum had been achieved ($p = 3 \times 10^{-10}$ torr), the surface of the substrate was cleaned from natural oxides and inclusions of silicon carbide using a flashing to 1300°C for 1.5 min by directly passing alternating current through the sample. Then, the substrate temperature was lowered to 400°C and a buffer layer of silicon 50 nm thick was applied. After this, the crystal was annealed at a temperature of 1250°C for 0.5–2 min by passing direct current to produce the desired surface profile in the form of terraces and nanosteps [1]. Then, by decreasing the density of the direct current passed through the sample, the temperature of the substrate was decreased to 550–600°C, and a germanium film 5 to 15 monolayers (ML) thick was deposited and subjected to postgrowth annealing
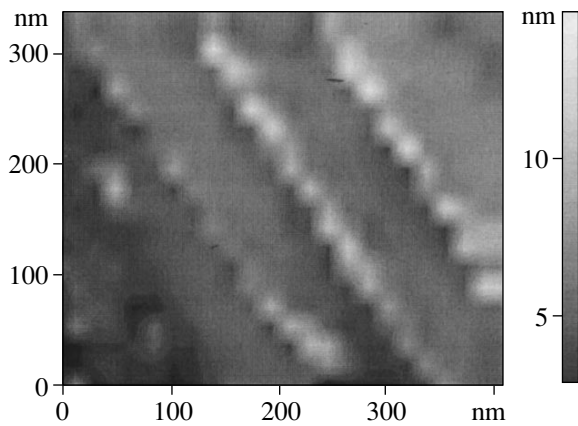
Fig. 1. System of highly ordered Ge nanoislands on a profiled Si(111) surface. The effective thickness of the germanium film is 7 ML; the postgrowth annealing of the sample is performed by passing direct current for 10 min at a temperature of 600°C.
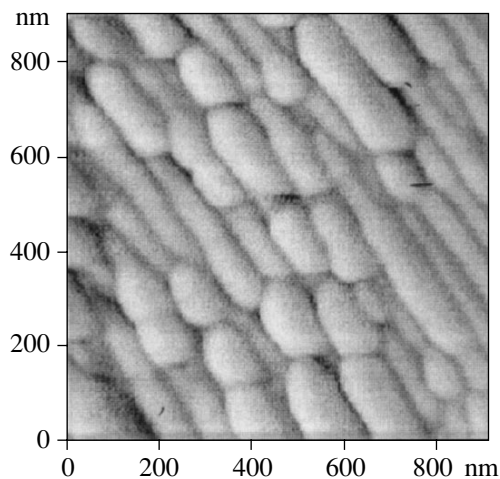

Fig. 2. Decomposition of an elastically strained germanium film deposited on a profiled substrate into isolated islands. The effective thickness of the germanium film is 12 ML; the postgrowth annealing of the sample is performed by passing direct current for 3 min at a temperature of 550°C.

for several minutes without changing the direction and intensity of the current. The rate of deposition of germanium was 0.05 nm/s. The investigation of the surface topography of the structures was performed using a Solver P47-SPM MDT atomic-force microscope.

The above heat treatment of the substrate resulted in the formation of a system of nanosteps with periodically located kinks, which served upon subsequent deposition of germanium as the sites of pinning of adatoms, initiating the development of a nanoisland. The process of electrotransfer due to the passage of a dc

electric current during the deposition of the film was an additional factor stimulating the motion of germanium adatoms toward edges of nanosteps and pinning them along the fronts of moving nanosteps on the substrate. The use of this technique permitted us to obtain arrays of highly ordered nanoislands of germanium 10–20 nm in diameter, whose density was 2–6×10^{10} cm^{-2}. The surface topography of typical structures with highly ordered germanium nanoislands observed in an atomic-force microscope is displayed in Fig. 1. Note that the spatial characteristics of the produced islands are determined by both the dimensions of the steps on the substrate and the effective thickness of the germanium film and the duration of the postgrowth annealing of the sample. By varying these growth parameters, we could observe various stages of the rearrangement of the surface, beginning from the breaking up of the film into isolated islands and ending by the formation of large clusters due to the coalescence of nanoislands. Figure 2 illustrates the stage of the decomposition of an elastically strained germanium film, which covers steps on the substrate, into isolated islands. In this case, after the deposition of a germanium film with an effective thickness close to 12 ML, the sample was annealed at a temperature of 550°C by passing direct current for 3 min. The deposition of germanium at sufficiently high temperatures of the substrate, in our opinion, prevents the formation of structural defects in nanoislands, which can take place when the epitaxial deposition is performed at low temperatures (e.g., 200°C) [2].

Thus, in this paper, we suggest a new original approach to the formation of highly ordered structures with germanium nanoislands on silicon, which permits one to produce dense arrays of islands with dimensions of 10–20 nm. The results obtained indicate the basic possibility of directionally affecting the process of self-organization of the surface of heterostructures by introducing an additional factor such as electromigration with the purpose of creating arrays of nanoislands with predetermined parameters.

REFERENCES

1. Y.-N. Yang, E. S. Fu, and E. D. Williams, Surf. Sci. **36**, 101 (1996).
2. O. P. Pchelyakov, Yu. B. Bolkhovityanov, A. V. Dvurechenskiĭ, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 1281 (2000) [Semiconductors **34**, 1229 (2000)].

Translated by S. Gorin

Coherent Tunneling between Elementary Conducting Layers in the NbSe₃ Charge-Density-Wave Conductor

Yu. I. Latyshev[1]*, A. A. Sinchenko[2, 3], L. N. Bulaevskiĭ[4], V. N. Pavlenko[1], and P. Monceau[3]

[1] *Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 101999 Russia*
*e-mail: lat@mail.cplire.ru
[2] *Moscow State Engineering Physics Institute, Moscow, 115409 Russia*
[3] *Centre de Recherches sur Les Trés Basses Températures, CNRS, 38042 Grenoble, France*
[4] *Los Alamos National Laboratory, Los Alamos, New Mexico NM 87545, USA*
Received November 5, 2001; in final form, December 17, 2001

Characteristic features of transverse transport along the a^* axis in the NbSe₃ charge-density-wave conductor are studied. At low temperatures, the I–V characteristics of both layered structures and NbSe₃–NbSe₃ point contacts exhibit a strong peak of dynamic conductivity at zero bias voltage. In addition, the I–V characteristics of layered structures exhibit a series of peaks that occur at voltages equal to multiples of the double Peierls gap. The conductivity behavior observed in the experiment resembles that reported for the interlayer tunneling in Bi-2212 high-T_c superconductors. The conductivity peak at zero bias is explained using the model of almost coherent interlayer tunneling of the charge carriers that are not condensed in the charge density wave. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 71.45.Lr; 72.15.Nj; 74.50.+r

It is well known that the crystal structure of BSCCO high-T_c superconductors consists of atomically thin superconducting cuprate layers spatially separated by atomically thin insulating layers of BiO and SrO. The corresponding spatial modulation of the superconducting order parameter in the direction across the layers (along the c axis) leads to a discrete description of the transverse transport (the Lawrence–Doniach model [1]) with the neighboring superconducting layers being coupled through Josephson tunneling junctions. The validity of this approach is confirmed by the results of the experiments on natural Bi-2212 layered structures of small lateral size [2, 3]. Currently, the study of the interlayer tunneling of Cooper pairs and quasiparticles [4, 5] is one of the new original methods for investigating high-T_c superconductors.

In this paper, we study the interlayer tunneling in a layered system with an electron condensate of a different type, namely, in a charge-density-wave (CDW) conductor. The material chosen for our experiments is NbSe₃. This compound is characterized by two Peierls transitions, which occur at the temperatures $T_{p1} = 145$ K and $T_{p2} = 59$ K. In the low-temperature Peierls state, the Fermi surface retains some regions where the nesting conditions are not satisfied (the "pockets") and, hence, the Peierls gap is absent. Therefore, NbSe₃ does not undergo transition to the insulating state and retains its metallic properties down to the lowest temperatures [6]. From the analysis of both the crystal structure of

NbSe₃ and the anisotropy of its properties, it follows that this material can be classed with quasi-two-dimensional layered compounds. In fact, its conductivity anisotropy in the (b–c) plane is determined by the chain conductivity along the b axis and is estimated as $\sigma_b/\sigma_c \sim 10$, whereas the conductivity ratio σ_b/σ_{a^*} is determined by the layered character of the structure and reaches the values $\sim 10^4$ at low temperatures [7, 8]. Figure 1a shows the crystal structure of NbSe₃ in the (a–c) plane. One can see that, in this material, the layered structure is formed as a result of a pairwise coordination of selenium prisms with the predominant orientation of their bases in the (b–c) plane. In Fig. 1b, the shaded areas indicate the elementary conducting layers in which the prisms are rotated and shifted with their edges toward each other. In these layers, the distances between the niobium chains are relatively small, whereas the neighboring conducting layers are separated by an insulating layer formed as a double barrier by the bases of the selenium prisms.

This type of layered structure in combination with the conductivity anisotropy offers the possibility for the CDW condensation in the elementary conducting layers spatially separated by atomically thin insulating layers. In this case, as in layered high-T_c superconductors, one can expect that the CDW order parameter will be modulated along the a^* axis, and the transport across the layers will be determined by the intrinsic interlayer
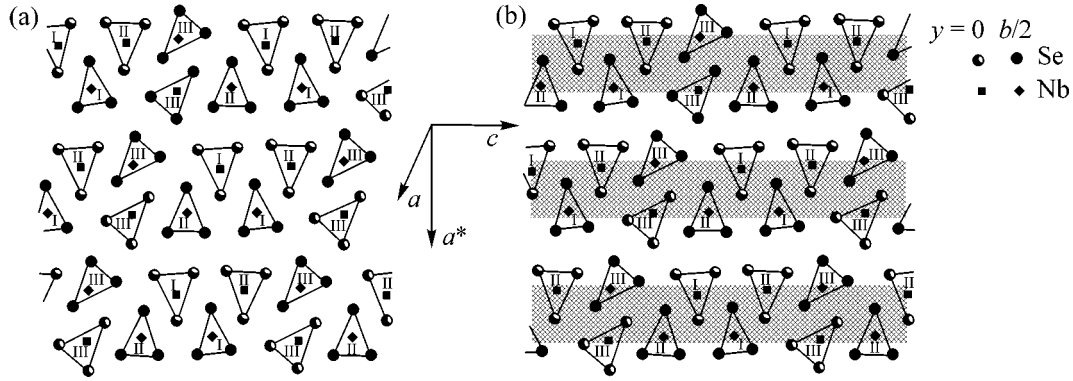
Fig. 1. (a) Schematic representation of the NbSe$_3$ structure in the (a–c) plane; (b) the same structure with the conducting planes indicated by shading.

tunneling between the elementary layers with the CDW.

To verify these speculations, we carried out an experimental study of the transport across the layers in NbSe$_3$ in the condensed CDW state.
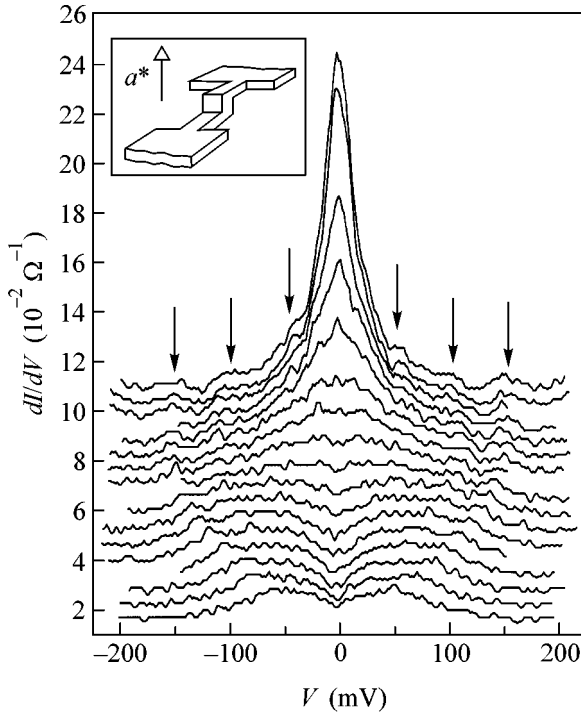

Fig. 2. Dependences of dI/dV on the voltage V for an overlap junction (sample No. 1) at different temperatures: 59.5, 55.6, 52.1, 48.0, 43.7, 40.1, 37.0, 34.3, 31.0, 28.0, 25.1, 22.6, 19.1, 15.7, 12.8, 8.0, and 4.2 K. The dynamic conductivity scale corresponds to the curve at $T = 59.5$ K, and other curves are shifted upwards for clarity. The inset shows the sample configuration.

The samples used in the experiment were layered structures with a small area through which the current flows across the layers, $S = 2 \times 2$ μm, and with the number of elementary layers ~ 30 (the overlap junctions). They were made from a thin NbSe$_3$ single crystal by the focused ion beam processing technique developed for fabricating similar structures from Bi-2212 [9]. Schematically, the structure under investigation is shown in the inset of Fig. 2. In addition, we studied the characteristics of NbSe$_3$–NbSe$_3$ point contacts oriented in the direction of the a^* axis. The configuration of such a contact is shown in the inset of Fig. 3. The contact was formed directly at low temperature by bringing two NbSe$_3$ whiskers together with high accuracy. In both cases, we performed four-terminal measurements of the I–V characteristics and their derivatives at low temperatures, i.e., below the second Peierls transition $T_{p2} = 59$ K.

The main results of the experiment are shown in Fig. 2, which, for one of the virtually perfect structures (no. 3), displays the dynamic conductivity along the a^* axis versus the bias voltage V at different temperatures from 59.5 to 4.2 K. It can be seen that, for $T < T_{p2}$, the differential I–V characteristics are of a tunneling character. When the temperature decreases below ~35 K, the I–V characteristics begin to exhibit a conductivity peak at zero bias voltage. As the temperature decreases further, this peak becomes a dominant feature. The peak amplitude reaches saturation when the temperature decreases below 6–8 K, and at $T = 4.2$ K, it is almost 20 times as great as the conductivity observed at high bias voltages. Note that this anomaly cannot be attributed to Joule heating. The estimate of the sample heating for a typical value of heat transfer to helium yields a conductivity decrease of less than 10% for the whole range of measured voltages. In addition, the I–V characteristics show a clearly defined set of conductivity peaks that are symmetric about zero voltage and at low temperature correspond to $|V| = 50$, 100, and 150 mV, i.e., to $|V| = nV_0$, where $V_0 = 50$ mV and $n = 1$, 2, 3. As the temperature increases above 25 K, these
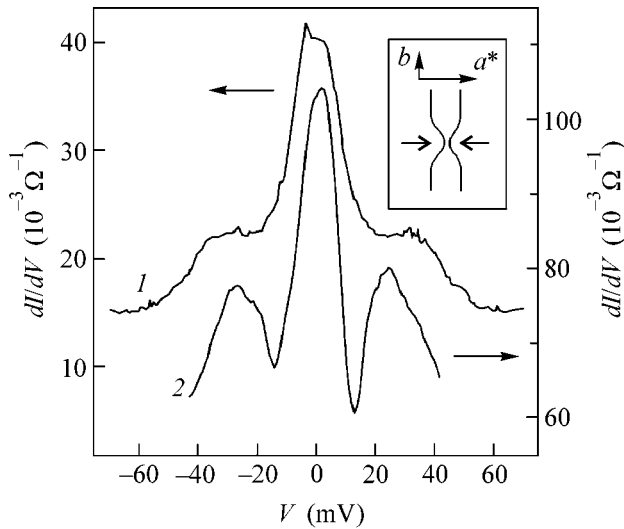
Fig. 3. Dependences $dI/dV(V)$ (*1*) for a NbSe$_3$–NbSe$_3$ point contact and (*2*) for an overlap junction (sample No. 2) at $T = 4.2$ K. The inset shows the point contact configuration.

peaks move to lower energies, and, at the temperature $T \simeq 59$ K corresponding to the second Peierls transition for NbSe$_3$, the values of V_n vanish.

The picture described above was observed only for perfect structures. The presence of structure defects, such as a twin boundary, leads to a considerable reduction of the conductivity peak at zero bias and to the appearance of a peak at $V = 25$–27 mV $\approx V_0/2$ (curve *2*

in Fig. 3). Qualitatively similar dependences were observed for some of the NbSe$_3$–NbSe$_3$ point contacts. Their *I–V* characteristics also exhibit a conductivity peak at zero bias with an amplitude approximately equal to that observed for defect overlap junctions, as well as a conductivity peak at $V \approx 25$ mV (curve *1* in Fig. 3). Figure 4 presents the temperature dependences of the normalized dynamic conductivity at zero bias voltage for all types of samples studied in the experiment. One can see that, for a perfect overlap junction, the amplitude of the zero bias anomaly is more than three times greater than the amplitudes observed for the defect layered structures and point contacts. Note that the voltage value $V_0 = 50$ mV is close to twice the value of the low-temperature energy gap, $2\Delta_{p2}/e$, for NbSe$_3$ [10, 11], while the dependence $V_n(T)/V_n(0)$ obtained for both overlap junctions and point contacts (Fig. 5) agrees well with the temperature dependence of the energy gap predicted by the BCS theory (the dashed line in Fig. 5). This result suggests that the conductivity features observed in the experiment are governed by a gap mechanism.

It is important to note that all effects described above were observed exclusively for the transport across the layers (along the *a** axis). The extra experiments performed by us on specially fabricated bridges and point contacts oriented along the *c* axis (with the transport across the chains in the layer plane) showed no conductivity peak at $V = 0$.

Let us first analyze the results obtained for perfect overlap junctions. The most prominent feature of the *I–V* characteristics of these structures is the strong con-
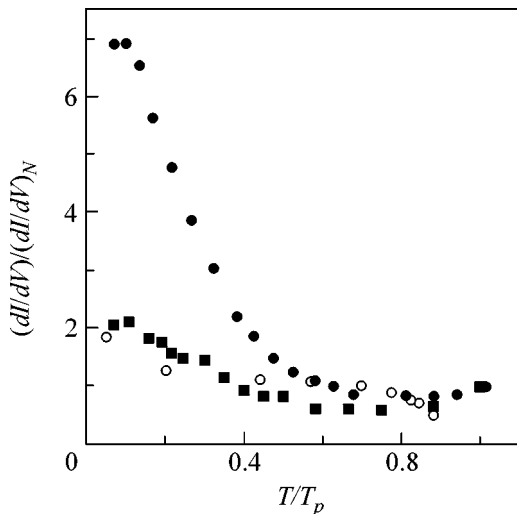

Fig. 4. Temperature dependences of the dynamic conductivity at zero bias voltage for two overlap junctions: sample No. 3 (full circles) and sample No. 1 (full squares), and for a NbSe$_3$–NbSe$_3$ point contact (empty circles). The conductivity values are normalized to the value of dI/dV ($T = 62$ K).
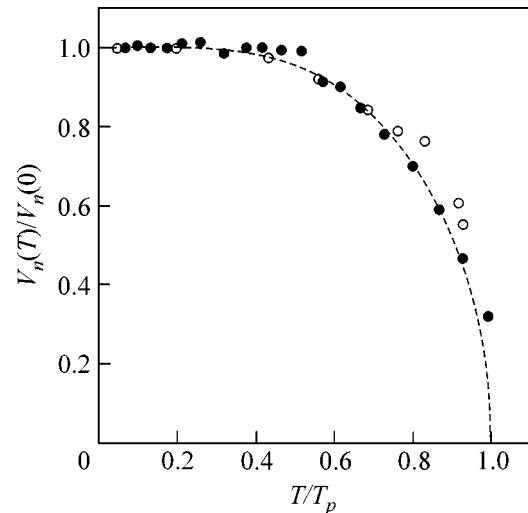

Fig. 5. Temperature dependence of the second Peierls energy gap for an overlap junction (sample No. 3, full circles) and a NbSe$_3$–NbSe$_3$ point contact (empty circles).

ductivity peak at zero bias voltage. In addition, the characteristics exhibit a periodic sequence of peaks at $|V| = nV_0$, which resembles the series of quasiparticle branches observed in the *I*–*V* characteristics of Bi-2212 layered structures when the measurements are performed across the layers [3]. As was noted above, the value $V_0 = 50$ mV is close to twice the value of the low-temperature CDW gap in NbSe$_3$, and the temperature dependence $V_0(T)$ follows the prediction of the BCS theory. However, in contrast to the case of a superconductor, the conductivity observed for NbSe$_3$ at zero bias voltage is finite. This fact suggests that the interlayer conduction at zero bias does not originate from the collective CDW contribution of the Josephson tunneling type.

The conductivity peak observed at zero bias cannot be explained by the regular (incoherent) single-particle tunneling as well. If this were the case, the conductivity peak could be attributed to the energy dependence of the density of states of electrons that are not condensed in the CDW. Then, the conductivity feature under discussion should also be observed for man-made N–I–CDW tunnel junctions. However, no such effects were revealed by the detailed experimental studies of this kind of tunnel junctions oriented along the *a** axis [10, 11].

We believe that the conductivity peak at zero bias is caused by the coherent interlayer tunneling of the charge carriers localized in the pockets of the Fermi surface, and below we thoroughly substantiate this statement.

In the general case, the tunnel current between two nonsuperconducting layers, *a* and *b*, is described by the expression [12]

$$I = \frac{4\pi es}{\hbar} \sum_{\mathbf{p}, \mathbf{q}} |t_{ab}(\mathbf{p}, \mathbf{q})|^2$$

$$\times \int_{-\infty}^{+\infty} d\omega \left(\tanh\frac{\omega}{2T} - \tanh\frac{\omega - eV}{2T} \right) \qquad (1)$$

$$\times \mathrm{Im}G_a^R(\mathbf{p}, \omega)\mathrm{Im}G_b^R(\mathbf{q}, \omega - eV),$$

where G^R is the retarded Green function of an electron with the momentum $\mathbf{p}$ in the layer *a*, $\mathrm{Im}G^R(\mathbf{p}, \omega)$ is the spectral density of the Green function, $t_{ab}(\mathbf{p}, \mathbf{q})$ is the matrix element characterizing the tunneling from the state with the momentum $\mathbf{p}$ in the layer *a* to the state with the momentum $\mathbf{q}$ in the layer *b*, *V* is the voltage between the adjacent layers, and *T* is the temperature.

Note that the contribution of the collective (moving) CDW mode [13] to the interlayer tunneling is absent, because the tunnel current is determined by the Green functions in different layers, whereas the collective mode propagates within the layers.

Below, we consider the coherent tunneling, when the electron momentum in the plane does not change under the tunneling: $\mathbf{p} = \mathbf{q}$. Then, we have

$$I = \frac{es}{\pi\hbar}\int d\mathbf{p}|t_{ab}(\mathbf{p})|^2$$

$$\times \int_{-\infty}^{+\infty} d\omega \left(\tanh\frac{\omega}{2T} - \tanh\frac{\omega - eV}{2T} \right) \qquad (2)$$

$$\times \mathrm{Im}G_a^R(\mathbf{p}, \omega)\mathrm{Im}G_b^R(\mathbf{p}, \omega - eV).$$

Taking into account that the scattering in a layer is determined by the collision frequency ν, we obtain

$$\mathrm{Im}G^R(\mathbf{p}, \omega) = \frac{\gamma}{\pi(\omega - \epsilon(\mathbf{p})^2 + \gamma^2)}, \qquad (3)$$

where $\epsilon(\mathbf{p})$ is the electron spectrum and $\gamma = \hbar\nu$. We can use this expression with $\gamma = \gamma_{sc} + \gamma_{inc}$, which takes into account the change in the momentum because of the scattering within the layers and the change due to the tunneling $\gamma_{inc} = \langle \epsilon(\mathbf{p}) - \epsilon(\mathbf{q}) \rangle$. In both cases, we take into account the energy uncertainty for the state characterized by the momentum $\mathbf{p}$. We define the tunneling as almost coherent when γ_{inc} and γ are small relative to other energy parameters of the electron system. In our case, such parameters are the Peierls gap and the width of the electron band in the pockets. Replacing $t_{ab}(\mathbf{p})$ by the momentum-independent quantity *t* and integrating with respect to $\mathbf{p}$, we obtain

$$\frac{1}{4\pi^2}\int d\mathbf{p}|t_{ab}(\mathbf{p})|^2\mathrm{Im}G_a^R(\mathbf{p}, \omega)\mathrm{Im}G_b^R(\mathbf{p}, \omega - eV)$$

$$= \frac{2\gamma N(0)|t|^2}{e^2V^2 + 4\gamma^2}, \qquad (4)$$

where $N(0)$ is the density of states of electrons in the pockets. Finally, for the interlayer current when $eV < 2\Delta$, we obtain the expression

$$I(V) = \frac{N(0)|t|^2\gamma eV}{2\pi^3(e^2V^2 + 4\gamma^2)}, \qquad (5)$$

and the dynamic conductivity has the form

$$\frac{\sigma(V)}{\sigma(0)} = 4\gamma^2 \frac{4\gamma^2 - e^2V^2}{(e^2V^2 + 4\gamma^2)^2}. \qquad (6)$$

Note that the temperature dependence is present in $\gamma(T)$ only. One can see that the dynamic conductivity has a peak of width $\approx\gamma$ at $V = 0$ and becomes negative and unstable when $eV > 2\gamma$. For a thirty-layer structure, the experimentally observed peak width ≈10 mV corresponds to $\gamma \approx 0.3$ meV. For comparison, we estimate the parameter γ_{sc} characterizing the scattering within the layers. Using the known mobility data $\mu = e/\nu_{sc}m^* \approx 4 \times 10^4$ cm^2/V s [7, 14], where $m^* = 0.24m_e$ [15], we obtain $\gamma_{sc} \approx 0.13$ meV; i.e., the changes in the electron

momentum because of tunneling are approximately equal to changes caused by the scattering within the layers.

Now, we explain why the interlayer current decreases with increasing voltage V. The electron tunneling between the layers must obey the energy conservation law; i.e., $\epsilon(\mathbf{p}) = \epsilon(\mathbf{q}) - eV$ to within 2γ. For coherent tunneling, we have $\epsilon(\mathbf{p}) = \epsilon(\mathbf{q})$, and for $eV \ll \gamma$, tunneling is possible and we have the conventional Ohm's law. When $eV > 2\gamma$, tunneling is impossible up to the voltage V reaching $2\Delta/e$. In this case, electrons condensed in the CDW begin to contribute to the interlayer current in the form of a regular tunneling of CDW quasiparticles through the double Peierls gap 2Δ. Thus, the interlayer current can be realized by means of only one of the two aforementioned mechanisms. In an actual multilayer structure, because of the geometric nonuniformity of individual layers (the areas of the layers are somewhat different), the voltage value $V \sim 2\gamma/e$ will not be reached simultaneously for different individual tunneling junctions. When $V > 2\gamma/e$, some of the junctions can be in the coherent tunneling regime whereas other junctions can be in the regime of single-particle tunneling through the gap. As a result, a sequence of conductivity peaks must appear in the I–V characteristic of the compound under study at the voltages $V = 2n\Delta_p/e$, where $n = 1, 2, \ldots$.

In the presence of defects in an overlap junction or in the case of a point contact, the incoherence of tunneling is enhanced, although in the best structures the tunneling remains almost coherent. Therefore, the I–V characteristics of these structures, along with the conductivity peak at zero bias, contain an additional feature at $V = \Delta_p/e$, which is related to the single-particle tunneling of the N–I–CDW type.

Thus, the results of this study show that the interlayer tunneling in natural layered structures obtained from NbSe$_3$ represents an independent efficient method for investigating the Peierls state in this compound. The tunneling conductivity peak observed at zero voltage can be self-consistently explained by the almost coherent interlayer tunneling of charge carriers that are not condensed in the CDW and are localized in the Fermi surface pockets not covered by the gap. The series of equidistant peaks in the I–V characteristic can be explained by the quasiparticle tunneling through the CDW gap under sequential transitions of individual tunneling junctions to the resistive state. Note that the conductivity peak observed at zero bias is a unique manifestation of coherent single-particle transport in solids.

REFERENCES

1. W. E. Lawrence and S. Doniach, in *Proceedings of the 12th International Conference on Low Temperature Physics, Kyoto, Japan, 1971*, Ed. by E. Kanda, p. 361.
2. R. Kleiner and P. Müller, Phys. Rev. B **49**, 1327 (1994).
3. R. Kleiner, P. Müller, H. Kohlstedt, *et al.*, Phys. Rev. B **50**, 3942 (1994).
4. K. Tanabe, Y. Hidaka, S. Karimoto, and M. Suzuki, Phys. Rev. B **53**, 9348 (1996).
5. Yu. I. Latyshev, T. Yamashita, L. N. Bulaevskii, *et al.*, Phys. Rev. Lett. **82**, 5345 (1999).
6. G. Grüner, *Density Waves in Solids* (Addison-Wesley, Reading, 1994).
7. N. P. Ong and J. W. Brill, Phys. Rev. B **18**, 5265 (1978).
8. Yu. I. Latyshev, P. Monceau, O. Laborde, *et al.*, J. Phys. IV **9**, 165 (1999).
9. S.-J. Kim, Yu. I. Latyshev, and T. Yamashita, Appl. Phys. Lett. **74**, 1156 (1999).
10. T. Ekino and J. Akimitsu, Physica B (Amsterdam) **194–196**, 1221 (1994).
11. A. Fournel, J. P. Sorbier, M. Konczykovski, and P. Monceau, Phys. Rev. Lett. **57**, 2199 (1986).
12. I. O. Kulik and I. K. Yanson, in *Josephson Effect in Superconducting Tunnel Structures* (Nauka, Moscow, 1970).
13. P. A. Lee, T. M. Rice, and P. W. Anderson, Solid State Commun. **14**, 703 (1974).
14. N. P. Ong, Phys. Rev. B **18**, 5272 (1978).
15. R. V. Coleman, M. P. Everson, Hao-An Lu, *et al.*, Phys. Rev. B **41**, 460 (1990).

Translated by E. Golyamina

Threshold Nucleation of a Nanometer-Scale Periodic Adatom Structure with the Participation of a Static Surface Acoustic Wave

V. I. Emel'yanov* and K. I. Eremin

International Laser Center, Moscow State University, Moscow, 119899 Russia

*e-mail: emel@em.msk.ru

Received December 20, 2001

A theory is developed to describe the nucleation of a periodic nanometer-scale adatom structure formed as a result of the instability caused by the interaction of adatoms with a self-consistent quasi-Rayleigh acoustic wave. The periods of the structure are determined as functions of adatom concentration and temperature. On the basis of the theoretical results, a mechanism is proposed that explains the recently observed effect of nanometer-scale structurization in the course of low-temperature molecular beam epitaxy of Ga under laser irradiation. © 2002 MAIK "Nauka/Interperiodica".

PACS numbers: 68.43.-h; 79.20.Ds; 81.07.-b; 81.16.Rf

1. The spontaneous formation of periodic nanometer-scale adatom structures on crystal surfaces has recently become an object of intensive study. The period of such a nanostructure is usually determined from the calculation of the stationary (equilibrium) free energy of this structure [1]. For some problems, the study of the initial stage of the periodic adatom structure formation (the nucleation) is of particular interest (see below). In this paper, we develop a kinetic theory of the spontaneous nucleation of a nanometer-scale surface grating that occurs as a result of the instability in the system of adatoms interacting through the deformation field.

The atoms adsorbed in the course of molecular beam epitaxy or deposition can be considered as surface defects (elastic inclusions). Owing to the deformation potential, and due to the local renormalization of the surface energy, the defects cause an inhomogeneous deformation of the surface and the underlying elastic continuum. In its turn, this self-consistent inhomogeneous deformation, also through the deformation potential, redistributes the adatoms along the surface.

This paper shows that, when the adatom concentration exceeds some critical value, such defect–deformation (DD) feedback leads to the development of a DD instability, which gives rise to a nanometer-scale periodic modulation of the surface relief with the accumulation of adatoms at the maxima or minima of the modulated structure. In this kind of DD structure, the components of the displacement vector of the medium are determined by the expressions that coincide with the corresponding formulas for the displacement vector in

a Rayleigh surface acoustic wave (SAW) in the static limit (i.e., when the wave frequency tends to $\omega = 0$).

2. Let the surface of an isotropic solid coincide with the plane $z = 0$ and the z axis be directed from the surface into the depth of the medium.

The equation for the displacement vector **u** of the medium has the form

$$\partial^2 \mathbf{u}/\partial t^2 = c_t^2 \Delta \mathbf{u} + (c_l^2 - c_t^2)\mathrm{grad}(\mathrm{div}\,\mathbf{u}), \qquad (1)$$

where c_l and c_t are the longitudinal and transverse velocities of sound, respectively.

Assume that a spatially inhomogeneous surface perturbation of the elastic continuum arises along a given direction x. We represent this perturbation in the form of a static ($\omega = 0$) quasi-Rayleigh SAW [2] with an amplitude exponentially growing with time:

$$\begin{aligned} u_x = &-iqR\exp(iqx + \lambda t - k_l z) \\ &- ik_t Q\exp(iqx + \lambda t - k_t z), \end{aligned} \qquad (2a)$$

$$\begin{aligned} u_z = &\ k_l R\exp(iqx + \lambda t - k_l z) \\ &+ qQ\exp(iqx + \lambda t - k_t z), \end{aligned} \qquad (2b)$$

$$k_{l,t}^2 = q^2 + \lambda^2/c_{l,t}^2, \qquad (3)$$

where λ is the increment of the DD instability and R and Q are the initial amplitudes of the SAW. Expressions (2) and (3) define the solution to Eq. (1).

The strain ξ at the surface is related to the components of the displacement vector as follows:

$$\xi = \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} = \xi(q)\exp(iqx + \lambda t),$$

$$\xi(q) = -\frac{\lambda^2}{c_l^2}R. \tag{4}$$

The spatially inhomogeneous surface strain ξ gives rise to the perturbation of the homogeneous distribution of adatoms:

$$N_d = N_{d0} + N_{d1}(x) = N_{d0} + N_{d1}(q)\exp(iqx + \lambda t), \tag{5}$$

where N_{d0} is the spatially homogeneous part and $N_{d1}(q)$ is the amplitude of the spatially inhomogeneous perturbation ($N_{d1}(x) \ll N_{d0}$).

The equation for N_d has a form similar to the equation for the volume concentration of defects [3]:

$$\frac{\partial N_d}{\partial t} = D_d\frac{\partial^2 N_d}{\partial x^2} - D_d\frac{\theta_d}{k_B T}\frac{\partial}{\partial x}\left[N_d\frac{\partial}{\partial x}\left(\xi + l_d^2\frac{\partial^2\xi}{\partial x^2}\right)\right], \tag{6}$$

where D_d is the surface diffusion coefficient for adatoms, θ_d is the surface deformation potential of an adatom, and l_d is the characteristic interaction length for the adatom interaction with the atoms of the crystal lattice [3]. The term proportional to l_d takes into account the nonlocal character of the interaction between the adatoms and the surface atoms [3, 4].

Linearizing Eq. (6) with allowance for the condition $N_{d1}(x) \ll N_{d0}$ and using Eq. (5), we obtain

$$(\lambda + D_d q^2)N_{d1}(q) = D_d N_{d0}\frac{\theta_d}{k_B T}q^2[1 - l_d^2 q^2]\xi(q). \tag{7}$$

The inhomogeneous adatom concentration $N_{d1}(x)$ modulates the surface energy $\sigma(x) = \sigma_0 + (\partial\sigma/\partial N_d)N_{d1}(x)$, which leads to the appearance of a shear (lateral) surface stress $\partial\sigma(x)/\partial x$, which is compensated by the shear stress arising in the medium. The boundary condition for the balance of lateral stresses in the linear approximation in strain has the form

$$\mu\left[\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right]_{z=0} = \left(\frac{\partial\sigma}{\partial N_d}\right)\frac{\partial N_{d1}(x)}{\partial x}, \tag{8}$$

where μ is the shear modulus of the medium. Below, we consider the coefficient $(\partial\sigma/\partial N_d)$ as a preset phenomenological parameter.

The interaction of adatoms with the surface of the medium through the deformation potential also gives rise to a normal stress at the surface. The corresponding (linear in strain) boundary condition has the form

$$\left[\frac{\partial u_z}{\partial z} + (1 - 2\beta)\frac{\partial u_x}{\partial x}\right]_{z=0} = \frac{\theta_d N_{d1}(x)}{\rho c_l^2 a}, \tag{9}$$

where a is the crystal lattice constant of the surface, $\beta = c_t^2/c_l^2$, and ρ is the density of the medium.

Substituting Eq. (2) and Eq. (7) with allowance for Eqs. (4) in Eqs. (8) and (9), we obtain a system of two linear equations for the amplitudes R and Q. From the condition that the determinant of this system is zero, we derive a dispersion equation for the DD instability:

$$[q^2 + k_t^2]^2 - 4q^2 k_l k_t = \frac{2\lambda^2}{\beta c_l^2}\frac{\theta_d N_{d0}}{k_B T\rho c_l^2}\frac{D_d q^2}{\lambda + D_d q^2}$$

$$\times (1 - l_d^2 q^2)\left[q^2 k_t\frac{\partial\sigma}{\partial N_d} + (q^2 + k_t^2)\frac{\theta_d}{2a}\right]. \tag{10}$$

After the substitution $\lambda \longrightarrow i\omega$ the left-hand side of Eq. (10) coincides with the Rayleigh determinant, which, being set equal to zero, determines the dispersion law for the Rayleigh acoustic wave [2]. The right-hand side of Eq. (10) describes the force action ($\sim\theta_d$) of adatoms, which deforms the surface and leads to the DD instability.

Expanding k_l and k_t in powers of the small parameter $(\lambda^2/c_{l,t}^2 q^2)$, we derive from Eq. (10) the expression for the DD instability increment:

$$\lambda(q) = D_d q^2$$

$$\times\left[\frac{\theta_d N_{d0}}{(1 - \beta)k_B T\rho c_l^2}(1 - l_d^2 q^2)\left(q\frac{\partial\sigma}{\partial N_d} + \frac{\theta_d}{a}\right) - 1\right]. \tag{11}$$

3. Consider two limiting cases of Eq. (11).

A. When the shear stress can be neglected ($\theta_d/a \gg q(\partial\sigma/\partial N_d)$), from Eq. (11) at $T = $ const, we obtain

$$\lambda(q) = D_d q^2\left[\frac{N_{d0}}{N_{dc}}(1 - l_d^2 q^2) - 1\right], \tag{12}$$

where N_{dc} is the critical concentration of adatoms:

$$N_{dc} = (1 - \beta)\frac{a\rho c_l^2 k_B T}{\theta_d^2}. \tag{13}$$

Setting $\theta_d = 10$ eV, $T = 100$ K, $a = 5 \times 10^{-8}$ cm, and $\rho c_l^2 = 10^{12}$ erg/cm^3, from Eq. (13) we obtain the estimate $N_{dc} \approx 1.8 \times 10^{12}$ cm^{-2}.

Function (12) has a maximum at $q = q_m$. This value determines the period $\Lambda = 2\pi/q_m$ of the dominant DD grating:

$$\Lambda(N_{d0}) = \sqrt{8}\pi l_d[1 - N_{dc}/N_{d0}]^{-1/2}. \tag{14}$$

The formation of the DD grating on the surface occurs in a threshold manner when the adatom concentration exceeds the critical value: $N_{d0} > N_{dc}$. At the DD instability threshold ($N_{d0} \longrightarrow N_{dc}$), the period tends to infinity, $\Lambda \longrightarrow \infty$. At high concentrations $N_{d0} \gg N_{dc}$, the

period asymptotically tends to its minimal value $\Lambda = \sqrt{8}\,\pi l_d$.

At a constant adatom concentration $N_{d0} = $ const, the period is determined by the temperature:

$$\Lambda(T) = \sqrt{8}\pi l_d[1 - T/T_c]^{-1/2}, \tag{15}$$

where T_c is the critical temperature:

$$T_c = \frac{1}{(1-\beta)}\frac{\theta_d^2 N_{d0}}{a\rho c_l^2 k_B}. \tag{16}$$

At $N_{d0} = 2 \times 10^{12}$ cm^{-2}, from Eq. (16) we obtain an estimate for the critical temperature $T_c \approx 110$ K.

According to Eq. (15), the formation of the DD grating on the surface is possible only when $T < T_c$. As $T \longrightarrow 0$, the period tends to the minimal value $\Lambda = \sqrt{8}\pi l_d$, and when $T \longrightarrow T_c$, the period tends to infinity, $\Lambda \longrightarrow \infty$.

B. Consider the second limiting case of Eq. (11) where the normal stress can be neglected ($\theta_d/a \ll q(\partial\sigma/\partial N_d)$). The necessary condition for the DD instability is the condition $\theta_d(\partial\sigma/\partial N_d) > 0$. In this case, from Eq. (11), we obtain

$$\lambda(q) = D_d q^2[pql_d(1 - l_d^2 q^2) - 1], \tag{17}$$

where p is the dimensionless control parameter:

$$p = \frac{1}{(1-\beta)}\frac{\theta_d(\partial\sigma/\partial N_d)N_{d0}}{l_d\rho c_l^2 k_B T}. \tag{18}$$

Function (17) has a maximum at $q = q_m$. This value is determined from the condition $\partial\lambda(q)/\partial q = 0$ and is obtained as a solution to the equation

$$5pl_d^3 q_m^3 - 3pl_d q_m + 2 = 0. \tag{19}$$

The corresponding period of the dominant DD grating is

$$\Lambda = \Lambda(p) = \sqrt{5}\pi l_d / \cos\left[\frac{\arccos(-\sqrt{5}/p)}{3}\right]. \tag{20}$$

The condition $\lambda(q_m) \geq 0$ determines the critical value of the control parameter $p \geq p_c = 2.597$. When this value is exceeded, the DD instability begins to develop.

Note that, from Eqs. (14), (15), and (20), it follows that the period of the adatom superlattice formed on the surface is proportional to the characteristic length of interaction between the adatoms and the atoms of the lattice, l_d, which lies in the nanometer range. The dispersion parameter l_d was introduced in the theory of the self-organization of three-dimensional DD nanostructures in an earlier publication [3]. In comparing the theoretical predictions with the experimental results, the parameter l_d can be considered as a fitting parameter (as in [3]). In another publication [4] devoted to the dynamics of the formation of three-dimensional DD nanostructures, it was found that this dynamics is also gov-

erned by l_d. Later [5], it was shown that the parameter l_d determines the screening length for the elastic interaction of defects with each other in media with high concentrations of point defects.

4. The x direction on the crystal surface has a preference because of the elastic anisotropy, and, on an isotropic surface, because of the external action inducing an elastic anisotropy, or owing to a spontaneous violation of the symmetry of the DD system (as in [6]). The growth of the amplitudes of the corresponding DD grating, which are given by Eqs. (2), (4), and (5), becomes saturated owing to the elastic nonlinearity of the DD system. In this case, a stationary one-dimensional adatom superlattice (a single-mode DD structure) is formed on the surface. Its amplitude can be calculated by taking into account the nonlinearity in the boundary conditions (8) and (9). By analogy with [6, 7], one can expect that, when the excess over the one-dimensional structure generation threshold is large enough, the formation of a two-mode or three-mode (hexagonal) nanostructure is most probable. In this case, the absolute extrema of the surface deformation form a cellular structure and can serve as the nucleation centers of a cellular surface nanostructure formed by adatoms.

In connection with this prediction, we note that recently, the group of Prof. N.I. Zheludev (from Southampton University) reported on the formation of a virtually close packing of Ga nanoparticles with a narrow size distribution on the illuminated (central) part of the end surface of a quartz optical waveguide as a result of the low-temperature (100 K) molecular beam epitaxy of Ga [8]. At the same time, the part of the substrate where the laser radiation was absent (corresponding to the peripheral part of the waveguide) was characterized by Ga particles with a wide distribution in size (up to microns).

The consideration presented above suggests the following possible scenario for the formation of an ensemble of Ga nanoparticles with a narrow size distribution under laser irradiation.

The radiation with a wavelength of 1550 nm [8], which lies within the molecular absorption band of covalently bonded Ga dimers [9], excites the dimers from the bonding state to the antibonding one [10], thus increasing the mobility of dimers and small Ga clusters. This provides the possibility for the development of a DD instability on the illuminated part of the end surface of the optical waveguide, and this instability induces the nucleation of a cellular adatom nanostructure. The periodic surface deformation field produced in this case gives rise to flows of randomly nucleated mobile Ga clusters in the direction toward the deformation extrema. As a result, only those clusters grow that are located at the extrema of the surface deformation field. Thus, in the illuminated region of the surface, within some time of cluster growth, a close packing of nanoparticles is formed with a narrow size distribution centered at a value that is proportional to the DD nano-

structure period given by Eq. (14), (15), or (20). From our consideration, it follows that the nucleation of the DD nanostructure is possible only at a sufficiently low temperature, below the critical one. This condition agrees with the results of the experiment [8], where the nanometer-scale structurization of Ga occurred at $T = 100$ K and was not observed at $T = 300$ K.

In the region where light is absent (the peripheral region), the deposition products are almost immobile and the DD instability is impossible. In this region, the particles nucleate in a random way and their subsequent growth also occurs randomly. As a result, in the nonilluminated region, Ga particles are formed with a wide distribution in size.

REFERENCES

1. V. A. Shchukin and D. Bimberg, Rev. Mod. Phys. **71**, 1125 (1999).

2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1987; Pergamon, New York, 1986).

3. V. I. Emel'yanov and I. M. Panin, Fiz. Tverd. Tela (St. Petersburg) **39**, 2029 (1997) [Phys. Solid State **39**, 1815 (1997)].

4. V. I. Emel'yanov and I. M. Panin, Fiz. Tverd. Tela (St. Petersburg) **42**, 1026 (2000) [Phys. Solid State **42**, 1058 (2000)].

5. V. I. Emel'yanov, Fiz. Tverd. Tela (St. Petersburg) **43**, 637 (2001) [Phys. Solid State **43**, 663 (2001)].

6. H. Haken, *Synergetics: an Introduction* (Springer-Verlag, Berlin, 1977; Mir, Moscow, 1980).

7. D. Walgraef, N. M. Ghoniem, and J. Lauzeral, Phys. Rev. B **56**, 15361 (1997).

8. K. F. MacDonald, W. S. Brocklesby, V. I. Emel'yanov, *et al.*, http://xxx.soton.ac.uk/abs/physics/0105042 (2001).

9. X. G. Gong, G. L. Chiarrotti, M. Parinello, and E. Tosatti, Phys. Rev. B **43**, 14277 (1991).

10. K. F. MacDonald, V. A. Fedotov, R. W. Eason, *et al.*, J. Opt. Soc. Am. B **18**, 331 (2001).

Translated by E. Golyamina