JETP Letters, Vol. 79, No. 11, 2004, pp. 515–518. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 643–646.
Original Russian Text Copyright © 2004 by Lonin, Mensov, Polushtaĭtsev.

Causes of the Filamentary Instability of Optical-Beam Self-Channeling in Photopolymerizable Composites

A. L. Lonin, S. N. Mensov, and Yu. V. Polushtaĭtsev
Nizhni Novgorod State University, Nizhni Novgorod, 603950 Russia

Received February 11, 2004; in final form, March 29, 2004

The model of the formation of filamentary inhomogeneities in the waveguide structures formed upon optical-beam self-channeling in photopolymerizable composites is constructed and numerically studied. The conditions for the formation of a homogeneous waveguide channel are determined and confirmed by the experimental results. *© 2004 MAIK “Nauka/Interperiodica”.*

PACS numbers: 42.65.Sf

The possibility of waveguide structure self-formation in nonlinear nonreversible materials (such as photopolymerizable composites) was pointed out in many works [1–5]. In such systems, a nondivergent polymeric channel is formed due to the radiation self-channeling even in the case of narrow optical beams possessing large diffraction divergence. Photopolymerizable composite (PPC) is a mixture of monomers and photoinitiators. In such a mixture, optical radiation induces the irreversible liquid monomer–solid polymer transition accompanied by an increase in the density of material and, hence, in its refractive index. For this reason, contrary to the conventional materials with Kerr nonlinearity, the resulting waveguide is retained after the radiation is terminated. The experimental results suggest that high-intensity radiation is not needed for the formation of such a waveguide; quite the contrary, the radiation self-channeling is observed only for low-intensity beams. With an increase in intensity, the waveguide regime is disturbed and the channel separates into individual filaments with sizes from 10 to 50 μm [3, 5]. For beams with the same intensity, the broader ones are less stable.

In this work, this phenomenon is numerically simulated by the example of the interaction of a regular Gaussian optical beam with a photopolymerizable composite containing small inhomogeneities of refractive index that are introduced, e.g., by thermodynamic fluctuations. The conditions for establishing stable radiation self-channeling regimes in PPCs are determined and confirmed by the experimental results.

The numerical simulation was performed under the following conditions. The PPC exposure characteristic (dependence of a change in the refractive index on the radiation amplitude $E(x, z, t)$) was approximated by the modified Rayleigh function [6]

$$\Delta n(x, z, t) = n(H) - n_0 = \Delta n_{\max}\left[1 - \exp\left(-\left(\frac{H(x, z, t)}{H_0}\right)^{\gamma}\right)\right], \quad (1)$$

where γ and H_0 are the constants characterizing the composite contrast, $\Delta n_{\max}$ is the maximal change in the refractive index, n_0 is the refractive index of the monomer, and $H(x, z, t)$ is the exposure measure defined by the expression

$$H(x, z, t) = \int_0^t |E(x, z, \tau)|^2 d\tau. \quad (2)$$

Considering the sluggishness of the polymerization process, the evolution of the complex field amplitude $E(x, z, t)$ of the optical beam with the initial intensity half-width a in PPC was analyzed using the quasi-stationary parabolic equation with the boundary condition $E(x, 0, t) = E_0\exp(-x^2/2a^2)$ $(E_0 = E(0\ 0, t))$, where time is a parameter [3]. In the dimensionless variables $x' = x/a$, $z' = z/ka^2$, and $t' = tE_0^2/H_0$, where k is the wave number, this equation has the form

$$2i\frac{\partial E'(x', z', t')}{\partial z'} = \Delta_\perp E'(x', z', t') + 2V^2\Delta n'(x', z', t')E'(x', z', t'), \quad (3)$$

where $E' = E/E_0$, $\Delta n' = \Delta n/\Delta n_{\max}$, and $V = ka(\Delta n_{\max}/n_0)^{0.5}$.

Inasmuch as the medium is in the liquid state at the initial polymerization stage, the change in its density is determined not only by the action of initiating radiation; one should also take into account the diffusion of composite components from the adjacent regions.

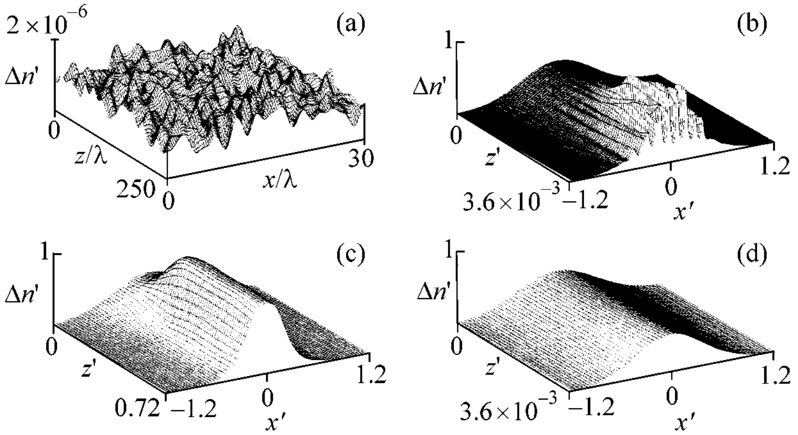


Fig. 1. Distribution of the nonlinear addition to the refractive index: (a) initial, (b) at time $t' = 0.9$ for $\gamma = 3$ and parameters $a = 100$ μm and $\beta = 1.8 \times 10^{-2}$, (c) the same for $a = 7.5$ μm and $\beta = 0.26$, and (d) the same for $a = 100$ μm and $\beta = 0.26$.

Because of this, the nonlinear addition to the refractive index was determined using, instead of Eq. (1), the kinetic equation [7], which has the following form in the dimensionless variables:

$$\frac{\partial \Delta n'(x', z', t')}{\partial t'} = \left(f(\Delta n'(x', z', t')) + \beta^2 \frac{\partial^2 \Delta n'(x', z', t')}{\partial x'^2}\right), \tag{4}$$

where $f(\Delta n')$ is the photopolymerization rate, and the parameter

$$\beta = \sqrt{\frac{t_p}{t_d}} = \frac{1}{a}\sqrt{\frac{DH_0}{E_0^2}} \tag{5}$$

determines the ratio between the characteristic polymerization time $t_p = H_0/E_0^2$ and the diffusion time $t_d = a^2/D$, where D is the monomer diffusivity in the monomer–polymer system. In accordance with local model (1),

$$f(\Delta n') = |E'|^2 \gamma (1 - \Delta n')[-\ln(1 - \Delta n')]^{1 - 1/\gamma}. \tag{6}$$

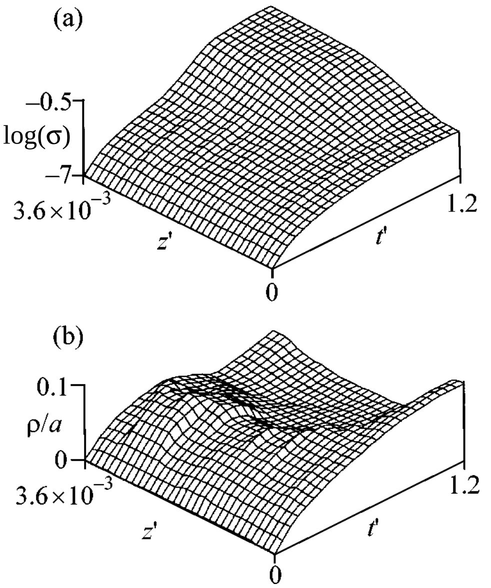


Fig. 2. (a) Amplitude of inhomogeneities and (b) their correlation radius as functions of the longitudinal coordinate and exposure time.

The results of numerical simulation showed that diffusion processes (4) do not affect the radiation self-channeling regime in the initially homogeneous material ($\Delta n(x, z, 0) = \text{const}$); for $\beta < 1$, $\gamma > 2$, and $\Delta n_{\max} \sim 1.5 \times 10^{-2}$, the optical beam forms a nondivergent homogeneous waveguide channel. However, even small-amplitude and smallest-scale random inhomogeneities $\Delta n(x, z, 0)$ (Fig. 1a) can noticeably change the structure of the radiation-induced polymeric track. For small values of parameter β (typical of the broad beams), the channel is homogeneous only in its initial sections, while the refractive index becomes transversely modulated as one moves away from the entrance plane (Fig. 1b). It is significant that the fluctuations of refractive index are enhanced in an avalanche-like manner almost to a level of Δn on the beam path, and the radius of inhomogeneity correlations noticeably increases. As an illustration, the dependences of the amplitude σ of these inhomogeneities and their correlation radius ρ are shown in Figs. 2a and 2b, respectively, as functions of time and the longitudinal coordinate for the initial values $\sigma = 10^{-6}$ and $\rho = 2\pi/k$ and for $a = 100$ μm, $\beta = 1.8 \times 10^{-2}$, $n_0 = 1.5$, and $\Delta n_{\max} = 1.5 \times 10^{-2}$. One can see that, in contrast to the instability scale in materials with the Kerr nonlinearity [8], the value of ρ is determined only by the parameter β (for a given composite contrast γ) and starts to exceed the wave scales during the process of photopolymerization. One can see from the graph in Fig. 2b that the transverse size of the filamentary inhomogeneities is virtually homogeneous. Their slight narrowing on the radiation propagation path (Figs. 1b and 2a) is caused by the specific features of radiation self-channeling and is typical of the entire channel for the paths exceeding the channel diffraction length (Fig. 1c). For this reason, to estimate the value of ρ, one may not solve the wave problem but restrict oneself to the solution of problem (4) for the section $z = 0$. The corresponding ρ dynamics is demonstrated in Fig. 3 for various values of β. For a PPC with the parameters $H_0 = 30$ (mW s)/mm^2, $D = 5 \times 10^{-5}$ mm^2/s, $n_0 = 1.5$, and $\Delta n_{\max} = 1.5 \times 10^{-2}$, one has $\rho \approx 20$ μm for $E_0^2 = 60$ mW/mm^2 and $t' = 0.9$ (Fig. 3). An analogous result is also given by the linearized solution to Eq. (4), which was used in [9] to estimate the magnitude of the intrinsic concentrational inhomogeneities arising in viscous media at the stage of fast density transformation. Then, in our notation,

$$\rho \approx \frac{a\beta}{\sqrt{\Delta n_{\max}\langle df/d\Delta n'\rangle_{\Delta n'}}}, \qquad (7)$$

where $\langle df/d\Delta n'\rangle_{\Delta n'}$ is the average polymerization acceleration in a time of PPC exposure.

At large β values (typical of the narrow beams acting on PPC), only one waveguide channel forms (Fig. 1c). Numerical studies of the radiation propagation for various β values showed that the criterion for the stable self-channeling regime is that the beam half-width be smaller than the sizes of self-formed inhomogeneities ($\rho/a > 1$) that arise in the broader beams at the same β value.

Taking into account the invariant dependence of β on the medium and radiation characteristic parameters (5), we note that the stable regime ($\rho/a > 1$) can also be realized for broad beams (Fig. 1d) at either lower radiation intensities E_0^2 or higher diffusion coefficients D. At the same time, the self-channeling of a narrow beam can become unstable upon an increase in the light intensity or composite viscosity.

Therefore, one can separate the regions of the stable and unstable regimes of optically induced formation of a waveguide channel in PPC by the level line $\rho/a = 1$ in the graph (Fig. 3) of ρ/a versus the parameters β (5) and t' that include the main characteristics of the polymerizable material and initiating radiation. Since ρ/a increases with increasing exposure time, one might assume that sufficiently long times t' would always provide the stable radiation self-channeling regime. However, in a real situation, this is prevented by an increase in the viscosity of a photopolymerizable material.

Within the framework of linearized model (7), one can also analytically write the approximate requirement on the material and radiation parameters that exclude the instability of the formed polymeric channel. Since $\langle df/d\Delta n'\rangle_{\Delta n'} = f(\Delta n')/\Delta n'$, setting $|E'|^2 = |E'_{\max}|^2 = 1$ and using Eqs. (1) and (6), one has

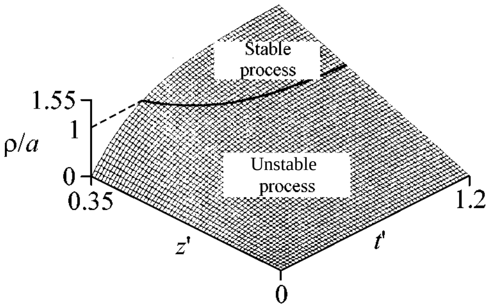


Fig. 3. Dependence of the correlation radius of inhomogeneities at the initial section on time and parameter β for $\gamma = 3$.

$$\beta\sqrt{\frac{1-\exp(-t'^{\gamma})}{\Delta n_{\max}\gamma(t')^{\gamma-1}\exp(-r'^{\gamma})}} \approx \frac{\rho}{a} > 1. \qquad (8)$$

EXPERIMENTAL STUDY OF THE STABILITY OF SELF-CHANNELING REGIMES

The possibility of predicting (Fig. 3) the stable regimes of optical-beam self-channeling in photopolymerizable composites was also studied experimentally for the interaction of He–Ne laser radiation with an OKM-2-based PPC containing quinone photoinitiator [10]. The parameters of exposure characteristic (1) for this composite in the range of intensities from 5 to 50 mW/mm^2 were the following: $\gamma = 3$, $H_0 = 30$ (mW s)/mm^2, $n_0 = 1.5$, and $\Delta n_{\max} = 1.5 \times 10^{-2}$. The laser intensity was adjusted using a polarization attenuator and focused onto the end of a cell with the photopolymer. The diameter of the input $E^2(x, 0, t)$ distribution was 100 μm. The optically formed polymeric channel was visualized using a video camera with lateral illumination by infrared light that did not modify the material. The exposure time was chosen on the condition that the monomer diffusivity be constant during the polymerization and was equal to $t' = tE_0^2/H_0 = 0.5$. This gives $D \approx 5 \times 10^{-5}$ mm^2/s. It follows from the graph in Fig. 3 that, in this case, the critical ($\rho/a = 1$) value of parameter β is 0.28. The corresponding intensity (see Eq. (5)) is $E_0^2 = 8$ mW/mm^2.

Experiment showed that, for an intensity of 50 mW/mm^2, which corresponds to $\beta = 0.11$ (see Eq. (5)), the channel split up into filaments with a transverse size of 20 μm. The photograph of the corresponding polymeric structure is given in Fig. 4a. For a beam intensity of 5 mW/mm^2, the self-channeling is stable, and only one waveguide channel is formed. Its image is presented in Fig. 4b ($\beta = 0.34$).

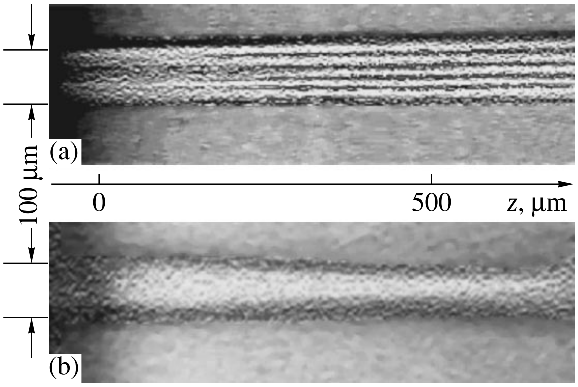


Fig. 4. Photographs of the polymeric channels formed in PPC by the optical beams with a width of 100 μm and intensities of (a) 50 and (b) 5 mW/mm^2.

The intrinsic inhomogeneities in such an unstable nonlinear system can also be generated by the inhomogeneities of the beam intensity. However, it can easily be seen that the self-consistency of the photopolymerization model considered in this work is based on mapping, by virtue of the diffraction effects, the inhomogeneities of refractive index of the material onto the inhomogeneities of beam intensity, and vice versa. It is the positive optical feedback which is the cause of the avalanche-like increase in the amplitude of intrinsic inhomogeneities in the polymeric channel. Because of this, the final result is independent of whether the small initial spatial inhomogeneities of $n(x, z, 0)$ or $E^2(x, z, 0)$ are primary or not.

This work was supported by the Grants of the Leading Scientific Schools of the Russian Federation (no. 1649.2003.3 from the G.A. Abakumov school and no. 1641.2003.2 from the V.A. Zverev and N.S. Stepanov school).

REFERENCES

1. A. S. Kewitsch and A. Yariv, Opt. Lett. **21**, 24 (1996).
2. T. M. Monro, L. Poladian, and C. M. de Sterke, Phys. Rev. E **57**, 1104 (1998).
3. V. A. Vdovin, A. L. Lonin, and S. N. Mensov, Zh. Tekh. Fiz. **71** (7), 67 (2001) [Tech. Phys. **46**, 853 (2001)].
4. M. Kagami, T. Yamashita, and H. Ito, Appl. Phys. Lett. **79**, 1079 (2001).
5. K. Dorkenoo, O. Cregut, L. Mager, *et al.*, Opt. Lett. **27**, 1782 (2002).
6. V. P. Roshchupkin, B. V. Ozerkovskiĭ, and Z. A. Karapetyan, Vysokomol. Soedin., Ser. A **19**, 2239 (1977).
7. G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: from Dissipative Structures to Order through Fluctuations* (Wiley, New York, 1977; Mir, Moscow, 1979).
8. M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *The Theory of Waves* (Nauka, Moscow, 1979) [in Russian].
9. M. I. Rabinovich and A. B. Ezerskiĭ, *Dynamic Theory of Forming* (Yanus-K, Moscow, 1998) [in Russian].
10. G. A. Abakumov, O. N. Mamysheva, V. A. Muraev, *et al.*, RF Patent No. 2138070 (1999).

Translated by V. Sakun

JETP Letters, Vol. 79, No. 11, 2004, pp. 519–522. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 647–650.
Original English Text Copyright © 2004 by K. Bliokh, Yu. Bliokh.

Optical Magnus Effect as a Consequence of Berry Phase Anisotropy[¶]

K. Yu. Bliokh[1, 2, *] and Yu. P. Bliokh[3]
[1] *Institute of Radio Astronomy, 4 Krasnoznamyonnaya ul., Kharkov, 61002 Ukraine*
[2] *Department of Physics, Bar-Ilan University, Ramat-Gan, 52900 Israel*
**e-mail: kostya@bliokh.kharkiv.com, k_bliokh@mail.ru*
[3] *Department of Physics, Technion, Haifa, 32000 Israel*

Received April 6, 2004

Presented in this work is a modified geometric optics of smoothly inhomogeneous isotropic medium, which takes into account weak anisotropy introduced by inhomogeneity. Pointed out is the common nature of two fundamental phenomena: Berry's geometrical phase and the optical Magnus effect, that is, propagation of rays of right and left circular polarization along different trajectories. Shown is that the former phenomenon can be explained by the difference in phase velocity of waves of right-hand and left-hand polarizations, while the latter one is the result of the difference in their group velocity. This work demonstrates that the optical Magnus effect is quite a topological effect that exclusively depends on the geometry of the system's contour in the momentum space. We predict the effect of the splitting of a ray of mixed polarization into two circularly polarized rays and propose a scheme for the experimental observation of this phenomenon. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 42.25.Ja; 42.15.-i; 03.65.Vf

1. INTRODUCTION

The first consistent analysis of asymptotic approximation of geometrical optics (GO) of electromagnetic waves in smoothly inhomogeneous isotropic media was made in a work by Rytov [1]. There he studied the GO in first approximation; also he gave a well-known law of rotation of polarization plane for electromagnetic wave. Geometrical attributes of this law were described further in detail by Vladimirskiy [2]. In the 1980s, Berry and his followers showed that geometrical phases are a general fundamental attribute of dynamic systems, and the Rytov–Vladimirskiy polarization plane rotation is an example of Berry phase (see [3] and references therein).

Another fundamental phenomenon that is not contained in Rytov's GO was discovered by Zel'dovich *et al.* in 1990 [4]. It was named the "optical Magnus effect" and consists in that waves of right and left circular polarization propagate in a smoothly inhomogeneous medium along different trajectories. Experimentally, this phenomenon was confirmed in waveguides in the mode limit. The optical Magnus effect required a theoretical justification in the framework of GO, and a phenomenological theory of this phenomenon was proposed in [5], introducing some additional correction terms in GO equations.

In this work, we demonstrate that the effect of splitting of different polarization rays ensues directly from the equations of classical GO, provided the arising Berry's geometric phases are correctly accounted for. The following argument is principal for this. The fact of the matter is that, as a consequence of their nonlocal nature, Berry phases can grow unlimitedly with growth of the independent variable, thus shifting eigenvalues (Lyapunov indices) of the system [6]. For waves, it means that increment of the Berry phase changes effective refractive indices and phase velocity of waves. These small changes in the refractive indices lead to weak anisotropy of the medium, birefringence, and, as a result, to propagation of eigenmodes along different trajectories.

Results obtained in this work allow one to conclude that the Magnus effect is not a local phenomenon but a fundamental topological effect, just like the Berry phase.

2. THEORY

Let us consider eikonals (phases) of left-hand and right-hand circularly polarized electromagnetic waves in smoothly inhomogeneous isotropic medium, which follow from Maxwell equations [1–3]:

$$\phi^{\pm} = \int_0^s k\,ds \pm \vartheta, \qquad (1)$$

where k is the current wave number, s is the length of ray arc, and ϑ is Berry's geometrical phase (Rytov–Vladimirskiy phase) arising in waves of right and left circular polarization with opposite signs. Here and

---

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

below, the medium's smoothness implies the short-wave asymptotic $k_0 \equiv \omega/c \longrightarrow \infty$, and Eq. (1) corresponds to the first approximation in k_0^{-1}. The Berry phase can be presented as follows [3]:

$$\vartheta = \int_0^s \mathbf{G}\dot{\mathbf{p}}ds = \int_L \mathbf{G}d\mathbf{p}. \qquad (2)$$

Here, $\mathbf{p} = \mathbf{k}/k_0$ is the dimensionless momentum of the wave, $\mathbf{G} = \mathbf{G}(\mathbf{p})$ is certain nonpotential field in $\mathbf{p}$ space, dot denotes differentiation with respect to s (that is, along the beam), and L is a contour along which the system is moving in $\mathbf{p}$ space. Physically, measurable variables relate to curl of field $\mathbf{G}$, which equals [3]

$$\left[\frac{\partial}{\partial \mathbf{p}} \times \mathbf{G}\right] = -\frac{\mathbf{p}}{p^3} \quad (p = |\mathbf{p}|).$$

Geometric phase (2) can be viewed as an integral along the trajectory or as a contour integral in $\mathbf{p}$ space.

The wave refractive indices can be determined from the eikonal equation for (1), (2):

$$n^{\pm} = k_0^{-1}\left|\frac{\partial \phi^{\pm}}{\partial \mathbf{r}}\right| = n_0(\mathbf{r}) + n_1(\mathbf{r}, \mathbf{p}),$$
$$n_1 = \pm k_0^{-1}\frac{d\vartheta}{ds} = \pm k_0^{-1}\mathbf{G}\dot{\mathbf{p}}, \qquad (3)$$

where $n_0(\mathbf{r})$ is the refractive index of isotropic medium under consideration. In traditional GO (see, for example, [7]), the corrective n_1 does not appear, since the eikonal is calculated in the zero approximation in k_0^{-1}. All terms of higher orders are included in transfer equations, i.e., in amplitude of the wave. Meanwhile, the geometrical phase is a *nonlocal* quantity and cannot be attributed to the wave amplitude. As will be shown below, the correction term n_1 leads to changes of GO equations, which are confirmed by experimental results, and, therefore, it has real physical foundations. It is worth noting that, owing to the noticeable dependence of the geometrical phase on the wave trajectory, the corrective n_1 depends not only on the current coordinates $\mathbf{r}$ but also on the *direction* of the wave vector (i.e., wave momentum). This indicates the emerging weak *anisotropy* of locally isotropic medium.

Let us write the Hamiltonian equations for ray propagation [7]. By choosing the Hamiltonian in the form $H = p - n^{\pm}(\mathbf{r}, \mathbf{p}) = 0$, we obtain

$$\frac{d\mathbf{p}}{ds} = \frac{\partial n^{\pm}}{\partial \mathbf{r}} = \frac{\partial n_0}{\partial \mathbf{r}} \pm k_0^{-1}\frac{\partial}{\partial \mathbf{r}}\left(\frac{d\vartheta}{ds}\right), \qquad (4)$$

$$\frac{d\mathbf{r}}{ds} = \mathbf{l} \mp k_0^{-1}\frac{\partial}{\partial \mathbf{p}}\left(\frac{d\vartheta}{ds}\right), \qquad (5)$$

where $\mathbf{l} = \mathbf{p}/p$ is the unit vector of the normal to the wave-phase front; it is also the unit tangent vector of the ray in the zero approximation in k_0^{-1}. The term $\dot{\mathbf{p}}$ in the ray Hamiltonian (see Eq. (3)) should be interpreted not as an independent quantity but only as expressed from the zero order equations.

One can analyze Eqs. (4) and (5) by using the perturbation method with small parameter k_0^{-1}. Then, the first terms of the right-hand sides of these equations correspond to zero approximation and well-known GO equations. The second terms of the right-hand sides of Eqs. (4) and (5) are the corrections on the order of k_0^{-1}, and they should be considered in the zero approximation solutions. In this way, one can find that the second term of the right-hand side of Eq. (4) is responsible for variations of the *phase* velocity *absolute value*, but not direction (see details in [8]). By using Eqs. (1)–(3), phase velocities of the right and left circularly polarized waves can be written as

$$\mathbf{v}_{\text{ph}}^{\pm} = \frac{c}{n_0}\left(1 \mp \frac{1}{n_0 k_0}\frac{d\vartheta}{ds}\right)\mathbf{l} = \frac{c}{n_0}\left(1 \mp \frac{\mathbf{G}\dot{\mathbf{p}}}{n_0 k_0}\right)\mathbf{l}. \qquad (6)$$

The second term on the right-hand side of equation (5) is responsible for deflections of ray trajectory, i.e., for variations of wave *group* velocity. It is shown below that this corrective is directed normally to the ray, and it changes the *direction* of the group velocity. Let us substitute Eq. (2) into Eq. (5) and write the equation for the ray shift $\delta\mathbf{r}$ relative to zero-approximation trajectory:

$$\frac{d\delta\mathbf{r}}{ds} = \mp k_0^{-1}\frac{\partial}{\partial \mathbf{p}}(\mathbf{G}\dot{\mathbf{p}})$$
$$= \mp k_0^{-1}\left[\dot{\mathbf{p}} \times \left[\frac{\partial}{\partial \mathbf{p}} \times \mathbf{G}\right]\right] \mp k_0^{-1}\left(\dot{\mathbf{p}}\frac{\partial}{\partial \mathbf{p}}\right)\mathbf{G}. \qquad (7)$$

Here, we used a well-known vector identity for gradient of scalar production. By integration of Eq. (7), we obtain

$$\delta\mathbf{r}^{\pm} = \mp k_0^{-1}\int_0^s\left\{\left[\dot{\mathbf{p}} \times \left[\frac{\partial}{\partial \mathbf{p}} \times \mathbf{G}\right]\right] + \left(\dot{\mathbf{p}}\frac{\partial}{\partial \mathbf{p}}\right)\mathbf{G}\right\}ds$$
$$= \pm k_0^{-1}\int_0^s \frac{[\dot{\mathbf{p}} \times \mathbf{p}]}{p^3}ds \mp k_0^{-1}\mathbf{G}\big|_{\mathbf{p}_0}^{\mathbf{p}}, \qquad (8)$$

where the equation $\left[\frac{\partial}{\partial \mathbf{p}} \times \mathbf{G}\right] = -\frac{\mathbf{p}}{p^3}$ [3] was used and $\mathbf{p}_0 = \mathbf{p}(0)$.

The ray shifting (8) contains two summands. The first one is nonlocal and can grow unlimitedly with the growth of s. The second term is a local function of the momentum $\mathbf{p}$. It cannot grow unlimitedly and, by the order of value, does not exceed the wavelength $\lambda \sim k_0^{-1}$. Obviously, the second term cannot lead to any observable trajectory effects. It should also be noted that the field **G** in the theory of geometrical phases is not defined uniquely, but only to a gauge transformation $\mathbf{G} \longrightarrow \mathbf{G} + \partial\varphi/\partial\mathbf{p}$, where $\varphi(\mathbf{p})$ is an arbitrary scalar potential. This arbitrariness does not influence the first summand in Eq. (8) but changes the second one. This circumstance is connected to the ambiguity of the definition of "ray trajectory" concept in a range of the wavelength. Note also that, if the current direction of the ray coincides with the initial one, $\mathbf{p} = \mathbf{p}_0$, all local terms turn to zero.

Thus, to study ray shifting, we should use only the first term in Eq. (8). Then, we come to the following result:

$$\delta\mathbf{r}^{\pm} = \pm k_0^{-1}\int_0^s \frac{[\dot{\mathbf{p}}\times\mathbf{p}]}{p^3}ds = \mp k_0^{-1}\int_L \frac{[\mathbf{p}\times d\mathbf{p}]}{p^3}. \qquad (9)$$

As one can see, the ray shifting is orthogonal to the ray, $\mathbf{p}d\delta\mathbf{r}^{\pm}/ds = 0$. Equation (9) shows that splitting of the rays, induced by optical Magnus effect, as well as Berry phase (2), can be presented in the form of a contour integral in **p** space. Therefore, the optical Magnus effect is a fundamental *topological* phenomenon. The Berry phase and the optical Magnus effect represent phase and trajectory divergence of the waves, respectively.

The following equation, similar to Eq. (5), corresponds to shift (9):

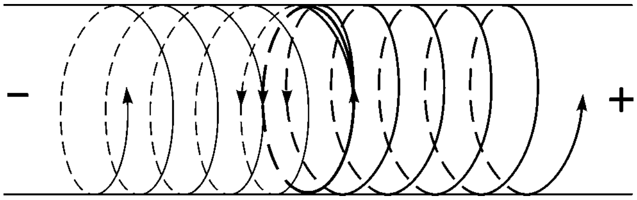
$$\frac{d\mathbf{r}}{ds} = \mathbf{l} \pm \frac{1}{k_0 p^2}[\dot{\mathbf{p}}\times\mathbf{l}] = \mathbf{l} \pm \frac{1}{k_0 n_0}\left[\frac{\partial \ln n_0}{\partial \mathbf{r}}\times\mathbf{l}\right], \qquad (10)$$

where we substituted $\dot{\mathbf{p}} = \partial n_0/\partial\mathbf{r}$ from the zero approximation of Eq. (4) and used $p \cong n_0$ from the connection $H = 0$. The expressions for group velocities of right and left circularly polarized waves follow from Eq. (10):

$$\begin{aligned}\mathbf{v}_g^{\pm} &= \frac{c}{n_0}\left(\mathbf{l} \pm \frac{1}{k_0 p^2}[\dot{\mathbf{p}}\times\mathbf{l}]\right)\\ &= \frac{c}{n_0}\left(\mathbf{l} \pm \frac{1}{k_0 n_0}\left[\frac{\partial \ln n_0}{\partial \mathbf{r}}\times\mathbf{l}\right]\right).\end{aligned} \qquad (11)$$

This equation indicates that group velocities of right and left waves in the given approximation are equal in absolute value, $|\mathbf{v}_g^{+}| = |\mathbf{v}_g^{-}| + O(k_0^{-2})$, but deflect in opposite sides from the zero-approximation ray.

Equation (10) corresponds exactly to equation that was obtained by Liberman and Zeldovich in [5]. This paper shows that the equation accurately describes the optical Magnus effect in a circular waveguide, in accordance with experiments [4]. Meanwhile, there exist fundamental differences between theory [5] and this work. First of all, attention may be drawn to the difference in the Hamiltonian for the waves in [5] and convenient GO Hamiltonian with refractive indices (3) of our work. It is evident that theory [5] involves many excess variables, which complicate the problem. In essence, the wave polarization was treated there as an independent dynamical variable with a *continuous* spectrum of values. At the same time (see [8]), only the right-hand and left-hand circularly polarized waves are the independent eigenmodes in the first approximation of GO (this follows already from their invariance relative to Rytov's polarization evolution). All other waves are only the superposition of circular eigenmodes. This corresponds to the *quantization of photon helicity* [9]. Therefore, the polarization (spin) variable for electromagnetic waves has only two eigenvalues, ±1, and this fact is reflected in the present work. As a consequence of this fact, an arbitrary wave of mixed polarization splits into two waves of right-hand and left-hand circular polarizations when propagating through an inhomogeneous medium. This phenomenon of wave splitting into two independent modes is absent in [5] (for example, according the theory [5], the trajectory of a linearly polarized wave remains unchanged).



Splitting of a finite ray of mixed (noncircular) polarization into two eigenrays of right and left circular polarizations.

An example of polarization splitting of a wave is presented in the figure. There, we can see the ray trajectories of wave propagation in an oversized circular waveguide. The initial linearly polarized finite ray (thick circular line) splits into two rays of right-hand (+) and left-hand (–) polarizations, which are gradually transported along the waveguide in opposite directions. A detailed consideration of this example and of optical Magnus effect for meridional rays from the topological viewpoint is given in [8]. The experiment presented in the figure can be performed in a cylindrical structure by exciting an oriented ray associated with the whispering gallery mode. This experiment will enable one to observe for the first time the *topological spin splitting of photons* (analogous effects for nonrelativistic particles are under intensive consideration now; see, for instance, [10] and references therein).

3. CONCLUSIONS

We have obtained GO ray equations of a smoothly inhomogeneous isotropic medium in the first approximation in k_0^{-1} from the first principles, such as the eikonal equation and the Hamiltonian approach. As the starting point, we have used the expression for wave eikonals, which were obtained many times before [1–3] from the Maxwell equations. The shift in wave eikonals introduced by arising geometric phases means that phase velocities of right-hand and left-hand polarized waves are different, that is, their refractive indices are different (3).[1] Correction terms for the refraction indices depend not only on the current coordinates of the space point but also on the direction of the wave vector of the wave, which implies weak anisotropy of the medium. In essence, the account for the terms of the order k_0^{-1}, proportional to $\partial n_0/\partial \mathbf{r}$, removes the polarization degeneracy of a homogeneous medium and makes it weakly anisotropic. The independent eigenmodes in this medium are the waves of right-hand and left-hand circular polarizations.

Hamiltonian ray equations (4), (5) and (10), written for refraction indices (3) contain correction terms that deflect the trajectories of right-hand and left-hand polarized waves in different directions. This means that a ray of an arbitrary mixed polarization splits during its propagation into two rays of right-hand and left-hand polarization. The deflections of differently polarized rays are proportional to the derivative of the Berry phase by direction, that is, in fact, to the anisotropy of the geometric phase. Thus, splitting of differently polarized rays (9) (the optical Magnus effect) is also a nonlocal topological effect, defined only by the system trajectory in **p** space. It follows from this, in particular, that, in a one-dimensionally inhomogeneous medium (where all ray trajectories are plane and the Berry phase is absent), the ray splitting does not take place.

We conclude that the Berry phase and the so-called optical Magnus effect are accompanying phenomena existing in the same order of k_0^{-1} in equations of geometric optics. The Berry phase implies a difference in the *absolute values of phase velocities* of right and left circular waves, while the Magnus effect ensues from the difference in the *directions of group velocities*.

We are grateful to Yu.A. Kravtsov and V.A. Permyakov for helpful discussions and taking interest in this research. This work was supported in part by INTAS (grant no. 03-55-1921).

REFERENCES

1. S. M. Rytov, Dokl. Akad. Nauk SSSR **18**, 263 (1938).
2. V. V. Vladimirskiĭ, Dokl. Akad. Nauk SSSR **31**, 222 (1941).
3. S. I. Vinitskiĭ, V. L. Debrov, V. M. Dubovik, *et al.*, Usp. Fiz. Nauk **160** (6), 1 (1990) [Sov. Phys. Usp. **33**, 403 (1990)]; *Geometric Phases in Physics*, Ed. by A. Shapere and F. Wilczek (World Sci., Singapore, 1989).
4. B. Ya. Zel'dovich and V. S. Liberman, Kvantovaya Élektron. (Moscow) **17**, 493 (1990); A. V. Dugin, B. Ya. Zel'dovich, N. D. Kundikova, and V. S. Liberman, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 186 (1991) [JETP Lett. **53**, 197 (1991)]; Zh. Éksp. Teor. Fiz. **100**, 1474 (1991) [Sov. Phys. JETP **73**, 816 (1991)]; A. V. Dooghin, N. D. Kundikova, V. S. Liberman, and B. Ya. Zel'dovich, Phys. Rev. A **45**, 8204 (1992).
5. V. S. Liberman and B. Ya. Zel'dovich, Phys. Rev. A **46**, 5199 (1992).
6. K. Yu. Bliokh, J. Math. Phys. **43**, 25 (2002); J. Math. Phys. **43**, 5624 (2002).
7. Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Nauka, Moscow, 1980) [in Russian].
8. K. Yu. Bliokh and Yu. P. Bliokh, physics/0402014 (2004).
9. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Relativistic Quantum Theory* (Nauka, Moscow, 1968; Pergamon, Oxford, 1971), Vol. 1.
10. F. Zhou, cond-mat/0311612 (2003).

---

[1] The conclusion about the difference in the phase velocities of the right- and left-polarized waves was made as early as in Rytov's pioneering work [1]. However, the corresponding formula presented in that work is incorrect, because the calculations carried out in the noninertial reference system associated with the rotating Frenet trihedron of the ray.

JETP Letters, Vol. 79, No. 11, 2004, pp. 523–528. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 651–656.
Original English Text Copyright © 2004 by Shevchenko.

On the Maximum Lyapunov Exponent of the Motion in a Chaotic Layer[¶]

I. I. Shevchenko
Pulkovo Observatory, Russian Academy of Sciences, St. Petersburg, 196140 Russia
e-mail: iis@gao.spb.ru

Received April 5, 2004; in final form, May 5, 2004

The maximum Lyapunov exponent (referred to the mean half-period of phase libration) of the motion in the chaotic layer of a nonlinear resonance subject to symmetric periodic perturbation, in the limit of infinitely high frequency of the perturbation, has been numerically estimated by two independent methods. The newly derived value of this constant is 0.80, with precision presumably better than 0.01. © *2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 05.45.Pq; 45.20.Jj

1. INTRODUCTION

On the basis of results of extended numerical experiments, Chirikov [1, 2] noted that the maximum Lyapunov exponent, referred to the mean half-period of phase libration, of the motion in the chaotic layer of a nonlinear resonance subject to symmetric periodic perturbation is approximately constant in a wide range of a parameter characterizing the perturbation frequency. In this paper, we estimate the least upper bound for the maximum Lyapunov exponent of the separatrix map. We show that this bound coincides with the value of the maximum Lyapunov exponent in the mentioned problem in the limit of infinitely high frequency of perturbation, and its value does not depend on the amplitude of the perturbation, i.e., it is defined robustly. In what follows, this quantity is called Chirikov's constant. The knowledge of the value of Chirikov's constant is important for accurate analytical estimation of Lyapunov exponents in applications in mechanics and physics [3, 4].

Nonlinear resonances are ubiquitous in problems of modern mechanics and physics. Under general conditions [5, 2, 6], a model of a nonlinear resonance is provided by a nonlinear pendulum with periodic perturbations. A rigid pendulum with an oscillating suspension point is a paradigm in studies of nonlinear resonances and chaotic behavior in Hamiltonian dynamics. The Hamiltonian of this system, according to, e.g., [7], is

$$H = \frac{\mathcal{G}p^2}{2} - \mathcal{F}\cos\varphi + a(\cos(\varphi-\tau) + \cos(\varphi+\tau)), \tag{1}$$

where $\tau = \Omega t + \tau_0$. The first two terms in Eq. (1) represent the Hamiltonian H_0 of the unperturbed pendulum, while the two remaining ones represent the periodic perturbations.

The variable φ is the pendulum angle (this angle measures the deviation of the pendulum from the lower position of equilibrium), and τ is the phase angle of perturbation. The quantity Ω is the perturbation frequency, and τ_0 is the initial phase of the perturbation; p is the momentum; and $\mathcal{F}$, $\mathcal{G}$, and a are constants. In what follows, it is assumed that $\mathcal{F} > 0$, $\mathcal{G} > 0$.

Chirikov [5, 2] derived the so-called separatrix map describing the motion in the vicinity of the separatrices of Hamiltonian (1):

$$\begin{aligned} w_{i+1} &= w_i - W\sin\tau_i, \\ \tau_{i+1} &= \tau_i + \lambda\ln\frac{32}{|w_{i+1}|} \pmod{2\pi}, \end{aligned} \tag{2}$$

where w denotes the relative (with respect to the separatrix value) pendulum energy $w = H_0/\mathcal{F} - 1$ and τ retains its meaning of the phase angle of perturbation. Constants λ and W are parameters: λ is the ratio of Ω, the perturbation frequency, to $\omega_0 = (\mathcal{F}\mathcal{G})^{1/2}$, the frequency of the small-amplitude pendulum oscillations; and

$$W = \frac{a}{\mathcal{F}}\lambda(A_2(\lambda) + A_2(-\lambda)) = 4\pi\frac{a}{\mathcal{F}}\lambda^2\operatorname{cosech}\frac{\pi\lambda}{2}, \tag{3}$$

where

$$A_2(\lambda) = 4\pi\lambda\frac{\exp\frac{\pi\lambda}{2}}{\sinh(\pi\lambda)} \tag{4}$$

is the Melnikov–Arnold integral [2, 6, 8]. One iteration of map (2) corresponds to one period of the pendulum rotation, or a half-period of its libration. The motion of system (1) is mapped by Eqs. (2) asynchronously [8]:

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

0021-3640/04/7911-0523$26.00 © 2004 MAIK "Nauka/Interperiodica"

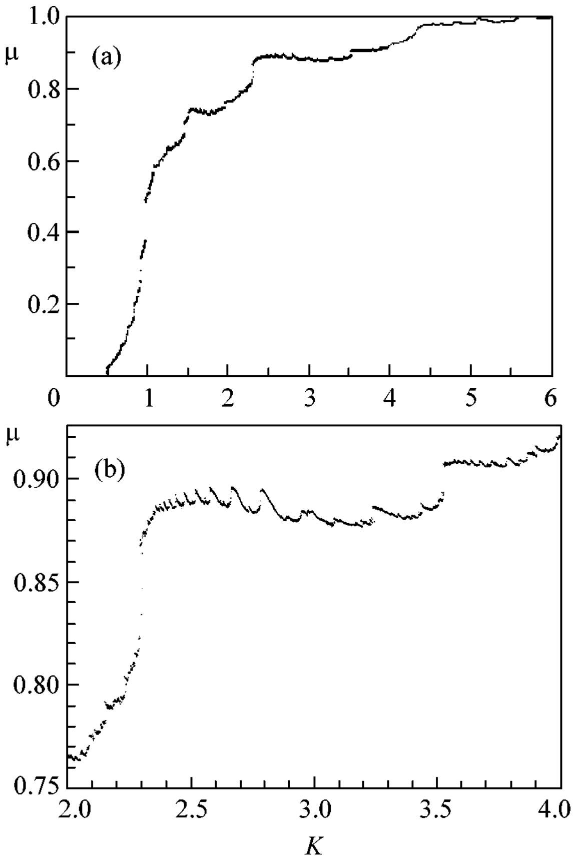


Fig. 1. (a) The dependence $\mu(K)$; (b) a detail enlarged.

the action-like variable w is taken at $\varphi = \pm\pi$, while the perturbation phase τ is taken at $\varphi = 0$. The desynchronization can be removed by a special procedure [8, 9].

An equivalent form of Eqs. (2), used, e.g., in [10], is

$$\begin{aligned} y_{i+1} &= y_i + \sin x_i, \\ x_{i+1} &= x_i - \lambda \ln|y_{i+1}| + c \quad (\mathrm{mod}\ 2\pi), \end{aligned} \tag{5}$$

where $y = w/W$, $x = \tau + \pi$; and the new parameter

$$c = \lambda \ln \frac{32}{|W|}. \tag{6}$$

The standard map represents linearization of the separatrix map in the action-like variable y near a fixed point; it is given by the equations

$$\begin{aligned} y_{i+1} &= y_i + K \sin x_i \quad (\mathrm{mod}\ 2\pi), \\ x_{i+1} &= x_i + y_{i+1} \quad (\mathrm{mod}\ 2\pi), \end{aligned} \tag{7}$$

where K is the so-called stochasticity parameter [1, 2].

2. LYAPUNOV EXPONENTS AND DYNAMICAL ENTROPY

The calculation of the Lyapunov characteristic exponents (LCEs) is one of the most important tools in the study of the chaotic motion. The LCEs characterize the rate of divergence of trajectories close to each other in phase space. A nonzero LCE indicates a chaotic character of motion, while the maximum LCE equal to zero is the signature of regular (periodic or quasiperiodic) motion. The quantity reciprocal to the maximum LCE characterizes the motion predictability time.

Let us consider two trajectories close to each other in phase space. One of them we shall refer to as *guiding*, and the other, as *shadow*. Let $d(t_0)$ be the length of the displacement vector directed from the guiding trajectory to the shadow one at an initial moment $t = t_0$. The LCE is defined by the formula [6]

$$L = \limsup_{\substack{t \to \infty \\ d(t_0) \to 0}} \frac{1}{t - t_0} \ln \frac{d(t)}{d(t_0)}.$$

In the case of a Hamiltonian system, the quantity L may take $2N$ different values (depending on the direction of the initial displacement), where N is the number of degrees of freedom; the LCEs divide into pairs: for each $L_k > 0$, there exists $L_{k+N} = -L_k < 0$, $k = 1, \ldots, N$.

The LCEs are closely related to the dynamical entropy [11, 12, 1, 2, 13]. For the Hamiltonian systems with 3/2 and 2 degrees of freedom, Benettin *et al.* proposed the relation $h = L\mu$ [12, Eq. (6)], where h is the dynamical entropy, L is the maximum LCE, and μ is the relative measure of the connected chaotic domain where the motion takes place. This formula is approximate. Benettin *et al.* [12] applied it in a study of the chaotic motion of the Hénon–Heiles system.

In what follows, our numerical data is presented on the measure μ of the main connected chaotic domain in phase space of the standard map, the maximum LCE L, and the product of μ and L for motion in this domain. Two methods for computation of the chaotic domain measure μ are used. A traditional "one-trajectory method" (OTM) consists in computing the number of cells explored by a single trajectory on a grid exposed on phase plane. A "current LCE segregation method" (CLSM) is based on an analysis of the differential distribution of the computed values of the Lyapunov exponents (current LCEs) of a set of trajectories with starting values on a grid on the phase plane. Both methods were proposed and used by Chirikov [1, 2] in computations of μ for the standard map. Analogous methods were used in [14] in computations of chaotic domain measure in the Henon–Heiles problem.

Figure 1 illustrates discontinuity of the obtained $\mu(K)$ function. The curve in Fig. 1 is obtained by the OTM on the grid 2000 × 2000 pixels on the phase plane $(x, y) \in [0, 2\pi] \times [0, 2\pi]$, and the number of iterations $n_{it} = 10^8$. A prominent bump in the dependence, shown in detail in Fig. 1b, is conditioned by the process of disintegration of the half-integer resonance in the course of a sequence of period-doubling bifurcations while K is increasing from ≈2 to ≈2.5. A similar but less pronounced bump is seen in Fig. 1a in the range $4 < K <$

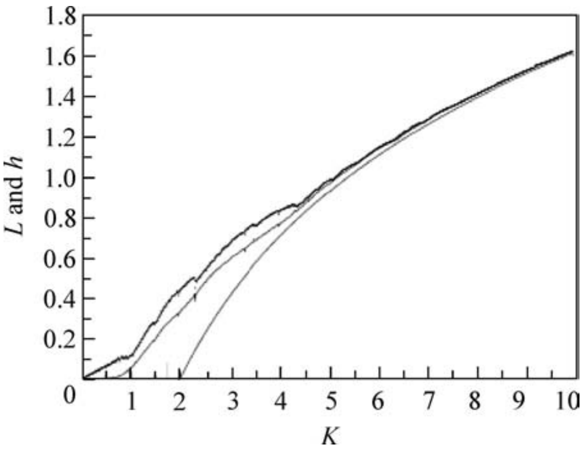


Fig. 2. $L(K)$ (the uppermost curve), $h(K)$, and $\ln(K/2)$ in a broad range of K.

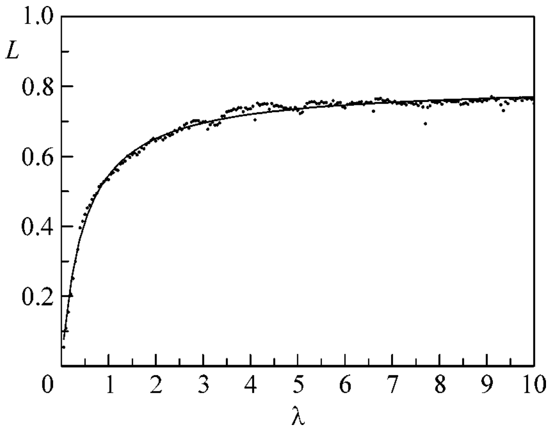


Fig. 3. The dependence $L(\lambda)$ for the separatrix map (the case of the least perturbed border of the chaotic layer) and its rational approximation.

4.5; this one is due to bifurcations of the integer resonance.

In general, at these moderate values of K (at $K < 6$), the discontinuities are conditioned by the process of absorption of minor chaotic domains by the main chaotic domain, while K increases.

In Fig. 2, we give the plot of the maximum LCE and the dynamical entropy in a broad range of K. Each value of L in Fig. 2 represents the mean value of the maximum LCEs over 10 trajectories of length $n_{it} = 10^7$ each. The initial data for these trajectories slightly differ from each other but in all cases are chosen to lie inside the main chaotic domain. Everywhere in this work (including the case of the separatrix maps considered below), the presented values of LCEs have been computed by the tangent map method described, e.g., in [1, 2].

The corresponding dependence of $h = L\mu$ on K is plotted in Fig. 2; the values of μ obtained by the OTM (Fig. 1) have been used. One can see that our numerical experiments suggest that the dynamical entropy of the standard map is continuous and monotonous in K, contrary to the discontinuous behavior of LCEs. This is no wonder, since the dynamical entropy is a more fundamental quantity.

For orientation, the function $\ln(K/2)$ is depicted in the same Fig. 2; this is the well-known logarithmic law derived by Chirikov [1, 2] analytically by means of averaging the largest eigenvalue of the tangent map in assumption that the relative measure of the regular component is small.

The downward spikes seen in the $L(K)$ dependence in Fig. 2 (and in Fig. 3 also) represent a manifestation of the so-called "stickiness effect" immanent to the chaotic Hamiltonian dynamics in conditions of divided phase space [15]: a chaotic trajectory may stick for a long time to the borders of the chaotic domain, where the motion is close to regular, and, therefore, the local LCEs are small. Since the computation time is always finite, the stickiness effect, in the case of deep stickings, leads to underestimated values of LCEs; see discussion in [15].

The $L(K)$ and $\mu(K)$ functions are discontinuous and evidently elude simple analytic representation in broad ranges of K. The case of $h(K)$ is different. With good accuracy, the numerical dependence at K greater than its critical value can be approximated by a function which is piecewise linear at moderate values of K (see Fig. 2) and logarithmic with a small power-law correction at higher values of K: $h(K) = \ln(K/2) + 1/K^2$ if $K > 4.5$, with accuracy better than 0.01 in absolute magnitude. So, the presented numerical data indicates that the high-K asymptotics of the $h(K)$ function contains a power-law component, in addition to the well-known logarithmic one. The same is valid for the $L(K)$ dependence if one ignores the small (and local in K) distortions of the function due to accelerator modes and periodic solutions of higher orders.

Let us consider in more detail the accuracy of the presented results. The LCE values can be effectively verified by controlling their saturation, taking various values of n_{it}. We compare the maximum LCEs computed taking $n_{it} = 10^7$ with those computed taking the number of iterations ten times less, $n_{it} = 10^6$. In both cases we average over ten trajectories. One finds that the difference between the LCEs in these two cases, averaged over the interval $1 \le K < 2$ (the step in K is 0.01, so 100 differences are averaged), is equal to only 0.0015. The saturation is faster with increasing K. Therefore, the saturation at $n_{it} = 10^7$ is practically complete at $K > 1$, and, therefore, there are no significant systematic errors in the determination of LCEs at such values of K.

The problem of accuracy of computation of μ is more difficult. The generic border of chaos in phase space of Hamiltonian systems is fractal [16, 17].

Umberger and Farmer [16] conjectured and numerically verified that the coarse-grained measure of the chaotic constituent of phase space of two-dimensional area-preserving map scales with the grid resolution ε, employed to estimate the measure, as a power law in the limit $\varepsilon \longrightarrow 0$:

$$\mu_\varepsilon = \mu_0 + A\varepsilon^\beta, \tag{8}$$

where μ_0 is the actual measure and A and β are constants characterizing the border; A, $\beta \geq 0$.

Numerical experiments provide values of μ_ε for a given ε. So, there are three undetermined quantities in Eq. (8): μ_0, A, and β. To obtain their numerical values, one needs to compute μ_ε at least thrice, i.e., at three different resolutions ε of the grid. Then, the system of three nonlinear equations (8) can be solved. We take three partitions of phase plane of the standard map: 2500×2500, 5000×5000, and 7500×7500 pixels; i.e., $\varepsilon = 1/2500$, $1/5000$, and $1/7500$. As in [16], the OTM is used; $n_{it} = 10^{10}$. The minimum ε and maximum n_{it} used in [16] were 1/4096 and 10^8, respectively. Thus, we should expect better estimates of the chaotic domain measure in the present study.

In our numerical experiment, we have computed μ_0 and β for ten values of K equally spaced in the interval [1.0, 5.5] and for ten values of K equally spaced in the interval [1.0, 1.9]. At all points, the computed value of $\mu_{1/7500}$ (μ_ε at the smallest $\varepsilon = 1/7500$) and the resulting actual μ_0 value differ by no more than 0.01; at $K > 1.5$, they coincide with an accuracy of two significant digits.

Our data for $K = 1.1$, 1.2, and 1.3 can be compared to the results of Umberger and Farmer [16], who computed the chaotic domain measure for these three values of K. The difference in their values of μ_0 and ours does not exceed ≈0.01. Our results on the chaotic domain measure are also in reasonable agreement with early estimates by Chirikov [1, 2]. The close proximity of the values of μ_0 to those of $\mu_{1/7500}$, as well as the agreement of them with the results [1, 2, 12], testifies that our estimates of the chaotic domain measure have an accuracy better than 0.01.

The power-law index β is related to the fractal dimension d_L of the set of all chaos borders inside the connected chaotic domain [17]: $d_L = 2 - \beta$. From our data, one has $\langle\beta\rangle = 0.63 \pm 0.13$ and $d_L \approx 1.37 \pm 0.13$. This agrees well with the theoretical estimate $d_L = 3/2$ by Chirikov [17].

An important constant of the standard map dynamics is the chaotic domain measure at the critical value of the stochasticity parameter $K = K_G = 0.971635406\ldots$ (on the critical value, see, e.g., [13]). Our calculation performed by the same algorithm as presented above gives $\mu(K_G) \approx 0.463$. The contribution of the chaotic domain around the integer resonance to this quantity is 3.5 times greater than that of the half-integer one ($0.463 \approx 0.359 + 0.103$). The calculated values of the parameter β in these domains are equal to ≈0.53 and ≈0.49, respectively; so, the border fractal dimension $d_L = 2 - \beta$ in the critical case $K = K_G$ is particularly close, as one could expect, to the theoretical estimate $d_L = 3/2$ by Chirikov [17].

3. ESTIMATION OF CHIRIKOV'S CONSTANT

The value of Chirikov's constant can be found by calculating the average of the local maximum LCE over the chaotic layer of map (5) in the limit $\lambda \longrightarrow \infty$. The local LCE must be taken with weight directly proportional to the time that the trajectory spends in a given part of the layer; this time is directly proportional to the local relative measure of the chaotic component. Therefore, one has the following formula:

$$C_h = \lim_{\lambda \to \infty} \frac{\int_0^{y_b} \tilde{L}_{sx}(y)\tilde{\mu}_{sx}(y)dy}{\int_0^{y_b} \tilde{\mu}_{sx}(y)dy}, \tag{9}$$

where $y_b = \lambda/K_G$ is the value of y at the border of the layer, $\tilde{L}_{sx}(y)$ is the local (with respect to y) value of the maximum LCE of the separatrix map, and $\tilde{\mu}_{sx}(y)$ is the local chaos measure. The tilde marks that the quantities are local. This formula is valid in the limit $\lambda \longrightarrow \infty$, since only in this limit can one reduce the sum over all integer resonances inside the layer to an integral. Moreover, the formula $y_b = \lambda/K_G$ (see [17, 8]) is accurate also only in this limit.

By means of the substitution $y = \lambda/K$ we introduce a new independent variable K, which is nothing but the value of the stochasticity parameter of the standard map locally approximating the separatrix map. The accuracy of the approximation improves with increasing λ [1, 2]; i.e., the local characteristics of the chaotic layer converge to those of the standard map locally approximating the motion: $\tilde{L}_{sx}(y = \lambda/K) \longrightarrow L(K)$ and $\tilde{\mu}_{sx}(y = \lambda/K) \longrightarrow \mu(K)$, where $L(K)$ and $\mu(K)$ are the maximum LCE and the measure of the main connected chaotic domain of the standard map in the function of K. The dependence on λ in the limit $\lambda \longrightarrow \infty$ is eliminated, and Eq. (9) is reduced to the final form

$$C_h = \frac{K_G}{\sigma} \int_{K_G}^{\infty} L(K)\mu(K)\frac{dK}{K^2}, \tag{10}$$

where the quantity

$$\sigma = \lim_{\lambda \to \infty} y_b^{-1} \int_0^{y_b} \tilde{\mu}_{sx}(y)dy = K_G \int_{K_G}^{\infty} \mu(K)\frac{dK}{K^2} \tag{11}$$

has the meaning of "porosity" of the chaotic layer. This is the ratio of the area of the chaotic component to the total area of the layer bounded by its external borders; the quantity $1 - \sigma$ is nothing but the total relative area of all regular islands inside the layer.

For $K \in [K_G, 10]$, we integrate Eqs. (10) and (11) numerically; the functions $L(K)$ and $\mu(K)$ are taken in numerical form, as presented in Figs. 1 and 2. The remainders for $K > 10$ are calculated analytically, with $h(K) = L(K)\mu(K)$ set equal to $\ln(K/2) + 1/K^2$ (as established above) and $\mu(K)$ set equal to unity. Adopting an accuracy of two significant digits, one has $C_h = 0.80$, $\sigma = 0.78$.

Our value of C_h differs significantly from the value of 0.663 got by Chirikov by integration of the dynamical entropy of the standard map in [1, 2] (where C_h is designated as h_W). This deviation is due to the sparsity of the numerical data obtained more than twenty years ago, as well as to ignoring the porosity of the chaotic layer in an approximate calculation in [1, 2].

As discussed in the previous section, the numerical dependence $\mu(K)$ is less certain than $L(K)$. Hence its uncertainty is the most likely source of errors in estimating C_h. The estimated error in determination of $\mu(K)$ does not exceed 0.01 (see above). At high values of K (at $K > 6$), the deviations are much less than 0.01, because $\mu(K)$ rapidly converges to unity. To estimate the accuracy of the obtained value of C_h, we recompute this value substituting $\mu(K) \pm 0.01$ instead of the original $\mu(K)$ in Eqs. (10) and (11) at $K \le 10$ (if $\mu(K) + 0.01 > 1$ we set $\mu(K) = 1$, of course). We find that both these negative and positive shifts in $\mu(K)$ change the resulting value of C_h by no more than 0.004. Therefore, if one takes three significant digits in C_h, the result is $C_h = 0.801 \pm 0.004$ in the described sense. The deviations in σ are greater: $\sigma = 0.780 \pm 0.009$. Finally, rounding up, we conclude that the accuracy of our estimate $C_h = 0.80$ is presumably better than 0.01.

One can verify this estimate of Chirikov's constant by means of a straightforward computation of the maximum LCE of the separatrix map at high values of λ. The λ dependence of the maximum LCE of the separatrix map is shown in Fig. 3. It has been obtained by a numerical experiment with map (5). The use of one and the same designation L for the LCE in both cases of the standard and separatrix maps should not cause confusion. The resolution (step) in $\lambda \in [0, 10]$ is equal to 0.05. At each step in λ, the values of L have been computed for 100 values of c equally spaced in the interval $[0, 2\pi]$. The number of iterations for each trajectory is $n_{it} = 10^7$ for $\lambda \in [0, 1)$ and $n_{it} = 10^8$ for $\lambda \in [1, 10]$ (the saturation time for numerical estimates of LCEs increases with λ and, therefore, more computational time is needed to get reliable values of LCEs at high values of λ). This is sufficient to saturate the computed values of L; see below. At each step in λ, we find the value of c corresponding to the minimum width of the layer (the case of the least perturbed border) and plot the value of L corresponding to this case. The case of the least perturbed border is generic in applications in the sense that strong perturbations of the border are local in c. Note that the parameter c is related to the amplitude of the periodic perturbation in Hamiltonian (1) at a given value of λ by Eq. (6).

Figure 3 also presents approximation of the observed dependence by the rational function

$$L(\lambda) = \frac{b + c\lambda}{1 + a\lambda} \tag{12}$$

with b set to zero in order that $L(0) = 0$. The resulting values of the parameters and their standard errors are $a = 2.097 \pm 0.033$, $b = 0$, and $c = 1.691 \pm 0.024$.

Chirikov's constant is given by the limit $L(\lambda \longrightarrow \infty)$; so, according to the described numerical experiment, $C_h \approx 0.806$, in good agreement with the result presented above ($C_h \approx 0.801$).

The integration time we used is sufficient for effective saturation of the computed LCE in the given interval of λ. Indeed, setting $n_{it} = 10^7$ for the whole interval $\lambda \in [0, 10]$ gives the resulting $C_h \approx 0.808$, i.e., the resulting C_h value is negligibly different from the value obtained with n_{it} raised to 10^8 at $\lambda \in [1, 10]$.

Variation of the parameter c in map (5) produces a scatter in the computed values of LCE due to emergence and disappearance of marginal resonances at the border of the chaotic layer (on the marginal resonances see [8]). Let us prove that the LCE scatter tends to zero in the limit $\lambda \longrightarrow \infty$; in other words, the limit of LCE is one and the same for all (though sufficiently small, for the map description to be valid) amplitudes of perturbation.

The largest variations are conditioned by integer marginal resonances. The y coordinates of the centers of the integer resonances satisfy the relation $y_{i+1}/y_i = \exp(2\pi/\lambda)$, where i is the number of the resonance. This relation follows from the second line of map (5). At the border of the layer $y \approx \lambda$ (see [1, 2]); therefore, in the case of $\lambda \gg 1$, the distance between the centers of two consecutive integer resonances at the border is $\Delta y \approx 2\pi$. The relative local measure μ_{marg} of the chaotic component associated with the separatrices of a marginal resonance depends on the value of the parameter K of the standard map locally approximating the motion near the marginal resonance; the maximum value is ≈ 0.46 at $K = K_G$ (see the concluding paragraph of Section 2), because any marginal resonance has $K < K_G$. This relative measure, referred to the total chaotic measure of the layer, is less than approximately $(\mu_{\text{marg}} \times 2\pi)/(\sigma\lambda) \approx 3.7/\lambda$. The largest value of the local LCE at the border is again associated with $K = K_G$; it equals approximately 0.11 (see Fig. 2). So, the contribution of the chaotic layer of the marginal resonance to the total value of the maximum LCE over the whole layer varies from zero up to $\approx 0.4/\lambda$, depending on the prominence of the marginal resonance, i.e., on the local value of K at the border. This contribution tends to zero with $\lambda \longrightarrow \infty$ and, therefore, the value of Chirikov's constant does not depend on the second parameter of the separatrix map, c, and, consequently, on the amplitude of the periodic perturbation.

4. CONCLUSIONS

Exploiting our high-precision data on the functions $h(K)$ and $\mu(K)$, we have calculated the value of Chirikov's constant C_h—the least upper bound for the maximum Lyapunov exponent of the separatrix map. This quantity is nothing but the maximum Lyapunov exponent (referred to the mean half-period of phase libration, or, equivalently, the mean period of phase rotation) of the motion in the chaotic layer of a nonlinear resonance subject to symmetric periodic perturbation in the limit of infinitely high frequency of the perturbation. We have shown that the value C_h does not depend on the second parameter of the separatrix map (or, equivalently, on the amplitude of the perturbation).

The newly derived value of C_h is 0.80, with precision presumably better than 0.01. The knowledge of this constant is important for correct analytical estimation of the value of the maximum Lyapunov exponent of the chaotic motion of a Hamiltonian system allowing description in the perturbed pendulum model.

The author is thankful to B.V. Chirikov and V.V. Vecheslavov for valuable discussions. This work was supported by the Russian Foundation for Basic Research (project no. 03-02-17356).

REFERENCES

1. B. V. Chirikov, *Interaction of Nonlinear Resonances* (Novosib. Gos. Univ., Novosibirsk, 1978) [in Russian].
2. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
3. I. I. Shevchenko, Izv. Gl. Astron. Obs. Pulkove **214**, 153 (2000).
4. I. I. Shevchenko, Kosm. Issled. **40**, 317 (2002) [Cosmic Res. **40**, 296 (2002)].
5. B. V. Chirikov, *Nonlinear Resonance* (Novosib. Gos. Univ., Novosibirsk, 1977) [in Russian].
6. A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics*, 1st ed. (Springer, New York, 1992; Mir, Moscow, 1984).
7. B. S. Bardin and A. P. Markeev, Prikl. Mat. Mekh. **59**, 922 (1995).
8. I. I. Shevchenko, Phys. Scr. **57**, 185 (1998).
9. I. I. Shevchenko, Zh. Éksp. Teor. Fiz. **118**, 707 (2000) [JETP **91**, 615 (2000)].
10. B. V. Chirikov and D. L. Shepelyansky, Physica D (Amsterdam) **13**, 395 (1984).
11. Ya. B. Pesin, Dokl. Akad. Nauk SSSR **226**, 774 (1976).
12. G. Benettin, L. Galgani, and J. M. Strelcyn, Phys. Rev. A **14**, 2338 (1976).
13. J. D. Meiss, Phys. Rep. **64**, 795 (1992).
14. I. I. Shevchenko and A. V. Melnikov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 772 (2003) [JETP Lett. **77**, 642 (2003)].
15. I. I. Shevchenko, Phys. Lett. A **241**, 53 (1998).
16. D. K. Umberger and J. D. Farmer, Phys. Rev. Lett. **55**, 661 (1985).
17. B. V. Chirikov, INP Preprint No. 90–109 (1990).

JETP Letters, Vol. 79, No. 11, 2004, pp. 529–531. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 657–659.
Original Russian Text Copyright © 2004 by Ashitkov, Ovchinnikov, Agranat.

Recombination of an Electron–Hole Plasma in Silicon under the Action of Femtosecond Laser Pulses

S. I. Ashitkov, A. V. Ovchinnikov, and M. B. Agranat

Institute of High Energy Densities, Associated Institute for High Temperatures, Russian Academy of Sciences, Moscow, 125412 Russia

Received March 31, 2004

Experimental data are obtained on the dynamics of conduction-electron relaxation at the stage preceding the melting of a silicon surface layer. The energy of a quantum of probe radiation is smaller than the band gap, making it possible to obtain information about the electron–phonon relaxation processes for an electron concentration of ~10^{21} cm^{-3} in the conduction band. © *2004 MAIK “Nauka/Interperiodica”.*

PACS numbers: 71.35.Ee; 79.20.Ds

A considerable number of publications have been devoted to the experimental study of the dynamics of formation of electron–hole plasmas and ultrafast melting and ablation during the heating of a surface layer of silicon and gallium arsenide by femtosecond laser pulses (see, for example, [1–6] and the literature cited therein). In spite of the fact that a large body of experimental data has been accumulated to date, no adequate theoretical model has been constructed for these processes as yet. The corresponding investigations are quite important from both theoretical and applied points of view.

In this work, we consider the results of experiments on the dynamics of the formation of an electron–hole plasma and free-carrier recombination occurring in Si subjected to femtosecond laser pulses with a fluence lower than the melting and ablation threshold of the surface layer. Such measurements have become possible after the development of a high-power femtosecond laser on the basis of a chromium–forsterite element at the Institute of High Energy Densities, Associated Institute for High Temperatures, Russian Academy of Sciences [7, 8]. This laser operates in the IR spectral range and has no analogues in Russia or abroad.

It was demonstrated earlier [1–5] that a femtosecond laser pulse produces in the surface layer an electron–hole plasma with an electron concentration up to 10^{22} cm^{-3} in the conduction band. In this case, thermal and nonthermal melting of the surface layer takes place, resulting in a transition to the metal state. Pump–probe measurements of the time dependences of optical reflection were made for various pump-pulse fluences. The relation between the reflectivity and the electron concentration in the conduction band determined with the help of the Drude model was found to be in good agreement with the experimental data. However, the use of a probe pulse in the visible spectral range in previous experiments provided information about the dynamics of these processes only for pump-pulse fluences near the melting threshold, which corresponded to an electron concentration in the conduction band of $\approx 5 \times 10^{21}$–10^{22} cm^{-3}. In this study, the area under investigation was probed at a wavelength of 1240 nm, which made it possible to study the carrier recombination in the semiconductor prior to the melting of surface layer for the carrier concentrations starting with ~10^{21} cm^{-3}.

Figure 1 shows the reflectivity at a wavelength of 1240 nm as a function of the concentration of conduction electrons, calculated with the help of the Drude model analogously to [4] and corresponding to a damping time of $\tau_p \approx 1$ fs and electron effective mass $m_{\text{eff}} = 0.2m_e$ (m_e is the free-electron mass) [4], in good agreement with the obtained results.

It can be seen from the figure that the reflectivity at a wavelength of 1240 nm differs from the initial value

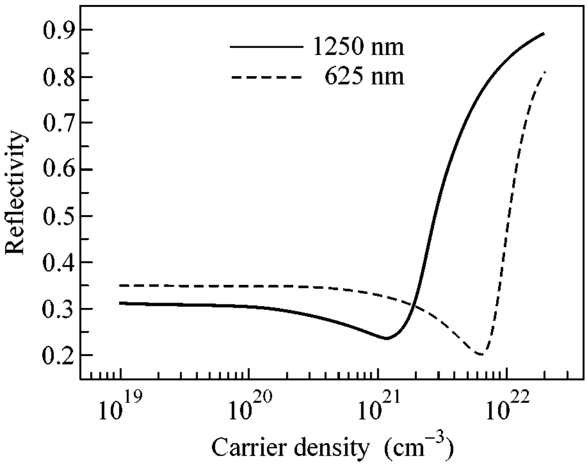


Fig. 1. Reflectivity at a wavelength of 1250 nm (solid curve) and 625 nm (dashed curve) as a function of the concentration of conduction electrons.

0021-3640/04/7911-0529$26.00 © 2004 MAIK “Nauka/Interperiodica”

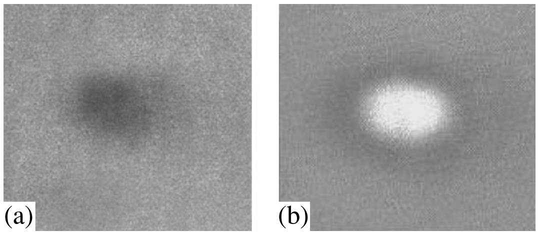


Fig. 2. Typical digital images of the silicon surface under the action of a pump pulse for fluences $F = 0.35F_a$ (a) and $F = 0.55F_a$ (b).

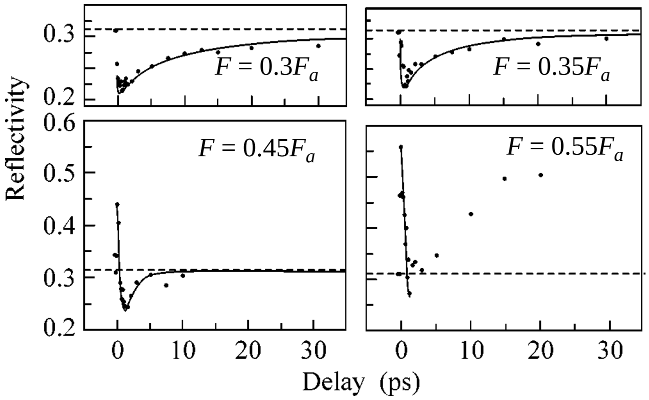


Fig. 3. Time dependences of the probe-pulse reflectivity measured for Si for various pump-pulse fluences (dots correspond to experimental values, solid curves are calculated using the Drude model, and dashed lines give the initial level of reflectivity).

for substantially lower (approximately by a factor of four) carrier densities, as compared to a probe-pulse wavelength of 620 nm [2, 4].

Measurements were made using the pump–probe scheme described in [6] for a probe-pulse duration of ≈80 fs, a pump-pulse duration of ≈120 fs, and a pump-pulse energy in the range $(0.25–0.75)F_a$, where F_a is the ablation threshold. The pump pulse was the second harmonic radiation ($\lambda = 620$ nm, $h\nu \approx 1.8$ eV), while the probe pulse was the fundamental radiation ($\lambda = 1240$ nm, $h\nu \approx 0.9$ eV) emitted by a chromium–forsterite femtosecond laser. Thus, we used a probe radiation with an energy of the quantum smaller than the semiconductor band gap E_g ($E_g = 1.1$ eV for Si).

Table

F/F_a	0.3	0.35	0.45	0.55
$N_{e\,\max}$ (10^{21} cm^{-3})	1.5	1.8	2.5	3.6
τ_e (10^{-12} s)	6	2.5	1.5	1.3
γ (10^{-31} cm^6/s)	0.8	1.2	1.0	0.8

Experiments were performed with a (100) Si single crystal. The pump photon penetration depth ($\lambda = 620$ nm) was ≈1 μm [9]. It should be noted that silicon is "transparent" to the probe pulse ($\lambda = 1240$ nm); consequently, the probe depth during pumping is determined by the absorption length of the pump pulse and the diffusion of hot electrons deep in the sample.

The pump and probe pulses were focused into spots of diameters 50 and 500 μm, respectively. The reflectivity was measured at the center of the focusing spot. Typical images of the pump area are shown in Fig. 2. The ablation threshold was determined from the appearance of a crater at the center of the focusing spot using the technique described in [5] and comprised $F_a = 0.43$ J/cm^2.

The results of measuring the time dependences of the probe-pulse reflectivity for various pump-pulse fluences (Fig. 3) can be summarized as follows.

The reflectivity as a function of time passes through several stages. At the first stage, coinciding in time with the pump pulse, the reflectivity abruptly changes due to the formation of a dense (with a carrier concentration $>10^{21}$ cm^{-3}) electron–hole plasma. At the second stage, as a result of recombination of electron–hole pairs, the reflectivity tends to its initial value. If the carrier concentration in this case exceeds a certain critical value [2], the reflectivity may considerably exceed the initial value (for $F = 0.45F_a$ and $F = 0.55F_a$). Otherwise, the reflectivity attains values smaller than the initial one due to the plasma compensation of the real part of the semiconductor dielectric function. The third stage is determined by the pump-pulse energy. For relatively low energy levels, the reflectivity returns to its initial value owing to a decrease in the carrier density. By contrast, at a higher level of the pump-pulse energy ($F = 0.55F_a$), the reflectivity begins to increase again to a level higher than the initial value after several picoseconds as the result of the phase transition to the liquid metal state [1–5].

As was noted in previous publications [2–4], such a process in silicon is due to the electron and hole recombination as a result of the Auger effect (Auger recombination). The table shows the maximal values of concentration $N_{e\,\max}$ of conduction electrons and characteristic time τ_e during which the concentration $N_{e\,\max}$ decreases by a factor of e^{-1} for various pump-pulse fluences as well as the Auger coefficient γ determined from the experimental data processed with the help of the Drude model for $\tau_p \approx 1$ fs and $m_{\mathrm{eff}} = 0.2m_e$.

It was pointed out in [1–4] that the error in the values of the concentration of conduction electrons and the Auger coefficient calculated on the basis of the Drude model is determined by the electron effective mass. If, however, we assume that the cause for the nonthermal melting observed for the pump-pulse fluences $F = 0.6–0.8F_a$ (both in our experiments and in [1-4]) is the formation of an electron–hole plasma with a hole concentration exceeding 5×10^{21} cm^{-3} (10% of the number of atoms [3]), then the above calculations will be in good agreement with this assumption.

It should be emphasized that, in the presence of Auger recombination, the number of free carriers rapidly decreases, but the energy of each carrier increases. The total energy of the ensemble of free carriers remains unchanged over a certain time interval [3]. Thus, the energy-transfer time from electrons to the lattice is at least longer than the carrier recombination time observed in our experiments.

REFERENCES

1. C. V. Shank, R. Yen, and C. Hirlimann, Phys. Rev. Lett. **51**, 900 (1983).
2. K. Sokolowski-Tinten, J. Bialkowski, and D. von der Linde, Phys. Rev. B **51**, 14186 (1995).
3. J. P. Callan, A. M.-T. Kim, L. Huang, and E. Mazur, J. Chem. Phys. **251**, 167 (2000).
4. K. Sokolowski-Tinten and D. von der Linde, Phys. Rev. B **61**, 2643 (2000).
5. D. von der Linde and H. Shuler, J. Opt. Soc. Am. B **13**, 216 (1996).
6. S. Ashitkov, M. Agranat, P. Kondratenko, *et al.*, JETP Lett. **76**, 461 (2002).
7. M. B. Agranat, S. I. Ashitkov, V. E. Fortov, *et al.*, in *Proceedings of XI Conference on Laser Optics (LO'2003)* (St. Petersburg, Russia, 2003).
8. M. B. Agranat, S. I. Ashitkov, A. A. Ivanov, *et al.*, Kvantovaya Élektron. (Moscow) **34** (6) (2004) (in press).
9. E. D. Palik, *Handbook of Optical Constants of Solids II* (Academic, San Diego, 1991).

Translated by N. Wadhwa

JETP Letters, Vol. 79, No. 11, 2004, pp. 532–536. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 660–664.

Electronic Structure of Carbon Nanohorns near the Fermi Level¶

D. V. Kolesnikov and V. A. Osipov

Joint Institute for Nuclear Research, Bogolyubov Laboratory of Theoretical Physics,
Dubna, Moscow region, 141980 Russia

e-mail: kolesnik@thsun1.jinr.ru; osipov@thsun1.jinr.ru

Received April 12, 2004

The effect of pentagonal defects on the electronic structure at the tip of carbon nanohorns is investigated within the continuum gauge field-theory model. It is found that the existence of a localized electron state at the Fermi level (a true zero-mode state) results in enhanced charge density near the tip. Using a self-consistent perturbation scheme, the eigenfunctions and the local density of states near the pentagonal defects are numerically calculated. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 73.22.-f

Nanohorns are novel synthesized members of a family of nanostructured carbon materials [1]. Experimentally [2], horn-shaped graphene sheets assemble to form dahlia flowerlike structures at the nanoscale. Some physical characteristics of nanohorns are unique, first of all, the prominent gas-adsorption phenomenon which makes it possible to use carbon nanohorns in fuel cells. It was also observed [3] that carbon nanohorn films have good field emission characteristics, making them promising candidates for field emission applications. It is interesting to note that the field amplification factor is found to depend on the geometrical shape of the emitter.

Geometrically, nanohorns are cone-shaped structures with a cone opening angle of about 19°, diameter of 2–4 nm, and length of about 50 nm [2]. Due to cone geometry, this opening angle corresponds to a 300° sector removed from the graphene sheet, which is equivalent to a creation of a $5\pi/3$ disclination. In turn, this means that nanohorns contain exactly five pentagons near the tip. Note, however, that the presence of the terminating cap at the tip of nanohorns does not allow the use of cone geometry for their description. In our opinion, a more appropriate geometry is the hyperboloid or, more precisely, the upper half of a two-sheet hyperboloid. Indeed, the hyperboloid has a cone asymptotic at large distances and a smoothing at the tip.

Both geometry and topological defects have a crucial influence on the electronic properties of carbon nanostructures. For instance, there is an important difference between electronic states at the apex of a cone and of a hyperboloid for the same number of pentagons. As was shown in [4], in the latter case, there is a possibility for the true zero-mode state. Namely, while for a cone normalized electron states at the Fermi level exist only for a finite system size and disappear in the infinite-size limit, they can exist for an infinite (full locus) hyperboloid. Likewise, a remarkable difference between electronic states near the tip of the pencil cap and bowl cap (containing six pentagons) was established [5] in metallic armchair nanotubes. At the same time, the different number of pentagons at the apex and even their relative location can lead to completely different behavior of electronic states [6–8].

The electronic properties of carbon nanohorns were theoretically studied in [6] using the parametrized linear combination of atomic orbitals (LCAO) technique. The optimum geometry and the electronic densities of states were determined for different nanohorn morphologies. This study shows that pentagonal sites determine all essential features of the electronic structure near the Fermi level at the tip. In particular, there appears negative excess charge near pentagons. It is interesting to study this problem within the continuum model. Despite the fact that the lattice disappears in the continuum description and, for example, nanohorns morphologies with different relative location of the five terminating pentagons cannot be taken into considerations, the continuum models were very effective in a study of electron states near the Fermi energy in fullerenes [9, 10], nanotubes [11], and nanocones [12, 13]. Moreover, the problem of specific electronic states at the Fermi level due to disclinations is similar to that of the fermion zero modes for planar systems in a magnetic field [14]. Generally, zero modes for fermions in topologically nontrivial background have been of current interest in both the field theory and condensed matter physics (see, e.g., review [15]). In addition, the continuum

---

¶ This article was submitted by the authors in English.

models allow us to study the electron states at large distances from the defects.

The first attempts to analyze the electronic states of carbon structures with a hyperboloid geometry within the continuum model were made in [4, 16]. The main finding was the existence of the normalized electron state at the Fermi level. However, any information about electronic states near the Fermi energy was lacking. In this paper, we adapt the continuum model to description of electronic states in nanohorns by considering the case of five pentagons at the tip (300° disclination) and taking into account a certain relation between the disclination power and the opening angle of the cone. Moreover, we develop a self-consistent perturbation scheme which allows us to calculate electronic eigenfunctions near the Fermi energy when the zero-mode solution is known. We also discuss asymptotic (long-range) solutions. Finally, we calculate the local density of electronic states (LDoS) near the Fermi energy.

Within the continuum approximation, we consider electrons on curved surfaces with disclinations taken into account. It was shown [17] that, for a flat graphene sheet, the self-consistent effective-mass theory can be reduced to the Dirac equation in (2 + 1) dimensions. This allows us to extend the consideration by formulating the Dirac equation on arbitrary curved surfaces incorporating topological defects by introducing $SO(2)$ gauge fields. In this context, disclination fields are accounted for by using the covariant derivative. As a result, the Dirac equation on an arbitrary surface in the presence of the $U(1)$ external gauge field W_μ is written as [10]

$$i\gamma^\alpha e^\mu_\alpha(\nabla_\mu - iW_\mu)\psi = E\psi, \tag{1}$$

where ψ is a two-component spinor and γ_μ are the Dirac matrices which can be chosen to be the Pauli matrices: $\gamma^1 = -\sigma^2$, $\gamma^2 = \sigma^1$; $\nabla_\mu = \partial_\mu + \Omega_\mu$, with Ω_μ being the spin connection term and e^μ_α being the zweibein. The Fermi energy in (1) is chosen to be zero, so that the states at $E = 0$ are the well-known zero-mode states in the field-theory context. Notice that for massless fermions σ^3 serves as a conjugation matrix, and the energy eigenmodes in (1) are symmetric around $E = 0$ ($\sigma^3\psi_E = \psi_{-E}$). The gauge field W_μ determines an elastic flow through a surface which is given by a circular integral

$$\frac{1}{2\pi}\oint \mathbf{W}d\mathbf{r} = \nu,$$

where ν describes the topological characteristics of the defect (the Frank index). Due to the symmetry group of the hexagonal lattice, the possible values of the Frank index are multiplies of 1/6.

The eigenfunctions in (1) are classified with respect to the eigenvalues of $J_z = j + 1/2$, $j = 0, \pm1, \pm2, \ldots$, and are to be taken in the form

$$\psi = \begin{pmatrix} u(r)e^{i\varphi j} \\ v(r)e^{i\varphi(j+1)} \end{pmatrix}. \tag{2}$$

Geometrically, the upper half of a hyperboloid can be regarded as the embedding

$$(\chi, \varphi) \longrightarrow (a\sinh\chi\cos\varphi, a\sinh\chi\sin\varphi, c\cosh\chi),$$
$$0 \le \chi < \infty, \quad 0 \le \varphi < 2\pi,$$

with a and c being the hyperboloid parameters. From this, the components of the induced metric can be obtained as (the details can be found in [4])

$$\begin{aligned} g_{\chi\chi} &= \alpha^2\cosh^2\chi + c^2\sinh^2\chi, \\ g_{\varphi\varphi} &= a^2\sinh^2\chi, \\ g_{\varphi\chi} &= g_{\chi\varphi} = 0, \end{aligned} \tag{3}$$

which yields

$$\begin{aligned} \Gamma^\chi_{\chi\chi} &= \frac{(a^2 + c^2)\sinh 2\chi}{2g_{\chi\chi}}, \\ \Gamma^\chi_{\varphi\varphi} &= -\frac{a^2\sinh 2\chi}{2g_{\chi\chi}}, \\ \Gamma^\varphi_{\varphi\chi} &= \Gamma^\varphi_{\chi\varphi} = \coth\chi \end{aligned} \tag{4}$$

for the nonvanishing coefficients of the connection. In a rotating $SO(2)$ frame, the zweibeins take the form

$$\begin{aligned} e^1_\chi &= \sqrt{g_{\chi\chi}}\cos\varphi, \quad e^2_\chi = \sqrt{g_{\chi\chi}}\sin\varphi, \\ e^1_\varphi &= -a\sinh\chi\sin\varphi, \quad e^2_\varphi = a\sinh\chi\cos\varphi, \end{aligned} \tag{5}$$

which gives

$$\omega^{12}_\chi = \omega^{21}_\chi = 0, \; \omega^{12}_\varphi = -\omega^{21}_\varphi = \left[1 - \frac{a\cosh\chi}{\sqrt{g_{\chi\chi}}}\right] = 2\omega \tag{6}$$

for the spin connection coefficients, so that

$$\Omega_\varphi = i\omega\sigma^3. \tag{7}$$

The external gauge potential was found [4] to be $W_\chi = 0$, $W_\varphi = \nu$. In this case, the substitution

$$\tilde{\psi} = \psi\sqrt{\sinh\chi}$$

reduces eigenvalue problem (1) to the system of equations

$$\begin{aligned} \partial_\chi\tilde{u} &= \sqrt{\coth^2\chi + b^2}\tilde{j}\tilde{u} = \tilde{E}\tilde{v}, \\ -\partial_\chi\tilde{v} - \sqrt{\coth^2\chi + b^2}\tilde{j}\tilde{v} &= \tilde{E}\tilde{u}, \end{aligned} \tag{8}$$

where $\tilde{E} = \sqrt{g_{\chi\chi}}E$, $b = c/a$, and $\tilde{j} = j - \nu + 1/2$.

Let us consider the zero-energy modes by setting $E = 0$ in (8). The general solution reads

$$\begin{aligned}\tilde{u}_0(\chi) &= A\left[(k\cosh\chi + \Delta)^{2k}\frac{\Delta - \cosh\chi}{\Delta + \cosh\chi}\right]^{\tilde{j}/2},\\ \tilde{v}_0(\chi) &= A\left[(k\cosh\chi + \Delta)^{2k}\frac{\Delta - \cosh\chi}{\Delta + \cosh\chi}\right]^{\tilde{j}/2},\end{aligned} \tag{9}$$

where $k = \sqrt{1 + b^2}$, $\Delta = \Delta(\chi) = \sqrt{1 + k^2\sinh^2\chi}$, and A is a normalization factor.

In the general case, for an unbounded hyperboloid (full locus), the normalization condition gives the following restrictions: $-1/2 < \tilde{j} < -1/2k$ for $u_0(\chi)$ and $1/2k < \tilde{j} < 1/2$ for $v_0(\chi)$. Thus, either $u_0(\chi)$ or $v_0(\chi)$ becomes normalizable on a hyperboloid of infinite volume except for the case $\nu = 1/2$. For $(c/a) \longrightarrow 0$, a normalizable solution does not exist. In fact, under this condition the hyperboloid is changing over to a plane. Consequently, our results are in accordance with the planar case.

Due to the cone asymptotic, it is quite reasonable to take into account the correspondence between the parameter k and the Frank index ν. Namely, like for a cone, one can specify $k = 1/(1 - \nu)$ (see, for example, [13]). In this case, only the $\tilde{u}_0$ mode becomes normalized and only for $j = 0$ and $3/4 < \nu < 1$. Therefore, a zero-mode solution exists for $\nu = 5/6$, that is, for the typical nanohorn configuration. This finding has an important physical consequence. Recall that there are two kinds of sublattice points in a unit cell of the graphene lattice due to degeneracy [17]. Hence, the presence of exactly one zero-mode state implies the enhancement of the electron densities only on one sort of points, while for two zero modes, both lattice points would have excess densities, thus allowing a metallization. Physically, this means that, for example, the field emission properties in the first case are less efficient. This agrees well with the experimental results in [3], where field emission properties of carbon nanohorn films were observed and compared with the best nanotube emitters.

The existence of zero-mode solutions allows us to formulate a self-consistent approximation procedure to obtain the eigenfunctions near the Fermi level. Namely, the substitution

$$\tilde{u}(\chi) = \tilde{u}_0(\chi)\mathcal{U}(\chi), \quad \tilde{v}(\chi) = \tilde{v}_0(\chi)\mathcal{V}(\chi)$$

reduces (8) to

$$\partial_\chi\mathcal{U} = \xi\Delta\mathcal{V}\frac{\tilde{v}_0}{\tilde{u}_0}, \quad \partial_\chi\mathcal{V} = -\xi\Delta\mathcal{U}\frac{\tilde{u}_0}{\tilde{v}_0}, \tag{10}$$

where $\xi = Ea$ or, in usual units, $\xi = Ea/\hbar v_F$ is a dimensionless parameter. For small ξ, $\mathcal{U}$ and $\mathcal{V}$ can be found by using the iteration procedure in the form

$$\mathcal{U} = \mathcal{U}^{(0)} + \xi\mathcal{U}^{(1)} + \ldots, \ \mathcal{V} = \mathcal{V}^{(0)} + \xi\mathcal{V}^{(1)} + \ldots, \tag{11}$$

where $\mathcal{U}^{(0)}$ and $\mathcal{V}^{(0)}$ are constants, and

$$\mathcal{U}^{(1)}(\chi) = \mathcal{U}^{(1)}(0) + \mathcal{V}^{(0)}\int_0^\chi\Delta(\eta)\frac{\tilde{v}_0(\eta)}{\tilde{u}_0(\eta)}d\eta,$$

$$\mathcal{V}^{(1)}(\chi) = \mathcal{V}^{(1)}(0) - \mathcal{U}^{(0)}\int_0^\chi\Delta(\eta)\frac{\tilde{u}_0(\eta)}{\tilde{v}_0(\eta)}d\eta.$$

Notice also that this approximation is valid for sufficiently small χ (when $\xi\Delta \ll 1$) because $\Delta(\chi)$ is a rapidly increasing function in (10).

Before we proceed, let us discuss asymptotic solutions. At large distances, the hyperboloid geometry becomes nearly the same as the geometry of a cone. For large χ (when $\sqrt{g_{\chi\chi}} \gg a$), Eqs. (8) are written as

$$\begin{aligned}\partial_r\tilde{u} - (\tilde{j}k/r)\tilde{u} &= Ek\tilde{v},\\ -\partial_r\tilde{v} - (\tilde{j}k/r)\tilde{v} &= Ek\tilde{u},\end{aligned} \tag{12}$$

where $r = ae^\chi/2$, which is exactly the polar coordinate r introduced for the cone geometry in [12]. The exact solution to (12) reads

$$\begin{aligned}\tilde{u}_\infty &= \sqrt{\frac{2r}{a}}\left(C_1J_{\left|\tilde{j}k - \frac{1}{2}\right|}(Ekr) + C_2J_{-\left|\tilde{j}k - \frac{1}{2}\right|}(Ekr)\right),\\ \tilde{v}_\infty &= \sqrt{\frac{2r}{a}}\left(C_1J_{\left|\tilde{j}k + \frac{1}{2}\right|}(Ekr) + C_2J_{-\left|\tilde{j}k + \frac{1}{2}\right|}(Ekr)\right),\end{aligned} \tag{13}$$

where $J_l(x)$ is the Bessel function and C_1 and C_2 are the constants. Naturally, when $r \longrightarrow \infty$ (13) can be approximated by

$$\tilde{u}_\infty \approx C\cos(Ekr + \varphi_0), \quad \tilde{v}_\infty \approx C\sin(Ekr + \varphi_0). \tag{14}$$

As follows from (14), $\tilde{u}_\infty^2 + \tilde{v}_\infty^2 \approx C^2 = \text{const}$, that is, the LDoS tends to a constant like for the cone (and like for a plane without disclinations). Notice that both Eqs. (12) and solutions (13) are also similar (but not identical) to those for a cone (cf. [4, 12]). Therefore, we may conclude that the influence of the disclination field on the density of states has a local character in the present geometry.

Since only one component in (9) becomes normalizable, one can set $\mathcal{V}^{(0)} = 0$ in (11). Then, in the leading approximation, one obtains

$$\begin{aligned}\tilde{u}(\chi) &= \mathcal{U}^{(0)}\tilde{u}_0(\chi),\\ \tilde{v}(\chi) &= -\xi\mathcal{U}^{(0)}\tilde{v}_0(\chi)\int_0^\chi\Delta(\eta)\frac{\tilde{u}_0(\eta)}{\tilde{v}_0(\eta)}d\eta.\end{aligned} \tag{15}$$

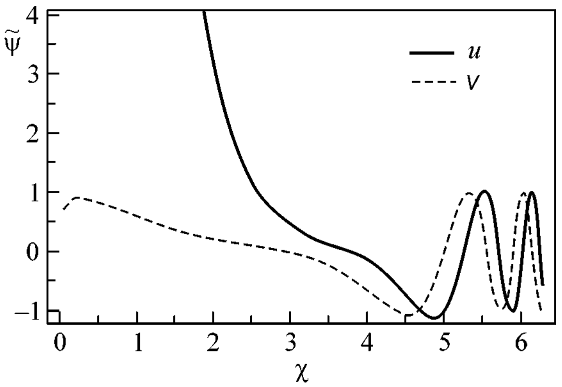


Fig. 1. The solution to (8) with the set of parameters $a = 1$, $j = 0$, $E = 0.01$, and $\nu = 5/6$.

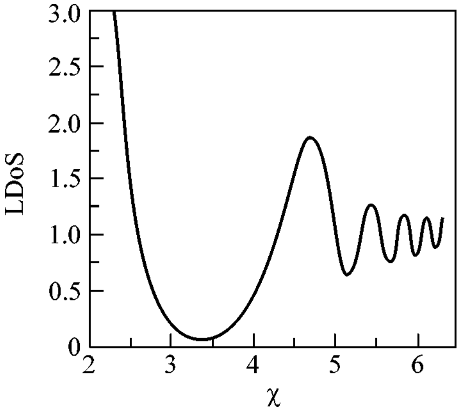


Fig. 2. Local density of states (per area) as a function of χ. The parameter set is the same as in Fig. 1.

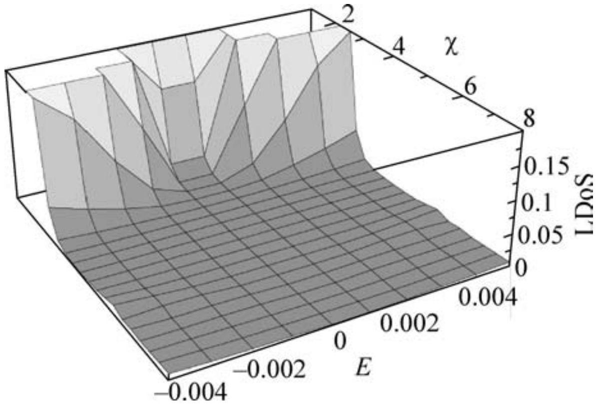


Fig. 3. Local density of states (per area) in arbitrary units as a function of χ and *E*. The parameter set (except *E*) is the same as in Fig. 1.

This allows us to integrate (8) numerically by using (15) as initial conditions. The results are shown in Figs. 1–3. As is seen from Fig. 1, the wave functions have an oscillatory behavior at large distances, in agreement with (14). With ξ decreasing, $\tilde{u}(\chi)$ rapidly increases while $\tilde{v}(\chi)$ decreases. Figure 2 shows the calculated LDoS (per area) as a function of χ at fixed energy. One can see a remarkable increase of the LDoS near the defect. The more detailed structure of the LDoS can be found in Fig. 3, where a 3D plot is presented. There is a finite density of electronic states at the Fermi level slowly growing with *E*.

Let us note once again that this behavior follows directly from the existence of the zero-mode state. For any other permissible values of the Frank indices (for a number of pentagons less than five), the zero mode does not exist and, correspondingly, no finite LDoS at the Fermi level is expected. It is interesting to compare our results with those obtained by the LCAO method [6], where only 25 atoms contained in the five terminating pentagons were considered for various tip morphologies. It is quite reasonable that the best agreement takes place for the most symmetrical arrangements of the pentagonal sites at the tip in [6]. Let us note also the good agreement of our results with [7] (at small *E*), where LDoS for five pentagons at the apex of a cone were investigated by using tight-binding calculations. In addition to [6, 7], the continuum description allows us to control the behavior of the LDoS at large distances from a tip which is linear in *E*, as for the flat graphene sheet.

To conclude, we have studied the electronic structure of carbon nanohorns within the continuum field-theory gauge model. The hyperboloid geometry is considered as the most appropriate for description of nanohorns. We have taken into account the natural correspondence between the parameters of the hyperboloid and the disclination power (the parameter *k* and the Frank index ν) and found that only for five pentagons at the tip does a normalized electron state appear at the Fermi level for an unbounded hyperboloid (the true zero-mode state). This finding allows us to formulate a self-consistent method for describing the electron states near the Fermi energy. The calculated LDoS rapidly grows in the tip region (near the disclination) and has a distinct nonzero minimum at the Fermi energy.

REFERENCES

1. S. Iijima, Physica B (Amsterdam) **323**, 1 (2002).
2. S. Iijima, M. Yudasaka, R. Yamada, *et al.*, Chem. Phys. Lett. **309**, 165 (1999).
3. J.-M. Bonard, R. Gaál, S. Garaj, *et al.*, J. Appl. Phys. **91**, 10107 (2002).
4. V. A. Osipov, E. A. Kochetov, and M. Pudlak, JETP **96**, 140 (2003).
5. T. Yaguchi and T. Ando, J. Phys. Soc. Jpn. **71**, 2224 (2002).

6. S. Berber, Y.-K. Kwon, and D. Tománek, Phys. Rev. B **62**, R2291 (2000).
7. J.-C. Charlier and G.-M. Rignanese, Phys. Rev. Lett. **86**, 5970 (2001).
8. P. E. Lammert and V. H. Crespi, Phys. Rev. B **69**, 035 406 (2004).
9. J. González, F. Guinea, and M. A. H. Vozmediano, Phys. Rev. Lett. **69**, 172 (1992); Nucl. Phys. B **406**, 771 (1993).
10. V. A. Osipov and E. A. Kochetov, JETP Lett. **72**, 199 (2000).
11. C. L. Kane and E. J. Mele, Phys. Rev. Lett. **78**, 1932 (1997).
12. P. E. Lammert and V. H. Crespi, Phys. Rev. Lett. **85**, 5190 (2000).
13. V. A. Osipov and E. A. Kochetov, JETP Lett. **73**, 562 (2001).
14. R. Jackiw, Phys. Rev. D **29**, 2375 (1984).
15. G. E. Volovik, Phys. Rep. **351**, 195 (2001).
16. R. Pincak and V. A. Osipov, Phys. Lett. A **314**, 315 (2003).
17. D. P. DiVincenzo and E. J. Mele, Phys. Rev. B **29**, 1685 (1984).

JETP Letters, Vol. 79, No. 11, 2004, pp. 537–541. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 665–669.
Original Russian Text Copyright © 2004 by Sin'ko, Smirnov.

Anomalous Behavior of α-Iron Zero-Temperature Isotherm in the Region of Negative Pressures

G. V. Sin'ko* and N. A. Smirnov

Russian Federal Nuclear Center, Zababakhin All-Russia Research Institute of Technical Physics,
Snezhinsk, Chelyabinsk region, 456770 Russia

**e-mail: g.v.sinko@vniitf.ru*

Received February 19, 2004; in final form, April 16, 2004

The detailed *ab initio* calculations of the electronic structure of α-iron crystals revealed the presence of a narrow interval of specific volumes in the $P(V, T = 0)$ isotherm where $dP/dV > 0$. According to the calculations, this anomaly occurs at pressures from –3 to –5 GPa and is due to the intensive energy-spectrum rearrangement accompanied by the almost simultaneous appearance of eight electronic topological transitions. *© 2004 MAIK “Nauka/Interperiodica”.*

PACS numbers: 71.20.Gj; 62.20.Dc; 64.30.+t; 75.50.Bb

Interest in the theoretical and experimental studies of the iron properties is not weakening because it is one of the most popular constructional materials and, in addition, is the material of Earth's core. In particular, of great importance is the reliable calculation of the cold curves and the pressure dependence of elastic constants of various iron crystal structures. Inasmuch as the accuracy of the calculated elastic constants can be affected by the presence of electronic topological transitions (ETTs) in the pressure range considered [1], we studied a change in the electronic spectrum of several iron crystal structures under pressure. The results of this study will be published in detail elsewhere. In this work, we draw attention only to one and, in our opinion, quite intriguing result that requires experimental verification. In the α-iron crystals possessing ferromagnetic properties and having the body-centered cubic structure (*bcc FM*), the regions of specific volumes with $dP/dV > 0$ were observed at relatively low negative pressures. Apart from the purely academic interest, this result, if confirmed experimentally, is also of practical importance, because it will shed new light on the mechanism of the formation of so-called smooth break-away upon the shock-wave interaction with the α-iron surface [2]. In the calculations of the crystal electronic structure, we used the FPLMTO method [7] that successfully recommended itself in our previous computations [3–6]. The exchange-correlation functional in the form used in [8] included gradient corrections [9]. All relativistic effects were taken into account, except for the spin–orbit interaction of valence electrons, which were described in the scalar-relativistic approximation. Special automated algorithm of choosing the internal parameters of method [7] and the refinement of its numerical realization provided a physical error of less than 0.1 mRy/atom in the specific-energy calculations, while the corresponding mathematical error was considerably lower. Electrons occupying the 3*s*, 3*p*, 3*d*, and 4*s* states in the iron atom were treated as valence electrons, and the half-core states were not set off. The basis set included the orbitals of the *s*, *p*, *d*, and *f* types. The grid for integrating over the **k** space using the linear tetrahedron method was constructed in the prism-shaped Brillouin zone by dividing each edge into 50 parts ($50 \times 50 \times 50$ grid). The variation of the number of grid points did not dramatically change our results.

The experimental values of density, bulk modulus, and pressure derivative of bulk modulus at normal conditions and the same values calculated by us and other

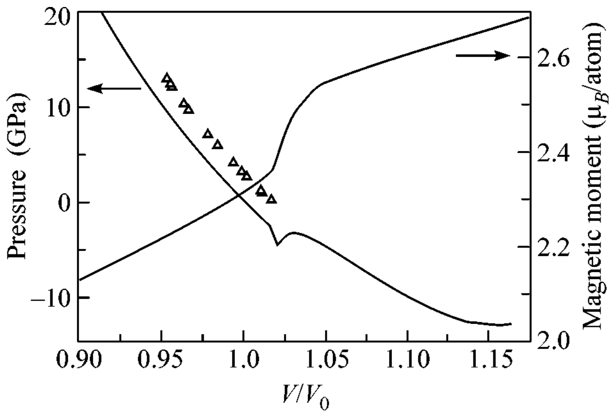


Fig. 1. Pressure and specific magnetic moment of α-iron as functions of volume fraction. Solid lines are for the calculation and triangles are for the experiment [10].

Calculated and experimental values of the specific volume V_0, bulk modulus B_0, and pressure derivative of bulk modulus

	V_0 (au)^3	B_0 (GPa)	B_0'	$\mu(\mu_B)$
This work	78.174	168.6	9.7	2.31
Calculation [11]	80.629	174.0	–	2.20
Calculation [12]	76.780	215.0	–	2.21
Calculation [13]	76.840	189.0	4.9	2.17
Experiment [14], T = 293 K	79.063	–	–	2.12
Experiment [15], T = 293 K	79.496	171.6	–	–
Experiment [16], T = 293 K	79.845	168.3	–	2.22
Experiment [17], T = 293 K	79.510	172.0	–	–
Experiment [18], T = 293 K	–	–	5.0	–
Experiment [19], T = 293 K	–	175.8	7.67	–
Extrapolation of the data [20] to T = 0 K	79.075	–	–	–

authors for the *bcc FM* structure at $T = 0$ and $P = 0$ are compiled in the table. The calculated volume dependences of pressure and specific magnetic moment are shown in Fig. 1. The results of the experiments on determining the iron compressibility under normal conditions are also shown in this figure. These data allow the conclusion to be drawn that the accuracy of our calculations is high enough.

In Fig. 1 and in what follows, V_0 denotes the value of specific volume calculated in this work for $T = 0$ and $P = 0$.

The calculated volume dependence of specific energy in the anomaly region is shown in Fig. 2. The volume dependences of the pressure and specific magnetic moment in the anomaly region are shown in Fig. 3 at a scale larger than in Fig. 1. The change in the density of states at the Fermi surface in the anomaly region is shown in Fig. 4. The complicated behavior of this curve is caused by the high ETT concentration in the anomalous region. Clearly, a 50 × 50 × 50 grid is insufficient for the accurate description of the instant of ETT. For this reason, the fine details of the curve in Fig. 4 depend on the number of grid points used for the integration over the **k** space, although the overall character of the curve is independent of the number of grid points.

According to the calculations, the *bcc FM* structure undergoes 12 ETTs in the interval $0.86 \le V/V_0 \le 1.23$. Eight of them are accompanied by a change of less than one percent in specific volume, indicating the strong energy-spectrum rearrangement and, as a result, the

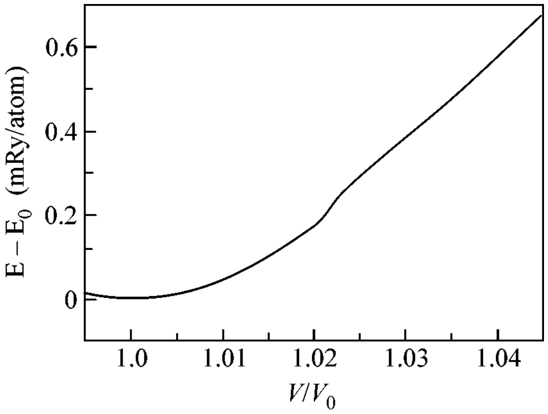


Fig. 2. Specific energy of α-iron vs. volume fraction in the anomalous region.

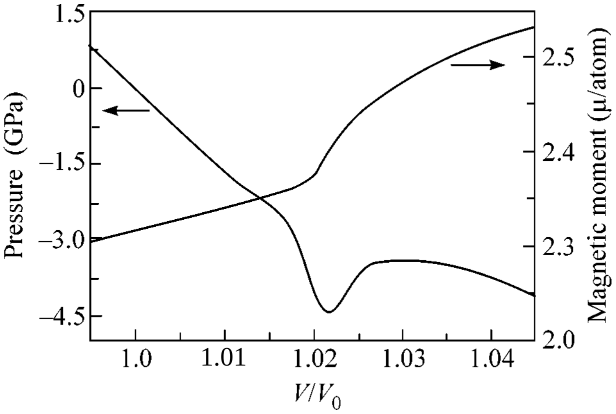


Fig. 3. Pressure and specific magnetic moment of α-iron vs. volume fraction in the anomalous region.

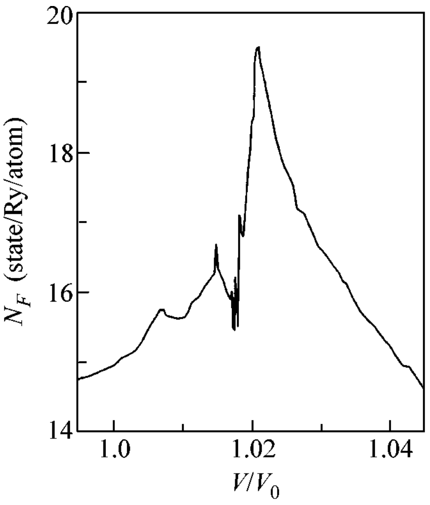


Fig. 4. Density of states at the Fermi surface of α-iron vs. volume fraction.

Fig. 5. Fermi-surface area corresponding to the 10^+ band of α-iron.

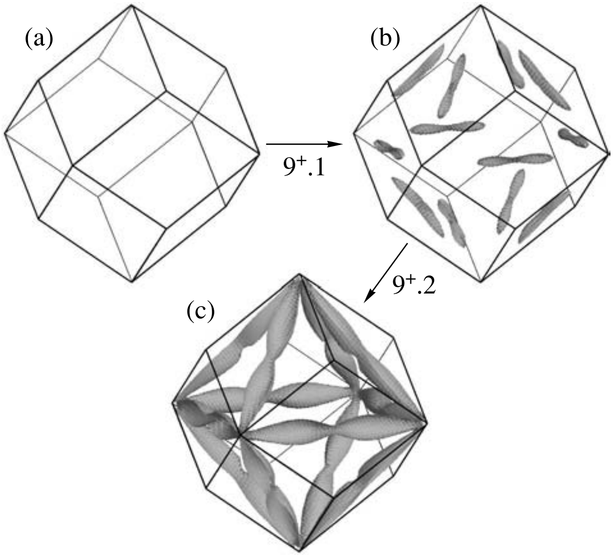


Fig. 6. Electronic topological transitions in the 9^+ band.

loss of crystal mechanical stability. The ratio between the numbers of electrons with the positive and negative spin projections changes drastically in the course of spectrum rearrangement, leading to an almost jumpwise change in the crystal magnetic moment, whose volume dependence is shown in Figs. 1 and 3.

In the interval of specific volumes considered, of the 16 valence electrons in the unit cell of *bcc FM*, electrons with the positive spin projection are distributed over ten bands. We will number them 1^+, …, 10^+ in the order of increasing energy. Electrons with the negative spin projection are distributed over eight bands. These bands will also be numbered 1^-, …, 8^- in the order of increasing energy. The 1^+–6^+ and 1^-–5^- bands are completely filled in the entire range of compressions considered, while the filling of other bands changes. The changes in the topology of the corresponding Fermi-surface regions are illustrated in Figs. 5–11.

The 10^+ band is partially filled over the entire range of compressions studied. The topology of the Fermi-surface area corresponding to the 10^+ band does not change, and its shape is shown in Fig. 5. In the 9+ band, which is completely filled at a sufficiently large negative pressure, holes appear with increasing pressure, and the corresponding region of the Fermi surface undergoes two ETTs (Fig. 6). With an increase in pressure, holes also appear in the 8^+ and 7^+ bands, which are completely filled at negative pressure, and the Fermi surfaces appear in each of these bands; the corresponding regions of these surfaces are shown in Figs. 7 and 8. In the 8^- band, which is empty at a sufficiently large negative pressure, electrons appear with pressure, and the Fermi-surface region corresponding to this band undergoes three ETTs (Fig. 9). Two ETTs occur in the 7^- band, which is partially filled over the entire range of compressions studied (Fig. 10). Three ETTs occur in the 6^- band, which is also partially filled at a sufficiently large negative pressure, and, as a result, this band becomes completely filled. The transformation of the

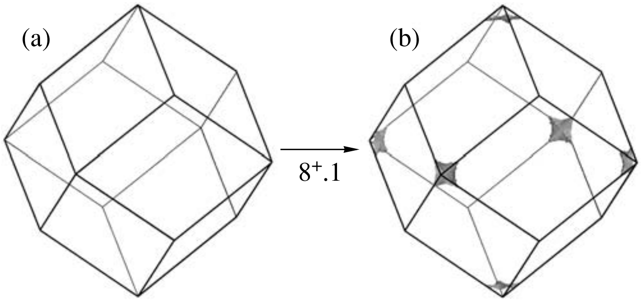


Fig. 7. Electronic topological transition in the 8^+ band.

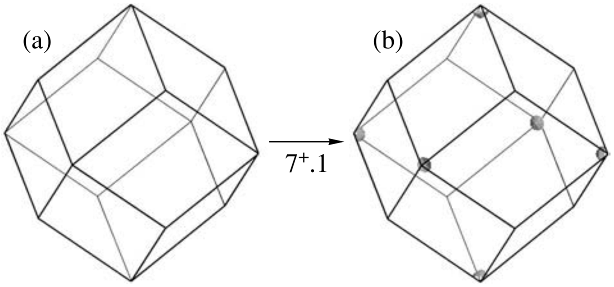


Fig. 8. Electronic topological transition in the 7^+ band.

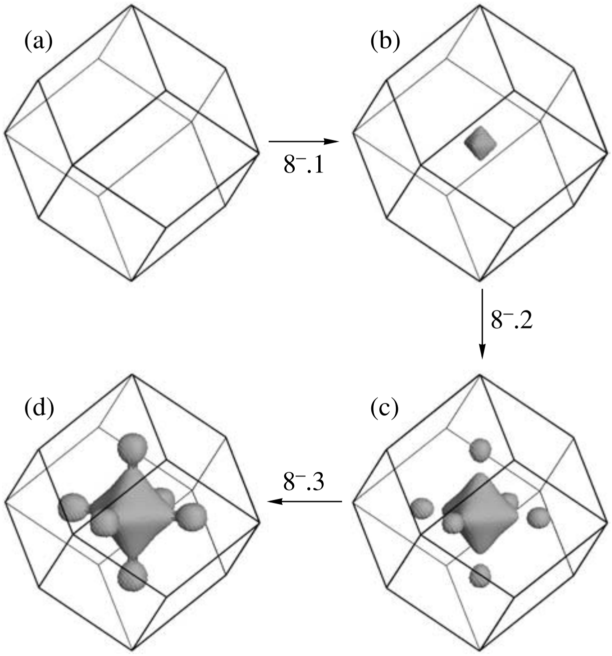


Fig. 9. Electronic topological transitions in the 8^- band.

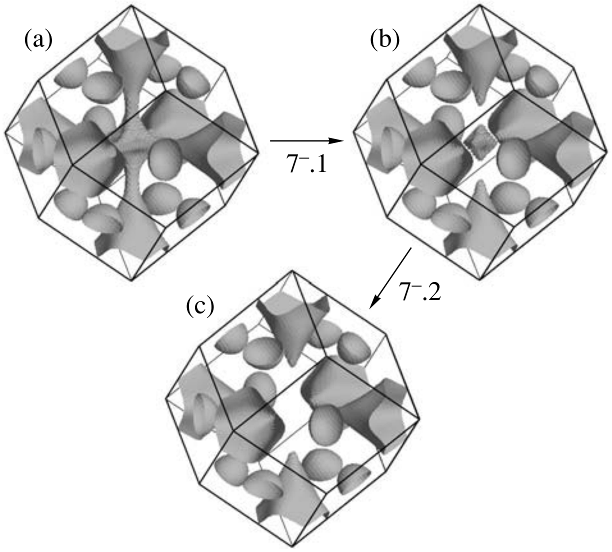


Fig. 10. Electronic topological transitions in the 7^- band.

corresponding Fermi-surface region can be seen in Fig. 11. For the ETTs shown in Figs. 5–11, the $9^+.1$ transition occurs at $V/V_0 \approx 1.057$, the $6^-.3$ transition occurs at $V/V_0 \approx 0.99$, the $8^-.2$ transition occurs at $V/V_0 \approx 0.96$, and the $8^-.3$ transition occurs at $V/V_0 \approx 0.88$. The eight remaining ETTs occur in the range $1.016 < V/V_0 < 1.023$ of volume fractions.

To our knowledge, such anomalies in the volume dependences of the crystal specific energy and pressure have been observed so far neither in the computational nor in the experimental works. For this reason, it is of great interest to study experimentally α-iron crystals at negative pressures and low temperatures, with the object of revealing the unusual behavior of their elastic properties. In this connection, a new method suggested in [21] for measuring the sound velocity and allowing it to be measured upon both crystal compression and tension can prove to be quite helpful. The experiments with iron whiskers can also be useful. The observation of an abrupt change in the magnetic moment of an iron sample upon the interaction of a shock wave with the surface can also serve as an indirect confirmation of the presence of anomaly in the elastic properties of α- iron.

This work was supported by the MNTTs (project no. 1181) and the Russian Foundation for Basic Research (project no. 04-02-17292).

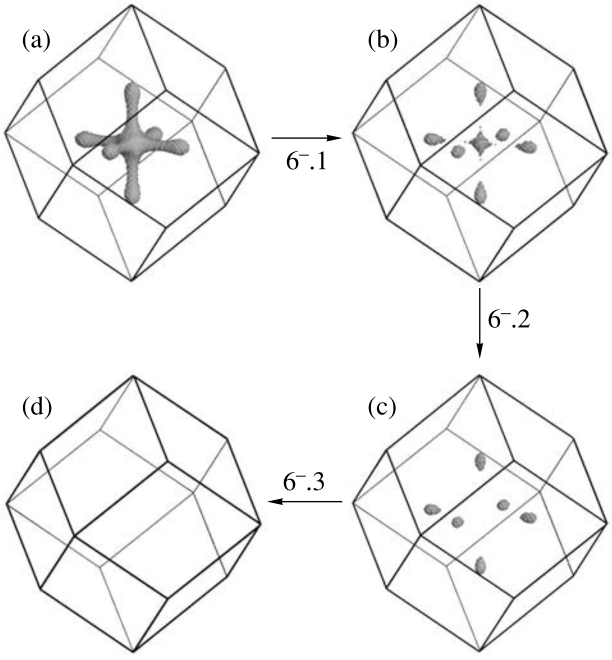


Fig. 11. Electronic topological transitions in the 6^- band.

REFERENCES

1. I. M. Lifshits, Zh. Éksp. Teor. Fiz. **38**, 1569 (1960) [Sov. Phys. JETP **11**, 1130 (1960)].
2. J. O. Erkman, J. Appl. Phys. **32**, 939 (1961).
3. M. I. Katsnelson, G. V. Sin'ko, N. A. Smirnov, *et al.*, Phys. Rev. B **61**, 14420 (2000).
4. G. V. Sin'ko and N. A. Smirnov, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 217 (2002) [JETP Lett. **75**, 184 (2002)].
5. G. V. Sin'ko and N. A. Smirnov, J. Phys.: Condens. Matter **14**, 6989 (2002).
6. G. V. Sin'ko and N. A. Smirnov, in *Physics of Extreme States of Substance—2003*, Ed. by V. E. Fortov *et al.* (IPKhF RAN, Chernogolovka, 2003).
7. S. Yu. Savrasov and D. Yu. Savrasov, Phys. Rev. B **46**, 12181 (1992).
8. J. F. Janak, V. L. Moruzzi, and A. R. Williams, Phys. Rev. B **12**, 1257 (1975).
9. J. P. Perdew, J. A. Chevary, S. H. Vosko, *et al.*, Phys. Rev. B **46**, 6671 (1992).
10. *Landolt–Börnstein. Structure Data of Elements and Intermetallic Phases* (Springer, Berlin, 1988), Vol. 14/III.
11. T. C. Leung, C. T. Chan, and B. N. Harmon, Phys. Rev. B **44**, 2923 (1991).
12. P. Dufek, P. Blaha, and K. Schwarz, Phys. Rev. B **50**, 7279 (1994).
13. L. Stixrude, R. E. Cohen, and D. J. Singh, Phys. Rev. B **50**, 6442 (1994).
14. W. B. Pearson, *Handbook of Lattice Spacings and Structures of Metals and Alloys* (Pergamon, New York, 1958).
15. K. A. Gschneidner, Jr., Solid State Phys. **16**, 275 (1964).
16. C. Kittel, *Introduction to Solid State Physics*, 5th ed. (Wiley, New York, 1976; Nauka, Moscow, 1978).
17. A. P. Jephcoat, H. K. Mao, and P. M. Bell, J. Geophys. Res. **91**, 4677 (1986).
18. M. W. Guinan and D. N. Beshers, J. Phys. Chem. Solids **29**, 541 (1968).
19. S. N. Vaidya and G. C. Kennedy, J. Phys. Chem. Solids **31**, 2329 (1970).
20. S. I. Novikova, *Thermal Expansion of Solids* (Nauka, Moscow, 1974) [in Russian].
21. G. S. Bezruchko, G. I. Kanel', and S. V. Razorenov, Teplofiz. Vys. Temp. **42** (2) (2004) [High Temp. **42**, 234 (2004)].

Translated by V. Sakun

JETP Letters, Vol. 79, No. 11, 2004, pp. 542–544. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 670–673.
Original Russian Text Copyright © 2004 by Markina, Klimov, Vasiliev, Freimut, Kordonis, Kreiner, Lorentz, Yamauchi, Ueda.

Sequence of Phase Transitions in a Quasi-One-Dimensional β-$Na_{0.33}V_2O_5$ Compound with Variable Valence

M. Markina[1, *], K. Klimov[1], A. N. Vasiliev[1], A. Freimut[2], K. Kordonis[2], M. Kreiner[2], T. Lorentz[2], T. Yamauchi[3], and Y. Ueda[3]
[1] *Faculty of Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
**e-mail: mmarkina@mail.ru*
[2] *2d Physical Institute, University of Köln, D-50937 Köln, Germany*
[3] *Institute for Solid State Physics, University of Tokyo, 277- 8581 Tokyo, Japan*

Received April 20, 2004

The thermal properties—specific heat, thermal conductivity, and thermal expansion coefficients—of a single crystal of quasi-one-dimensional variable-valence β-$Na_{0.33}V_2O_5$ compound were studied. With lowering temperature, it sequentially undergoes the structural ($T_S \sim 230$ K), charge ($T_C \sim 136$ K), and magnetic ($T_N \sim 22$ K) phase transitions. The structural transition at T_S, resulting in the ordering of the Na ions, and the charge ordering at T_C, resulting in the charge redistribution over the positions of V ions, are accompanied by the anomalies in the temperature dependences of all the studied properties. The magnetic ordering at T_N results in the appearance of the canted antiferromagnetic structure and manifests itself only in the anomaly in the temperature dependences of the thermal expansion coefficients. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 65.40.Ba, 65.40.De; 64.70.Kb; 75.30.Kz

Compounds relating to the family of vanadium bronzes of the β(β') type with general formula β(β')–$A_xV_2O_5$ (A = Na, Sr, Cu, etc.) are quasi-one-dimensional conductors [1]. The variable-valence sodium-based β-$Na_{0.33}V_2O_5$ compound belongs to this family. With lowering temperature, it undergoes the structural ($T_S \sim 230$ K), charge ($T_C \sim 136$ K), and magnetic ($T_N \sim 22$ K) phase transitions [2]. Under the hydrostatic pressure $P \sim 8$ GPa, the charge-ordered phase of β-$Na_{0.33}V_2O_5$ collapses and it undergoes the transition to the superconducting state at $T_{SC} = 8$ K [3]. The monoclinic crystal structure of this compound (Fig. 1) is typical of all vanadium bronzes in the β(β') phase and contains tunnels formed by the VO_6 and VO_5 complexes along the *b* axis. The Na ions are located inside the tunnels. At high temperatures, the vanadium ions in β-$Na_{0.33}V_2O_5$ occupy three crystallographically different positions: V1 in the octahedral oxygen environment forms zig-zag chains of edge-shared VO_6 octahedra along the *b* axis, V2 in the same environment forms double chains of angle-shared octahedra, and V3 forms zig-zag chains of edge-shared VO_5 pyramids [2]. The Na ions (*A*) also occupy two different crystallographic positions in the unit cell. With lowering temperature, the Na ions become structurally ordered at T_S and a superstructure of the $1 \times 2 \times 1$ type arises along the *b* axis.

β-$Na_{0.33}V_2O_5$ is a compound with variable valence, because its unit cell contains one V^{4+} ion and five V^{5+} ions. At $T > T_C$, all vanadium sites are equivalent and β-$Na_{0.33}V_2O_5$ is a conductor. With lowering temperature, this compound undergoes at T_C the metal–insulator transition of the charge-ordering type; nonmagnetic V^{+5} ions occupy the V3 positions, while the V1 and V2 positions can be occupied by both V^{5+} and V^{4+}. According to [4], the magnetically active V^{4+} ions are at the V1 positions. Optical studies of β-$Na_{0.33}V_2O_5$ showed that the charge redistribution in the lattice is accompanied by strong polaron effects [5].

Although the magnetic subsystem in $Na_{0.33}V_2O_5$ is highly diluted, the canted antiferromagnetic structure is formed in this material at T_N [6]. The very fact that the

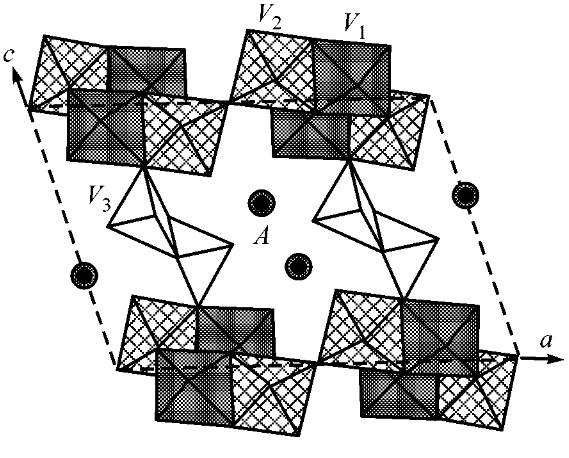


Fig. 1. Crystal structure of β-$Na_{0.33}V_2O_5$.

long-range magnetic order appears in a diluted structure indicates that, upon the charge ordering, the magnetically active V^{4+} ions are arranged in such a way as to provide for a rather strong exchange interaction between them.

To gain additional information on the kinetic and thermodynamic properties of this compound, the thermal capacity, thermal conductivity, resistivity, and thermal expansion coefficients of a β-$Na_{0.33}V_2O_5$ single crystal were measured in this work over a broad temperature range. Single crystals were grown from a melt using $NaVO_3$ as flux. The samples were parallelepipeds with characteristic sizes of 0.5 × 5 × 0.5 mm and with the largest size oriented along the *b* axis. The temperature dependences of the thermal capacity and resistivity were studied using a quasi-adiabatic microcalorimeter and an automated setup (Termis) for measuring resistivity. The thermal conductivity and the thermal expansion coefficients were studied by the methods of constant thermal flow and capacitive dilatometry, respectively.

The temperature dependence of the thermal capacity of β-$Na_{0.33}V_2O_5$ $C(T)$ in the range 5–300 K is shown in Fig. 2. A peak at the metal–insulator phase-transition temperature $T_C = 136$ K is clearly seen in this curve, and the slope of the $C(T)$ curve slightly changes after the temperature of structural phase transition at $T_S \sim 230$ K. The absence of anomaly in the thermal capacity at the magnetic ordering temperature $T_N \sim 22$ K can be explained in the following way. In the materials containing one-dimensional chains of exchange-coupled magnetic ions, the three-dimensional order in the whole magnetic subsystem appears due to the presence of a weak interchain interaction. This interaction and, hence, the ordering temperature are smaller than the intrachain interaction between the magnetic ions. Therefore, the magnetic entropy released at the three-dimensional ordering temperature is small and does not lead to the appearance of a noticeable anomaly in the thermal capacity. Moreover, the magnetic subsystem of β-$Na_{0.33}V_2O_5$ is highly diluted and its contribution to the thermal capacity is small on the background of the rapidly increasing phonon contribution.

The temperature dependence of the thermal capacity in the low-temperature range was approximated using two terms: $\beta T^3 + \gamma T^{3/2}$. The first term is responsible for the phonon contribution and the contribution from the magnetic subsystem of a three-dimensional antiferromagnet. The second term corresponds to the contribution from a three-dimensional ferromagnet, in accordance with the presence of a spontaneous magnetic moment below T_N. In this case, the estimate of the lower bound for the Debye temperature gives θ_D = 490 K. A satisfactory approximation can also be obtained with only the first term βT^3. In this case, the numerical value of the parameter β gives a value of 456 K for the lower bound of the Debye temperature.

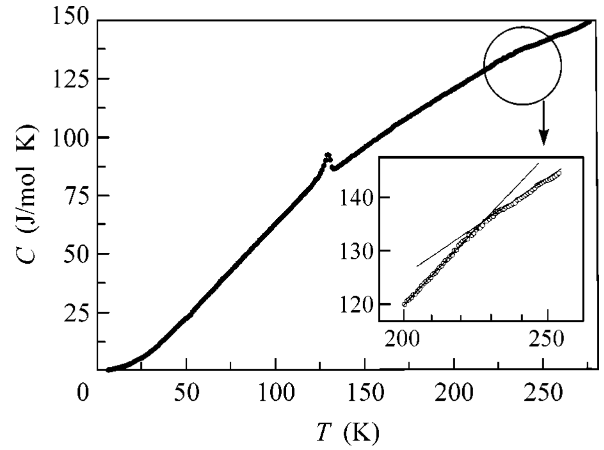


Fig. 2. Thermal capacity of β-$Na_{0.33}V_2O_5$.

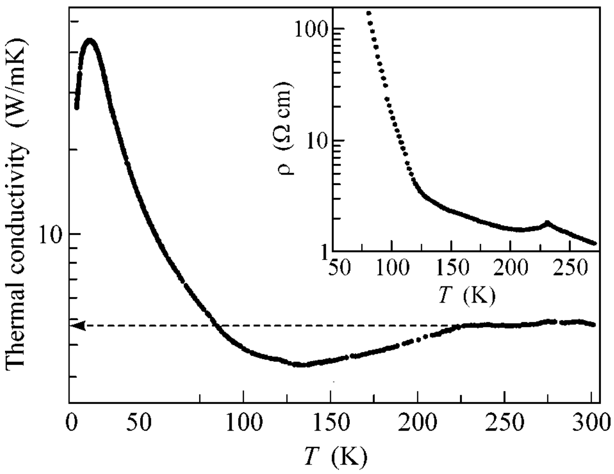


Fig. 3. Thermal conductivity of β-$Na_{0.33}V_2O_5$ along the *b* axis. The temperature dependence of the resistivity of β-$Na_{0.33}V_2O_5$ along the *b* axis is shown in the inset.

The temperature dependence of the thermal conductivity coefficient $\kappa(T)$ measured along the *b* axis of a β-$Na_{0.33}V_2O_5$ single crystal is presented in Fig. 3. With an increase in temperature, $\kappa(T)$ reaches its maximal value $\kappa_{max} \sim 45$ W/(m K) at $T \sim 13$ K. This maximum is characteristic of the phonon-assisted heat transfer and appears due to the increase in the number of phonons, on the one hand, and to a decrease in the phonon mean free path with lowering temperature, on the other. The metal–insulator transition at T_C is accompanied by a change in the sign of the $\kappa(T)$ slope in the interval T_C–T_S. The increase in the thermal conductivity of β-$Na_{0.33}V_2O_5$ at $T > T_C$ can be due to electron delocalization, i.e., to the electron redistribution over the vanadium sites. This process is likely terminated at the structural transition temperature T_S, above which the thermal conductivity is virtually independent of tem-

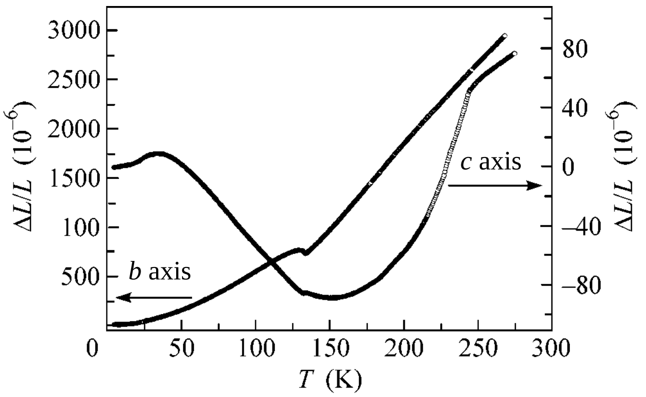


Fig. 4. The elongation per unit length along the *b* and *c* axes of β-$Na_{0.33}V_2O_5$.

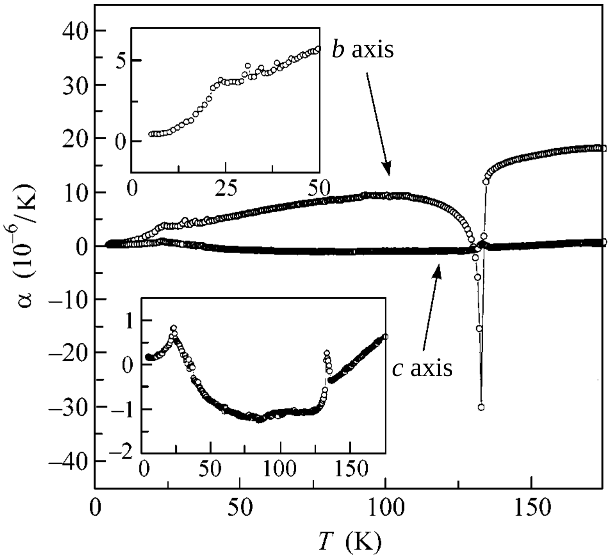


Fig. 5. Thermal expansion coefficients $\alpha_i(T)$ along the *b* and *c* axes of a β-$Na_{0.33}V_2O_5$ single crystal. The behavior of $\alpha_b(T)$ in the vicinity of T_N and $\alpha_c(T)$ over a broad temperature range are shown in the insets.

perature. The value $\kappa_0 \sim 5$ W/(m K) achieved by the thermal conductivity at $T > T_S$ corresponds, by the order of magnitude, to a "bad" metal.

The temperature dependence of the resistivity $\rho(T)$ of β-$Na_{0.33}V_2O_5$ measured along the *b* axis is shown in the inset in Fig. 3. The structural transition at T_S shows up in this curve as a break, and the metal–insulator transition at T_C is accompanied by a sharp increase in the resistivity.

The temperature dependences of the elongation per unit length along the crystallographic *b* and *c* axes of β-$Na_{0.33}V_2O_5$ are shown in Fig. 4. The thermal expansion in this crystal is highly anisotropic. With lowering temperature, the sample is strongly compressed along the *b* axis, while the size along the *c* axis slightly changes in the alternating manner. As the temperature lowers, the "invar" effect is observed along the *c* axis in the T_N–T_C interval: the sample length increases with cooling. This is likely due to the charge redistribution over the VI positions in the structure of β-$Na_{0.33}V_2O_5$. The thermal expansion coefficients α_i of the β-$Na_{0.33}V_2O_5$ single crystal are also highly anisotropic (Fig. 5). At $T < T_N$, they strongly change because of the spontaneous crystal striction in the magnetically ordered state.

The charge ordering at T_C is also accompanied by the anomalies in α_i, although they differ in magnitude and sign. A sharp anomaly is seen along the *b* axis, indicating that the single crystal is elongated jumpwise along this axis with lowering temperature. A charge-ordering peak observed for the *c* axis indicates that the single crystal is contracted along this direction. The anomaly in α_c is much smaller than in α_b, but both peak shapes suggest that the charge ordering has traits of the first-order phase transition. The fact that the charge and magnetic orderings are accompanied by the noticeable anomalies in the $\alpha_i(T)$ dependences points to the strong influence of the charge and magnetic subsystems on the lattice degrees of freedom in this compound.

We note in conclusion that the distribution of magnetically active V^{4+} ions over the crystallographic positions at low temperatures plays the decisive role in establishing the ground-state parameters of β-$Na_{0.33}V_2O_5$ and searching for the mechanisms responsible for the superconducting transition under pressure. This information can likely be obtained only from the resonance and neutron diffraction studies in conjunction with a certain model of the distribution of the V^{4+} and V^{5+} ions. The information obtained in this work and, in particular, the data on the thermal expansion anisotropy in β-$Na_{0.33}V_2O_5$ can be helpful in constructing such a model.

This work was supported by the Russian Foundation for Basic Research (project no. 03-02-16108), CDRF (grant no. RU-PI-2599-MO-04), and Deutsche Forschungsgemeinschaft (SFB 608).

REFERENCES

1. Y. Ueda, M. Isobe, and T. Yamauchi, J. Phys. Chem. Solids **63**, 951 (2002).
2. H. Yamada and Y. Ueda, J. Phys. Soc. Jpn. **68**, 2735 (1999).
3. T. Yamauchi, Y. Ueda, and N. Mori, Phys. Rev. Lett. **89**, 057002 (2002).
4. H. Onoda, T. Takahashi, and H. Nagasava, J. Phys. Soc. Jpn. **51**, 3868 (1982).
5. C. Presura, M. Popincius, P. H. M. van Loosdrecht, *et al.*, Phys. Rev. Lett. **90**, 026402 (2003).
6. A. N. Vasil'ev, V. I. Marchenko, A. I. Smirnov, *et al.*, Phys. Rev. B **64**, 174403 (2001).

Translated by V. Sakun

JETP Letters, Vol. 79, No. 11, 2004, pp. 545–549. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 674–679.

Effect of the Spin–Orbit Interaction on the Cyclotron Resonance of Two-Dimensional Electrons

Yu. B. Vasil'ev*, S. D. Suchalkin, S. V. Ivanov, B. Ya. Mel'tser, and P. S. Kop'ev

Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia

** e-mail: yu.vasilyev@mail.ioffe.ru*

Received April 29, 2004

In the cyclotron resonance (CR) spectra of two-dimensional (2D) electrons in InAs quantum wells, the CR line splitting is observed. The splitting is found to be an oscillating function of magnetic field. The oscillations do not correlate with the filling factor. The experimental results are interpreted in terms of the spin–orbit splitting in the presence of a built-in electric field appearing due to the asymmetry of the quantum-well potential. From the splitting of the CR line, the spin–orbit coupling constant α_{so} is determined. The resulting value agrees well with the value obtained for the same sample from the Shubnikov–de Haas oscillations. The role of the resonance interaction of charge carriers in the well with the interface donor states is discussed. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 73.21.Fg; 76.40.+b

In recent years, spin effects in low-dimensional semiconductor systems have attracted considerable interest in connection with the possibility of developing new devices based on the use of spin-polarized charge carriers [1]. In particular, the possibility of using the spin–orbit interaction in a semiconductor heterostructure for the control of the electron-gas spin polarization by an external electric field was investigated.

In a system without an inversion center [2], the Hamiltonian of the charge carriers can contain a component that is linear in the wave vector and caused by the presence of the effective electric field perpendicular to the direction of carrier motion. In this case, the carrier energy spectrum exhibits spin splitting even in the absence of an external magnetic field. The splitting energy is proportional to the charge-carrier wave vector. For the carriers in the quantum well, the effective electric field can be caused by the interatomic electric fields if the material of the quantum well has no inversion center. Another source of the built-in electric field may be the difference in the concentrations of ionized impurities on the quantum-well walls. In the latter case, the main parameter characterizing the strength of spin–orbit interaction is the spin–orbit coupling constant α_{so}, which is proportional to the built-in electric field. In an external magnetic field, instead of two spin subbands, two "ladders" of Landau levels appear with different cyclotron energies. The difference between the cyclotron energies of two subbands can be measured experimentally.

The cyclotron resonance (CR) allows the direct measurement of the cyclotron-energy difference in spin-split subbands under condition that the effect of spin–orbit interaction on the CR spectra is stronger than the effect of the conduction-band nonparabolicity. The results of such measurements can be used to obtain an independent (in addition to the conventional analysis of the Shubnikov–de Haas oscillation beats [3–9]) experimental estimate of α_{so} and also to gain information on the degree of carrier spin polarization in the quantum well. The narrow-band A3B5 heterostructures, which are characterized by a relatively small effective charge-carrier mass and, hence, a high cyclotron energy in a given magnetic field, seem to be particularly promising for such experiments.

In this paper, we report the observation of the spin–orbit splitting in the CR spectra of two-dimensional (2D) electrons in InAs quantum wells. The CR splitting is found to exhibit oscillations periodic in the inverse magnetic field which do not correlate with the variation of the filling factor. We show that the results of measurements are adequately described in terms of the Rashba model [2].

In the experiment, we measured the optical absorption at the cyclotron frequency in a magnetic field at a temperature of 4.2 K in the Faraday configuration with the use of a Fourier spectrometer. The samples were two InAs quantum wells grown by molecular beam epitaxy. Simultaneously with the CR, we measured the magnetoresistance in the two-probe geometry [10]. Each of the two structures under study contained a 20-nm-thick well bounded by AlSb barriers and a capping GaSb layer protecting the structure from oxidation. The distance between the well and the surface was different in different samples. Sample N1 contained a 2-nm-thick AlSb barrier separating the well and a GaSb surface layer with a thickness of 28 nm. Sample N2 had a more complex structure. Between the well and the

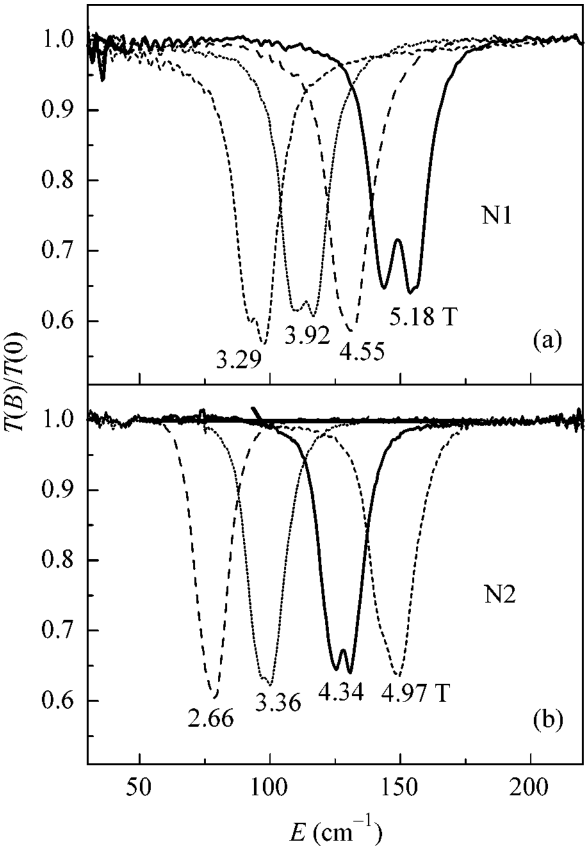


Fig. 1. Examples of CR spectra in different magnetic fields for two samples: (a) sample N1 and (b) sample N2. The numbers near the curves indicate the magnetic field in teslas.

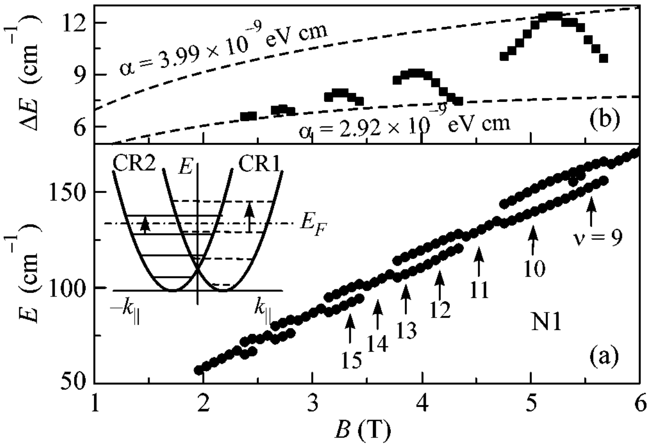


Fig. 2. (a) Dependence of the energy of absorption maximum on the magnetic field for sample N1. The arrows indicate the positions of the corresponding filling factors for the electron concentration determined from the areas of the CR curves. The inset schematically represents the electron energy spectrum. (b) The energy difference between two CR peaks as a function of magnetic field. The dots represent the experimental data, and the dashed lines are the result of calculations.

10-nm-thick GaSb surface layer, it had a 5-nm-thick AlSb layer and a 10-nm-thick AlGaSb layer. The electron concentration in the well was relatively high (on the order of 10^{12} cm^{-2}) in both samples, although no additional doping was used. Possible electron sources in the well could be the donor states at the surface of the GaSb antioxidant layer and the uncompensated deep donor levels in AlSb barriers [11, 12].

The CR spectra of both samples showed a considerable (on the order of 1 meV) CR line splitting (Fig. 1). All spectra in Fig. 1 are normalized to the zero-field spectrum. The number of peaks in a spectrum and the "distance" between them depend on the magnetic field.

Figures 2a and 3 show the magnetic-field dependences of the energy corresponding to the absorption maxima in samples N1 and N2, respectively. The dependences were measured by sweeping magnetic field with a step of 0.07 T. In Fig. 2a, the arrows indicate the positions of the corresponding filling factors for an electron concentration of 1.16×10^{12} cm^{-2} in the well, which was determined from the CR curve areas. Figure 2b shows the energy difference between the two CR peaks as a function of magnetic field for the case where the splitting of the CR line takes place. One can see that this dependence exhibits oscillations periodic in an inverse magnetic field. An unexpected experimental fact is that there is no correlation between the CR splitting and the filling factor. Figure 3 additionally shows the dependence of magnetoresistance on magnetic field for sample N2. The arrows indicate the positions of the corresponding filling factors for an electron concentration of 8×10^{11} cm^{-2} in the well, which was determined from the Shubnikov–de Haas oscillations. The electron concentration determined from the areas of the CR curves yields close values, although they are somewhat greater: (8.5–10) $\times 10^{11}$ cm^{-2}. For this sample, we also observed an oscillatory dependence of the CR-splitting energy on magnetic field (see inset in Fig. 3). One can see that the frequencies of the Shubnikov–de Haas oscillations and the oscillations of the CR-splitting energy are noticeably different for this sample. Thus, in both samples studied in our experiment, the CR line splitting proves to be unrelated to the filling of the Landau levels.

Many arguments can be found in favor of the assumption that the CR splitting in the samples under study is not determined by the nonparabolicity of the conduction band that was responsible for the previously observed CR line splitting in the InAs quantum wells (see [13, 14]). In the case of nonparabolicity [13, 14], the splitting necessarily occurs in the vicinity of integer (both even and odd) filling factors and in sufficiently high magnetic fields (above 4 T). The magnitude of this splitting depends on a number of parameters, including magnetic field and electron concentration. The splitting measured in our experiments is approximately five times greater than the splitting reported in [13, 14] for close values of magnetic field and electron concentration. The main distinction is that the splitting of the CR line in our experiments does not correlate with the filling of the Landau levels.

A possible explanation for the splitting could be the filling of the second size-quantization level. However, only the lower level is filled in our samples, which is confirmed both by the theoretical calculations and experimental data [15]: for a 20-nm well, the filling of the second size-quantization level occurs when the electron concentration exceeds 1.4×10^{12} cm^{-2}. In addition, as is seen from Fig. 1, the areas under the curves of two CR peaks and, hence, the electron concentrations in both subbands are approximately equal, which is impossible for two size-quantization subbands.

Another possible explanation of the CR splitting is the spin–orbit interaction, namely, the Rashba effect. An experimental confirmation of this hypothesis is the observation of the beats in the Shubnikov–de Haas oscillations that indicate the presence of a built-in electric field in the structures under study. In samples with identical quantum wells but different thicknesses of the barrier layers and different distances to the sample surface, the splittings are different. In addition, in similar samples fabricated on other technological setups and possessing weaker built-in electric fields, no such effects were observed (unpublished data). Hence, we can conclude that the splitting magnitude depends on the built-in electric field and on the degree of electron wave-function penetration into the barriers. The electron energy spectrum is adequately described by the Rashba model. According to this model, the spin–orbit interaction removes the spin degeneracy of electron levels and leads to the spin splitting near the Fermi level in the absence of magnetic field (inset in Fig. 2a). In a magnetic field, each spin subband is additionally split into a series of Landau levels. The radiation absorption at the spin-conserving transitions between the Landau levels is possible within each of the subbands. In this case, the cyclotron frequency will be different for different spin subbands, because it depends on the spin–orbit interaction. From the difference between the cyclotron frequencies, one can determine the spin–orbit coupling constant.

The spin–orbit interaction is described by the Hamiltonian [2]

$$H_{so} = \alpha_{so}[\sigma \times \mathbf{k}]\mathbf{v}, \tag{1}$$

where σ is the Pauli matrix, $\mathbf{v}$ is the unit vector perpendicular to the well plane, and $\mathbf{k}$ is the electron wave vector along this plane. The electron dispersion curve is split into two spin-up and spin-down branches:

$$E_n^{\pm}(k) = \frac{(\hbar^2)k^2}{2m(*)} \pm \alpha_{so}. \tag{2}$$

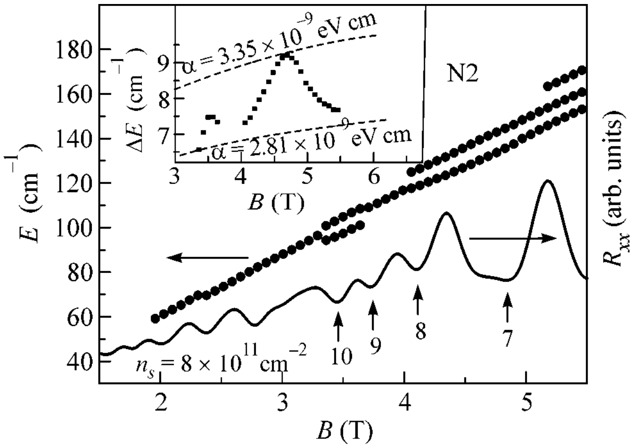


Fig. 3. Dependences of the energy of absorption maximum and the magnetoresistance on magnetic field for sample N2. The arrows indicate the positions of the corresponding filling factors. The inset shows the energy difference between two CR peaks as a function of magnetic field. The dots represent the experimental data, and the dashed lines are the result of calculations.

In a magnetic field perpendicular to the well plane, each branch has its own ladder of Landau levels:

$$E_N^{+} = \hbar\omega_c(N + \sqrt{(\delta^2 + (\gamma^2)N}),$$

$$E_M^{-} = \hbar\omega_c(M + \sqrt{(\delta^2 + (\gamma^2)M}),$$

where N and M are the integer numbers of Landau levels for the spin-up and spin-down subbands, respectively;

$$\delta = \frac{1}{2}\left(1 - \frac{g(*)m(*)}{2m_0}\right), \quad \gamma = \frac{2}{\hbar}\sqrt{\frac{m(*)\alpha_{so}^2}{2\hbar\omega_c}}.$$

The cyclotron transitions occur within each of the Landau-level ladders without changing spin. Since the CR energy depends on the spin–orbit interaction, the spin–orbit coupling constant can be determined from the CR line splitting:

$$\Delta E = \hbar\omega_c(\sqrt{(\delta^2 + (\gamma^2)(N+1)} - \sqrt{(\delta^2 + (\gamma^2)N}$$
$$+ \sqrt{(\delta^2 + (\gamma^2)(M+1)}\sqrt{(\delta^2 + (\gamma^2)M}).$$

Figure 2b and the inset in Fig. 3 show the results of fitting for the dependence describing the CR splitting as a function of magnetic field for samples N1 and N2, respectively. The symbols in the plots are for the experimental data, and the dashed lines are the results of calculation for the maximal and minimal CR line splitting according to Eq. (5) under the assumption that the effective g factor is equal to –13 [16]. The choice of the value of the g factor does not play any significant role in estimating the constant α_{so}, because the fitting procedure is highly sensitive to the choice of the value of α_{so} and shows much weaker dependence on the g factor.

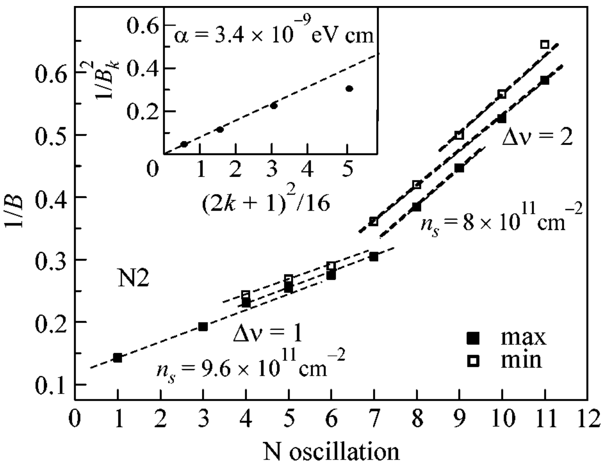


Fig. 4. Positions of the maxima and minima of magnetoresistance in the inverse magnetic field with arbitrarily chosen oscillation numbers. The inset shows the dependence of the beating node positions on the node number.

One can see from Fig. 2b and the inset in Fig. 3 that α_{so} varies from 2.92×10^{-9} to 3.99×10^{-9} eV cm for sample N1 and from 2.81×10^{-9} to 3.35×10^{-9} eV cm for sample N2.

The beats observed in the Shubnikov–de Haas oscillations for the structures under study are usually interpreted as a characteristic property of the systems with inversion asymmetry. Figure 3 shows a magnetoresistance curve whose shape differs from the classical oscillations presumably because the measured resistance contains the contributions from both longitudinal and transverse (Hall) components. One can clearly see the magnetoresistance oscillations. The positions of their maxima and minima in the inverse magnetic field are shown in Fig. 4, with the oscillation numbers chosen in an arbitrary way. From the slope of these curves, we determined the charge carrier concentration in the well. We obtained a value of 9.6×10^{11} cm^{-2} for magnetic fields above 3 T and approximately 8×10^{11} cm^{-2} for magnetic fields below 3 T. These values are close to the estimates of the carrier concentration from the areas under the CR curves: 10.1×10^{11} cm^{-2} for weak magnetic fields (B = 2.45 T) and (8.5–9) $\times 10^{11}$ cm^{-2} for strong fields (B = 4.8–5 T). An interesting feature of the observed oscillations is that the period of oscillations in the inverse magnetic field corresponds to a change of a factor of two in the filling factor when B is smaller than 3 T and a change of a factor of one when B is greater than 3 T. In addition, as the magnetic field increases, the oscillation phase varies: as is seen from Fig. 4, a continuous linear dependence on the inverse magnetic field takes place only for certain groups of extrema. The change in the phase can be associated with the appearance of a beating node in the corresponding magnetic field, even though the beats in the oscillations are weak presumably because of the two-probe geometry of our measurements. Using the method proposed in [17, 18], we constructed the plots of beating-node positions against the node number (inset in Fig. 4). One can see that the resulting dependence is linear (the straight line in the inset in Fig. 4). From the slope of this dependence, we determined the value $\alpha_{so} = 3.4 \times 10^{-9}$ eV cm, which agrees well with the value obtained from the CR line splitting.

An important result of our study consists in the experimental observation of the oscillations of the CR splitting and, hence, of the constant α_{so} with changing magnetic field, which are unrelated to the filling factor. This effect was observed by us for both samples studied. Although a detailed explanation of this effect requires further investigation, we can expect that the oscillations of α_{so} observed in our experiment could be due to the resonance interaction between the electrons in the well and the donor states in the barriers [10]. The penetration of the electron wave functions into the barrier increases when the energy of Landau level coincides with the energy of the donor level in the barrier. This changes the spin–orbit interaction and causes spin-splitting oscillations as the Landau levels sequentially cross the donor level upon sweeping the magnetic field. In this case, the role of Fermi energy is played by the energy of donor state measured from the size-quantization level in the well. This interaction is analogous to the resonance interwell tunneling in a double quantum well, where the effective mass in one of the wells is much greater than in the other. It is important to note that our interpretation does not contradict the existing concepts. Indeed, it is known that the spin–orbit interaction strongly depends on the strength of the built-in electric field and on the degree of electron wave-function penetration into the barrier. The latter property was studied in detail and experimentally verified in [8]. Moreover, it is known that the surface donor states can arise at the heteroboundary between the InAs well and the AlSb barriers. The energy of such donor states was estimated in [10] at approximately 45 meV above the bottom of the quantum-well band. This value agrees well with the period of oscillations of α_{so} measured in our experiments.

It is necessary to make the following comments. First, our consideration completely ignored the nonparabolicity effects. This is valid for weak (below 4 T) magnetic fields, when no corresponding CR splitting is observed. However, in strong magnetic fields, the role of nonparabolicity increases and can affect the CR spectra. For example, the double peak observed in the absorption spectrum at B = 5.18 T near 155 cm^{-1} (Fig. 2a) occurs because of the nonparabolicity of the conduction band. We also note that the calculated values of α_{so} shown in Figs. 2 and 3 by the dashed lines were obtained without taking into account the nonparabolicity of conduction band, because the effect of nonparabolicity is noticeable only in strong magnetic fields. Our estimates show that the neglect of nonparabolicity introduces an error of about 10%.

Second, one should not directly relate the spin–orbit coupling constant to the difference in the populations of spin subbands, as is done in some publications. If such a difference in population were present, it would cause a spontaneous polarization in zero magnetic field (while no preferred directions are present and no definite direction of such a polarization can be determined). In fact, as one can see from the inset in Fig. 2a, the populations of the two spin subbands are identical. We also note that the difference in the populations of subbands in a magnetic field cannot account for α_{so}. Electrons are redistributed between different states (e.g., in the spin subbands and the impurity states) according to the position of the Fermi level, and their redistribution is associated with the filling factor.

Third, the CR-splitting oscillations observed in our experiments cannot be associated with the modulation of the *g* factor, because the latter is determined by the charge-carrier redistribution of the Landau levels in magnetic field and, hence, must correlate with the filling factor. Such a correlation between the effective *g* factor and the filling factor was observed in the experiment [19].

We are grateful to R. Sergeev, R. Suris, and I. Kukushkin for useful discussions. This work was supported by the Russian Foundation for Basic Research and by the programs “Quantum-Dimensional Structures” and “The Physics of Solid-State Nanostructures.”

REFERENCES

1. B. Datta and S. Das, Appl. Phys. Lett. **56**, 665 (1990).
2. E. I. Rashba, Fiz. Tverd. Tela (Leningrad) **2**, 1224 (1960) [Sov. Phys. Solid State **2**, 1109 (1960)]; Yu. A. Bychkov and E. I. Rashba, Pis’ma Zh. Éksp. Teor. Fiz. **39**, 66 (1984) [JETP Lett. **39**, 78 (1984)].
3. J. Luo, H. Munekata, F. F. Fang, *et al.*, Phys. Rev. B **38**, 10142 (1988); J. Luo, H. Munekata, F. F. Fang, *et al.*, Phys. Rev. B **41**, 7685 (1990).
4. B. Das, D. C. Miller, S. Datta, *et al.*, Phys. Rev. B **39**, 1411 (1989); B. Das, S. Datta, and R. Reifenberger, Phys. Rev. B **41**, 8278 (1990).
5. G. Engels, J. Lange, Th. Schapers, *et al.*, Phys. Rev. B **55**, R1958 (1997).
6. J. Nitta, T. Akazaki, H. Takayanagi, *et al.*, Phys. Rev. Lett. **78**, 1335 (1997).
7. S. Brosig, K. Ensslin, R. J. Warburton, *et al.*, Phys. Rev. B **60**, R13989 (1999).
8. D. Grundler, Phys. Rev. Lett. **84**, 6074 (2000).
9. C.-M. Hu, J. Nitta, T. Akazaki, *et al.*, Phys. Rev. B **60**, 7736 (1999).
10. H. Kroemer, C. Nguyen, and B. Brar, J. Vac. Sci. Technol. B **10**, 1769 (1992).
11. D. J. Chadi, Phys. Rev. B **47**, 13478 (1993).
12. In the fabrication of Hall bars, there is a possibility that the built-in electric field formed in the sample may change after chemical etching.
13. J. Scriba, A. Wixforth, J. P. Kotthaus, *et al.*, Semicond. Sci. Technol. **8**, S133 (1993).
14. M. J. Yang, R. J. Wagner, B. V. Shanabrook, *et al.*, Phys. Rev. B **47**, 6807 (1993).
15. A. C. H. Rowe, J. Nehls, R. A. Stradling, *et al.*, Phys. Rev. B **63**, 201307 (2001).
16. C.-M. Hu, C. Zehnder, Ch. Heyn, and D. Heitmann, Phys. Rev. B **67**, 201302R (2003).
17. S. I. Dorozhkin and E. B. Ol’shanetskiĭ, Pis’ma Zh. Éksp. Teor. Fiz. **46**, 399 (1987) [JETP Lett. **46**, 502 (1987)].
18. S. I. Dorozhkin, Phys. Rev. B **41**, 3235 (1990).
19. R. J. Nicholas, R. J. Haug, K. von Klitzing, *et al.*, Phys. Rev. B **37**, 1294 (1988).

Translated by E. Golyamina

JETP Letters, Vol. 79, No. 11, 2004, pp. 550–554. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 680–685.

Photoconduction and Photocontrolled Collective Effects in the Peierls Conductor TaS_3[¶]

S. V. Zaitsev-Zotov* and V. E. Minakova

Institute of Radioengineering and Electronics, Russian Academy of Sciences,
Mokhovaya ul. 11, bld. 7, Moscow, 125009 Russia
** e-mail: SerZZ@cplire.ru*

Received April 29, 2004

Light illumination of thin crystals of CDW conductor TaS_3 is found to result in dramatic changes of both linear (G) and nonlinear conduction. The increase of G is accompanied by suppression of the collective conduction, growth of the threshold field E_T, and appearance of the switching and hysteretic behavior in the nonlinear conduction. The effects in the nonlinear conduction are associated with increase of CDW elasticity due to illumination that leads, in particular, to the appearance of a relation $E_T \propto G^{1/3}$ expected for the one-dimensional pinning. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 72.15.Nj; 72.40.+w

Quasi-one-dimensional (quasi-1D) conductors with charge-density-waves (CDWs) [1] are one of the most interesting physical systems with collective electron transport. The interaction between electrons condensed into the CDW dominates in elastic properties of the electron crystal–CDW. The elastic properties of the CDW affect such characteristics of quasi-1D conductors as the value of the threshold field for CDW sliding, E_T, phase-correlation length, dielectric constant, etc. In turn, the elastic properties are controlled by quasiparticles (electrons and holes) thermally excited over the Peierls gap in the energy spectrum and screening the electric fields caused by CDW deformations. Thus, a variation of the quasiparticle concentration (or the total carrier concentration) may be a tool for controlling the properties of CDW conductors. An attempt to vary the total carrier concentration has been undertaken in the field-effect experiment [2]. In particular, it was found that a 1% change in the total concentration of the current carriers by the transverse electric field leads to a 40% change in the threshold field value. Another well-known way to modify the carrier concentration is excitation of nonequilibrium current carriers by light. For example, illumination of a semiconductor may result in an increase of the carrier concentration by several orders of magnitude. Such a change can be easily detected as a variation of the conduction (photoconduction). Photoconduction is one of the most fruitful methods to study the details of the energy structure, current-carrier recombination time, and other semiconductor parameters. The similarity of numerous physical properties of the CDW conductors and semiconductors arising from the existence of the gap in the electron-state density is well known [3].

Several attempts of experimental search for photoconduction of CDW materials [4–7] reveal contradictory results. In [4, 5], no noticeable photoconduction in TaS_3 was observed. Instead, the bolometric response was found and employed for detailed study of the energy structure in TaS_3. In addition, an enhancement of the bolometric response was reported in the nonlinear regime [4]. In [6], photoinduced CDW conduction was observed in blue bronze $K_{0.3}MoO_3$. The red boundary of the effect was found to correspond to the Peierls gap. The phenomenon was associated with initiation of the CDW depinning by optically excited single electrons. No light-induced variation of the linear conduction or the threshold field was reported. In [7], photoinduced modification of the dynamic transition from slide to creep in $K_{0.3}MoO_3$ was reported: light illumination was found to increase E_T and the CDW-creep rate. The origin of the effect was attributed to a local destruction of the CDW, which led to the photoinduced phase slip and the redistribution of the CDW phase. No effect of illumination on the linear conduction was reported. Thus, despite some similarity between CDW conductors and semiconductors, no photoresponse in the linear conduction was found in 25 years of study of CDW materials. It is worth noting that the absence of photoconduction would agree with theories predicting very small quasiparticle lifetime of the order of 10^{-12} s [8]. Femtosecond spectroscopy study of $K_{0.3}MoO_3$ has shown that the electron–hole recombination time is short, indeed, on the subpicosecond scale [9]. From this point of view, quasiparticles (electrons and holes) are ill-defined physical objects, and the absence of photoconduction is a feature of CDW conductors. So, observation of photoconduction is essential for the physics of CDW conductors.

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

Here, we show that photoconduction of quasi-1D conductor TaS_3 can be directly observed in low-frequency conduction measurements. In particular, we show that the linear conduction may be increased up to an order of magnitude under light illumination. The resulting changes of the nonlinear conduction are dramatic and reveal themselves in a substantial growth of the threshold field, suppression of the nonlinear conduction near E_T, and appearance of the switching behavior. We also show that the observed growth of E_T can entirely be accounted for by photoinduced increase of the CDW elasticity. Thus, no exotic assumptions on photoinduced phase slip [7] or CDW depinning [6] are required. Moreover, our results clearly demonstrate the opposite effect—suppression of the collective conduction under light illumination.

Orthorhombic TaS_3 is a typical Peierls conductor. In this material, the CDW formation at $T_p = 220$ K is accompanied by the complete dielectrization of the electron spectrum. All five studied samples (fabricated by splitting of high-quality crystals) having cross-section areas of $0.002 \le s \le 0.15$ μm^2 demonstrated qualitatively similar behavior. Such thin crystals were chosen to enhance the contribution of the region affected by the illumination [10]. In our samples, the photoconduction developed in the entire sample, rather than in a surface layer shunted by bulk as in [4–7]. Using thin samples allows one also to suppress the heating effect due to exceptionally good thermal contact with sapphire substrate. In addition, growth of potential relief due to finite-size effects [13] leads to substantial photocondution growth due to spatial separation of photoexcited electrons and holes. Current terminals were made by indium cold soldering, and all measurements were performed in the two-terminal regime. IR LED with intensities of $(10^{-6}$–$1)$ W/cm^2 at the sample position and with a wavelength of $\lambda = 0.94$ μm was used. Thus, the photon energy was higher than the optical gap value of 125 meV [5].

Figure 1 shows a set of temperature dependences of conduction, $G(T, W)$, for various intensities W of steady light illumination for the thickest sample. Noticeable deviation from the darkness curve starts at $T < 70$ K. The inset shows the respective set of photoconductions $\delta G = G(T, W) - G(T, 0)$. The position of the maximum of $\delta G(T)$ varies with W in the range 40–65 K. The deviation starts at somewhat higher temperatures for thinner samples. For example, for the thinnest one having $s = 0.002$ μm^2, $\delta G/G \sim 0.01\%$ was observed at $T = 100$ K. The maximum of $\delta G(T)$ at $W = 110$ mW/cm^2 is at $T = 65$ K. The upper boundary of the light-induced heating may be estimated, using $\delta G(T)$ at $T = 100$ K, to be as small as $(\delta G/G)/(\delta G/dT) \sim 1$ mK, as a consequence of the exceptionally good thermal contact between the sample and substrate. All the results presented below were obtained for this thinnest sample having length $l = 340$ μm and room-temperature resistance $R_{300} = 430$ kΩ.

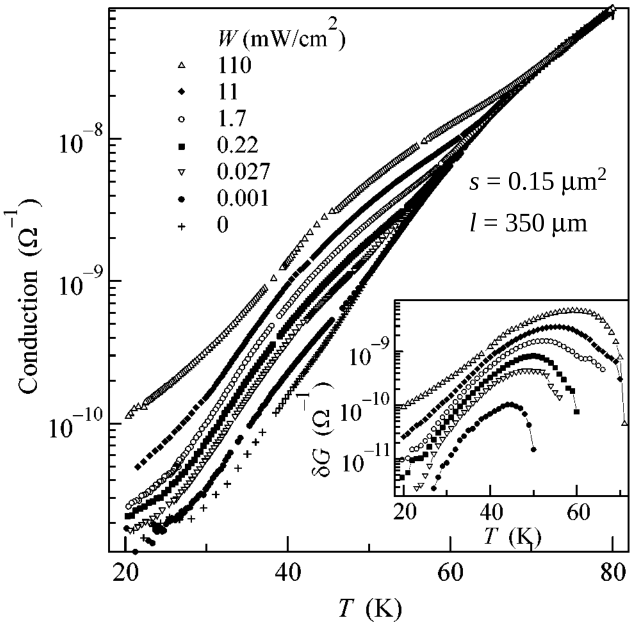


Fig. 1. Temperature dependences of conduction G for different levels of steady light intensity W. Inset: temperature dependences of photoconduction $\delta G = G(T, W) - G(T, 0)$. All curves were measured upon cooling to eliminate the effects associated with conduction hysteresis.

The dependence of the value of photoconduction δG on the light illumination intensity W was studied by two methods. In the temperature range 27–95 K, AC conduction was measured at the frequency $f = 4.5$ Hz as a function of intensity W of a steady light illumination. Such a low frequency value was chosen to enable measurements of very low conduction values. At temperatures $T > 90$ K, the ratio $\delta G/G$ becomes too small and comparable with the one resulting from temperature fluctuations. To improve measurement quality, the double modulation method was used at $T = 81$–100 K. Namely, the modulation of AC conduction due to the light illumination was detected at the light-chopping frequency $f_{ch} = 4.5$ Hz, AC conduction being measured at the frequency $f = 335$ Hz. The results are presented in Fig. 2. The dependences can be approximated by the power law $\delta I = W^{\alpha}$, where the photocurrent $\delta I = \delta GV$. These dependences are highly nonlinear in the middle of the temperature range ($\alpha < 1$) and approach linear ones ($\alpha \approx 1$) at its ends. The temperature dependence of the exponent α determined as $\alpha = d\ln\delta I/d\ln W$ at $W > 10$ mW/cm^2 is shown in the inset in Fig. 2. This dependence has a pronounced minimum at $T = 50$ K, where a change of up to four orders of magnitude in the light intensity leads to an increase of only an order of magnitude in the linear conduction.

At sufficiently low temperatures, the light-induced conduction variation becomes so large that it can be clearly seen in I–V curves. Figure 3 shows the evolution of I–V curves (plotted as $G = I/V$ versus V) caused by light illumination at $T = 41$ K. In the dark, the $G(V)$

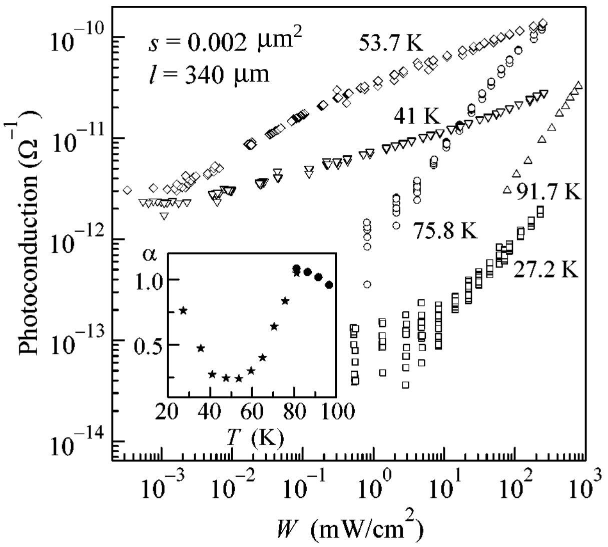


Fig. 2. Photoconduction δG as a function of light intensity W for different temperatures. Inset: temperature dependence of the exponent $\alpha = d\ln\delta I/d\ln W$ (different symbols correspond to different methods of measuring; see text).

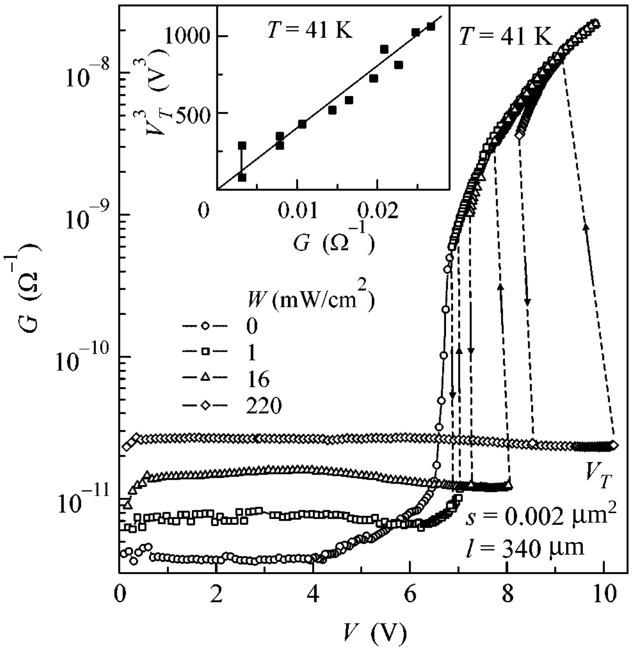


Fig. 3. The conduction $G = I/V$ as a function of voltage V on the sample for different levels of stationary illumination W. Arrows indicate the direction of changing current. Inset: V_T^3 as a function of linear conduction G (measured for a voltage of 100 mV on the sample) for different levels of illumination. Vertical bars correspond to the uncertainty in determining V_T for $W = 0$ and $W = 1$ mW/cm^2.

curve has the usual shape: there is a region of a constant (linear) conductivity in a small electric field, then a region of creep (weak nonlinearity), and a region of CDW sliding (strong nonlinearity). The curve has a smooth character without any switching or hysteresis. Dramatic changes of the shape of $G(V)$ curves in both the linear and nonlinear conduction take place under light illumination. It can be seen that the growth of the light intensity causes

(1) an increase of the linear conduction up to 10 times;

(2) a decrease of the conduction with voltage growth in the creeping region;

(3) an increase of up to 60% in the threshold field for the onset of CDW sliding (decrease of the nonlinear conduction near V_T); and

(4) the appearance of the switching regime (unusual for TaS$_3$ samples) with hysteretic character [11]. The growth of the threshold field and appearance of the switching under light illumination was also observed in K$_{0.3}$MoO$_3$ [7].

Figure 4 shows the temperature evolution of $G(V)$ dependences measured under steady light illumination ($W = 117$ mW/cm^2) and in the dark. At temperatures above 65 K (corresponding to the maximum value of photoconduction for this sample), changes of the shape of $G(V)$ curves under light illumination are seen only in the linear conduction region and in the region of creep. With temperature decrease, the dependences $G(V)$ under light illumination begin to deviate from the dark ones in the CDW sliding region as well. The transition to the sliding regime becomes sharper, the switching behavior appears, and the dramatic growth in the width of the hysteresis loop develops.

It is worth noting that all the $G(V)$ curves coincide in the high voltage region $E \gg E_T$. This indicates that (i) the relative variation of the nonlinear conduction due to illumination is much less than for the linear conduction and (ii) the heating of the sample due to illumination is small (heating leads to an increase of the conduction and decrease of the threshold field).

Our results allow one to estimate the recombination time τ of photoexcited carriers in TaS$_3$. The photoexcited current-carrier concentration Δn is determined by the balance between the rate of their photogeneration, $kaLn_{\mathrm{ph}}(1 - e^{-\beta b})$, and the relaxation rate $abL\Delta n/\tau$ (we assume $\Delta n \ll n$), where k is the quantum efficiency of photogeneration; n_{ph} is the number of incident photons per unit area and unit time; β is the absorption coefficient; and a, b, and L are the sample width, thickness, and length, respectively. Thus, for $\beta b \ll 1$, one gets $\tau = \Delta n/k\beta n_{\mathrm{ph}}$. At $T = 100$ K, $\Delta n(\propto\delta G) \propto n_{\mathrm{ph}}(\propto W)$, i.e., τ is independent of the light intensity W. As was mentioned above, $\delta G/G \sim 0.01\%$ for this sample at $T = 100$ K. As $\Delta n/n(T) = \delta G/G(T)$, $n_{100} \approx 10^{-3}n_{300}$, $n_{300} \approx 2 \times 10^{21}$ cm^{-3}, so that $\Delta n \approx 2 \times 10^{14}$ cm^{-3}. The value of k is unknown. Assuming $k = 1$ and $1/\beta = 0.3$ μm [10], for $W = 110$ mW/cm^2, one gets $\tau_{100} \approx 10^{-10}$ s. Moreover, this value is almost two orders of magnitude greater at 65 K and $W = 110$ mW/cm^2 (see Fig. 1). This time is much smaller than in pure semiconductors. This explains why the photoconduction in quasi-1D conductors was not found for a long time. On the other hand, τ in thin crystals of TaS_3 is several orders of magnitude greater than $\tau \approx 5 \times 10^{-13}$ s measured by optical methods in blue bronze at the same temperatures [9]. A possible physical mechanism providing enhancement of photoconduction in thin samples is given in the following paragraph.

In the temperature range 30 K < T < 70 K, there is a nonlinear relation between the light intensity and photoconduction (see inset in Fig. 2) and the photoconduction reaches its maximum (inset in Fig. 1). This means that, in this temperature range, the recombination time of photoexcited current carriers (i) depends on their concentration and (ii) is larger than outside this temperature range. These features are peculiar to so-called "persistent photoconduction" (also "delayed" or "frozen"), which is well known for some inhomogeneous semiconductors [12]. In our case, the barriers for recombination result from the potential relief caused by the pinned CDWs. The energy range for this relief can be estimated from the relation $\delta\zeta \sim E_T L_{\|}$, where $L_{\|}$ is the GDW phase-correlation length. At higher temperatures, E_T decreases (see Fig. 4), and finally one gets $\delta\zeta \ll T$. As a result, the relaxation barriers diminish and the relaxation becomes faster, with a rate practically independent of the light intensity. As the threshold field rapidly increases with temperature decrease, a situation where $\delta\zeta$ is on the order of the Peierls gap is achieved. In this case, one may expect nucleation of dislocations and opening of a new relaxation channel. This may explain the low-temperature decrease of photoconduction (see inset in Fig. 1). In addition, since E_T is illumination-dependent, the barrier height depends on the concentration of photoexcited current carriers. Taking $L_{\|}$ = 1–10 μm [13] and $E_T \approx 30$ mV/μm, for T = 41 K at W = 220 mW/cm^2 (see Fig. 3), one gets $\delta\zeta \simeq$ 300–3000 K $\gg T$, which is comparable to the Peierls energy gap of TaS_3 (1700 K). It is clear that such barriers have a pronounced effect on the recombination rate. Since E_T in thin samples is several orders of magnitude higher than its value in bulk, thin samples are preferable for observation of photoconduction.

Another interesting feature of I–V curves of illuminated samples is a decrease of conduction with growth of the voltage in the creep regime, which is clearly seen in Fig. 3 at $E \lesssim E_T$. From a formal point of view, this behavior corresponds to a negative contribution of creeping CDWs to the total conduction of the illuminated sample. We believe, however, that this behavior results from the decrease of the concentration of photoexcited current carriers in the creep regime and possible contribution caused by CDW configuration variation due to a change in current-carrier concentration (configurational photoconduction). Such a decrease corresponds to an increase of the recombination rate of photoexcited current carriers due to modification and a time evolution of the potential relief in the creep regime. Note that Ogawa *et al.* [7] attributed the growth of the total conduction at $E \lesssim E_T$ to the increase of CDW creep rate, i.e., to the opposite effect.

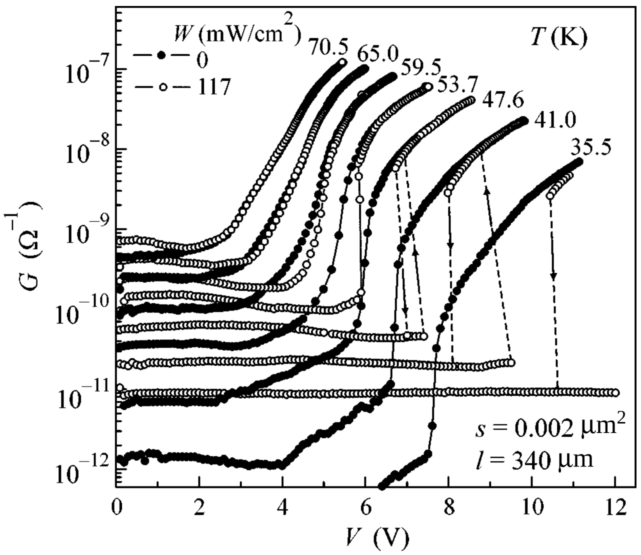


Fig. 4. The conduction G as a function of voltage V on the sample at different temperatures. Black circles correspond to the dark measurements, and light circles, to the steady illumination.

The illumination-induced variation of current-carrier concentration must affect all static and dynamic properties of the CDWs due to modification of the screening conditions. In particular, one can expect changes in the CDW wave vector, the relaxation rate of CDW metastable states, E_T, the dielectric constant ($\epsilon \lesssim 1/E_T$ [1]), the CDW transport coefficients, etc. Our results allow one to verify, for the first time, the relations between the screening carrier concentration, CDW elasticity, and E_T. As the transverse sizes of our samples are smaller than the transverse CDW phase-correlation length, CDW pinning is one-dimensional [13]. In this case, $E_T \approx (n_i w/K_{\|})^{1/3}$, where n_i is the impurity concentration, w is the pinning potential, and $K_{\|}$ is the elastic modulus of the CDW [1, 13]. Since $K_{\|} \propto 1/n$ [14], one has $E_T \propto n^{1/3}$. The inset in Fig. 3 shows the V_T^3 versus G dependence. The dependence is actually close to linear.

It is well known that the energy dissipation for sliding CDWs is provided by quasiparticles participating in the screening of the time-dependent CDW deformations. Thus, one could expect that growth of the current-carrier concentration due to illumination would enhance the CDW conduction. However, no noticeable modification of the CDW nonlinear conduction at $E \gg E_T$ is observed. Thus, we conclude that the concentration of photoexcited current carriers diminishes with growth of CDW velocity. Noticeable reduction of the photoconduction starts already in the CDW creep region (Fig. 3).

The origin of the switching regime is a subject of wide discussions [15]. This regime is observed in $NbSe_3$, where the dielectrization of the electron spectrum accompanied by CDW formation is not complete and leads to the presence of free carriers in the system at low temperatures. Our results prove that the origin of the switching effect deals with the appearance of extra current carriers in the system, irrespective of their nature—natural carriers, as in $NbSe_3$, or photoexcited, as in TaS_3 (present work) or in $K_{0.3}MoO_3$ [7].

In conclusion, it has been found that the light illumination of thin TaS_3 samples affects practically all their electrophysical properties. Namely, in illuminated samples we observed

(1) significant increase of the linear conduction;

(2) strong nonmonotonic temperature dependence of the photoconduction;

(3) highly nonlinear dependence of photoconduction on the light intensity;

(4) decrease of conduction with growth of the voltage in the creep regime;

(5) increase of the threshold field and suppression of the nonlinear conduction near V_T;

(6) negligible effect on the nonlinear conduction at $V \geq V_T$;

(7) appearance of switching in the I–V curves.

Observation of the photoresponse in the linear conduction is crucial for understanding the photoinduced changes in the CDW dynamics. Photoconduction in CDW conductors is a much more complex phenomenon than in usual semiconductors because of very strong coupling between the quasiparticles and the CDWs. The possibility of photocontrol of quasiparticle conduction opens wide prospects for investigations of various static and dynamic properties of quasi-1D conductors.

We are grateful to S.N. Artemenko and V.Ya. Pokrovskii for useful discussions and to R.E. Thorne for providing high-quality crystals. This work was supported by the RFBR (project no. 04-02-16509), Presidium of RAS, NWO, CRDF (RP2-2563-MO-03), and INTAS-01-0474.

REFERENCES

1. G. Grüner, Rev. Mod. Phys. **60**, 1129 (1988); P. Monceau, in *Electronic Properties of Inorganic Quasi-One-Dimensional Conductors*, Ed. by P. Monceau (Reidel, Dordrecht, 1985), Part 2.
2. T. L. Adelman, S. V. Zaitsev-Zotov, and R. E. Thorne, Phys. Rev. Lett. **74**, 5264 (1995).
3. S. N. Artemenko, V. Ya. Pokrovskiĭ, and S. V. Zaĭtsev-Zotov, Zh. Éksp. Teor. Fiz. **110**, 1069 (1996) [JETP **83**, 590 (1996)].
4. J. W. Brill and S. L. Herr, Phys. Rev. B **27**, 3916 (1983).
5. M. E. Itkis and F. Ya. Nad’, Pis’ma Zh. Éksp. Teor. Fiz. **39**, 373 (1984) [JETP Lett. **39**, 448 (1984)].
6. R. Gaál, S. Donovan, Zs. Sörlei, and G. Mihály, Phys. Rev. Lett. **69**, 1244 (1992).
7. N. Ogawa, A. Shiraga, R. Kondo, *et al.*, Phys. Rev. Lett. **87**, 256401 (2001).
8. S. A. Brazovskiĭ, Zh. Éksp. Teor. Fiz. **78**, 677 (1980) [Sov. Phys. JETP **51**, 342 (1980)].
9. J. Demsar, K. Biljakovic, and D. Mihailovic, Phys. Rev. Lett. **83**, 800 (1999).
10. The light penetration depth in TaS_3 is about 0.1–1 μm [M. E. Itkis, private communication].
11. I-V measurements were made in the voltage controlled regime with a load resistance 10 MOhm. As a result, the current jumps were not vertical for large currents.
12. B. A. Volkov, L. I. Ryabova, and D. R. Khokhlov, Usp. Fiz. Nauk **172**, 857 (2002) [Phys. Usp. **45**, 819 (2002)].
13. As a review of finite-size effects see: S. V. Zaitsev-Zotov, Usp. Fiz. Nauk **174**, 585 (2004).
14. S. N. Artemenko and A. N. Kruglov, Zh. Éksp. Teor. Fiz. **83**, 1134 (1982) [Sov. Phys. JETP **56**, 643 (1982)].
15. T. L. Adelman, J. McCarten, M. P. Maher, *et al.*, Phys. Rev. B **47**, 4033 (1993).

JETP Letters, Vol. 79, No. 11, 2004, pp. 555–558. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 686–690.

E–T Phase Diagram of Lead Magnesium Niobate Relaxor Ferroelectric in the Mandelstam–Brillouin Scattering Spectra

S. G. Lushnikov[1], *, J.-H. Ko[2], and Seiji Kojima[3]
[1] *Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia*
** e-mail: sergey.lushnikov@mail.ioffe.ru*
[2] *R&D Center, Samsung Corning Co., LTD, 472 Sin-dong, Paldal-gu, Suwon-si, Gyeonggi-do, Korea*
[3] *Institute of Materials Science, University of Tsukuba, Tsukuba city, Ibaraki 305-8573, Japan*
Received May 5, 2004

The results of studying the velocity of hypersonic longitudinal acoustic phonons over a wide temperature range under an applied constant electric field in the $PbMg_{1/3}Nb_{2/3}O_3$ (PMN) relaxor ferroelectric are reported. An analysis of the experimental data on the basis of the existing *E–T* phase diagrams shows the necessity of a substantial change in the current concepts. Two radically different states—with the reversible and irreversible behavior of velocity—have been observed experimentally in the induced ferroelectric phase. A modified *E–T* phase diagram of the PMN crystal is suggested. © *2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 77.84.-s; 77.84.Dy; 63.20.-e

$PbMg_{1/3}Nb_{2/3}O_3$ (PMN) crystals are used as models in studying the properties of a large family of perovskite-like ferroelectrics with the general formula $AB'B''O_3$ [1, 2], whose lattice structure and dynamics are actively studied by various methods [3–7]. In the PMN crystals, the aliovalent Mg^{5+} and Nb^{3+} ions are randomly distributed in the *B* sublattice of ABO_3 perovskite. As a result, the dynamics of phase transformations change radically: instead of the classical ferroelectric phase transition, a frequency-dependent anomaly smeared over hundreds of degrees and not associated with the structural ferroelectric phase transition appears in the dielectric response; a nonergodic phase arises; etc. These features of lattice dynamics served as a basis for naming the family of the crystals under discussion as relaxor ferroelectrics or relaxors. Interest in relaxors is caused by both the wide industrial use of their unique properties and the possibility of studying changes in the physical properties of a crystal upon disordering one of the sublattices.

The dynamics of phase transformations and, correspondingly, the phase diagram turn out to be even more complicated under an applied external electric field. The phase *E–T* diagram of a PMN crystal was constructed in [8] and developed in [9] on the basis of the dielectric measurements (Fig. 1a). For brevity, it will be named the Colla diagram. The analysis of the published data performed in review [10] allowed its author (Prof. Z.-G. Ye) to propose his own *E–T* diagram (Fig. 1b). The effect of various cooling regimes (field cooling, field-free cooling but field heating, etc.) on the experimental results was one of the main factors taken into account in [10]. This circumstance is crucial for partially disordered crystals. Therefore, the phase diagram constructed in [10] reflects not only the phase boundaries but also the properties that manifest themselves in various cooling regimes of the PMN crystal (for example, field-induced linear and nonlinear optical properties [11]). The appearance of an induced ferroelectric phase of the *R*3*m* symmetry is a common feature of the PMN crystal phase diagrams under discussion. However, according to [9], the ferroelectric (FE) phase arises upon cooling in an applied electric field with $E > E_{\mathrm{th}}$ (Fig. 1a), whereas two induced FE phases—with reversible and irreversible polarizations (regions *2* and *3* in Fig. 1b, respectively, separated by line *B*)—are recognized in the Ye phase diagram. Moreover, the boundary between the states with reversible and irreversible polarizations upon the reverse motion from region *2* to region *3* is shown inside region *2* with an additional dashed line. The question of the existence of a phase boundary separating the paraelectric (PE) and nonergodic (glasslike, GL) phases in low fields ($E < E_{\mathrm{th}}$) (Fig. 1a) remains open. Thus, at least two essentially different *E–T* phase diagrams have been published to date. Studies devoted to a direct experimental verification of the proposed PMN crystal phase diagram have not been reported so far. This paper presents the results of studying the behavior of longitudinal acoustic phonons in PMN crystals using Mandelstam–Brillouin light scattering (MBS) as a function of the applied electric field and invoking the analysis of the existing *E–T* phase diagrams of a PMN crystal.

A single crystal prepared in the form of a parallelepiped with sides 1 × 3 × 5 mm oriented as $[111]/[1\bar{1}0]/[11\bar{2}]$ was investigated in our experiments. The sample was oriented using a DRON-2 x-ray

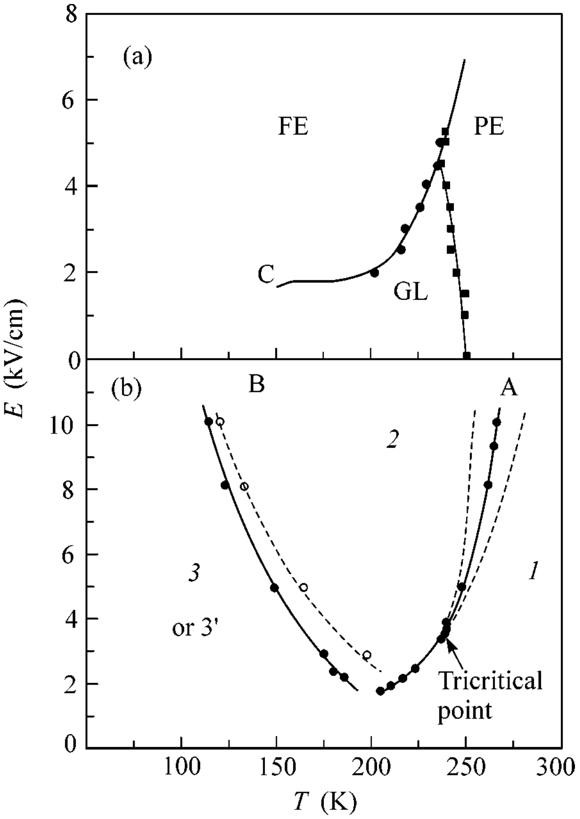


Fig. 1. PMN E–T phase diagram: (a) observed in the field-cooling regime (FC **E** || [111] regime), taken from [9]; (b) taken from [10], where A, B, and C are phase boundaries.

diffractometer. Two surfaces perpendicular to the [111] direction were coated with silver paste electrodes. High voltage from a stabilized power unit (HJPM-P, Matsusada Precision Inc.) was applied to the [111] direction of the PMN crystal. Light was excited using an ion Ar^+ laser with the wavelength $\lambda = 514.5$ nm, scattered from the sample, and analyzed by a high-contrast Sandercock 3+3 tandem Fabry–Perot interferometer. A 180° scattering geometry was used in the experiments, and the free spectral range of the tandem was 75 GHz. In the temperature measurements, the sample was placed in a cryostat cell (THMS 6000), whose temperature was varied from –190 to 600°C with a stabilization of ±0.1°C. In the course of measurements, we adhered to the following order: the sample was cooled in zero electric field, a constant electric field was applied upon reaching the specified temperature, after which the field magnitude was varied within the required limits. After the field measurements, the sample was heated to room temperature. Then, the cycle was repeated again.

In this work, scattering from the longitudinal acoustic (LA) phonons with $\mathbf{q}_{\mathrm{ph}} \parallel [1\bar{1}0]$ was studied. The frequency shift $\Delta\nu$, determined directly in the course of measurements, is related to the velocity of acoustic phonons V in our scattering geometry by the equation $V = \Delta\nu\lambda/(2n)$, where n is the refractive index of the crystal. The velocity of the LA phonon involved in scattering is determined by the equation $\rho V^2 = \frac{1}{2}(C_{11} + C_{12} + 2C_{44})$, where ρ is the crystal density and C_{11}, C_{12}, and C_{44} are elastic constants. In our earlier works, the relative variation of the velocity of hypersonic LA phonons was studied as a function of the direction of the applied electric field [12]. The largest variations of the velocity were observed for the field in the [111] direction. This fact determined the choice of the experimental geometry used in this work.

Let us consider the behavior of acoustic phonons upon varying the magnitude of applied external electric field at various temperatures. At temperature $T = 255$ K, the velocity of acoustic phonons as a function of the applied electric field exhibits two anomalies (Fig. 2a). The first anomaly is observed at $E_t \approx 1$ kV/cm, and the second is observed in the vicinity of 10 kV/cm. The reverse run, that is, the variation of the velocity upon a decrease in the applied field, is shown in Fig. 2a (and in the subsequent figures) by black circles. As is seen in the figure, it does not reproduce the initial field dependence of the velocity. Such a hysteresis corresponds to a first-order phase transition. At the next temperature point, $T = 240$ K, the first anomaly of E_t in the field dependence of the velocity is virtually absent. Only a weak kink is observed at 1.5 kV/cm (Fig. 2b), whereas the anomaly in the vicinity of 10 kV/cm does not change substantially. Let us compare the dependences under discussion with the corresponding region of the E–T phase diagrams (Fig. 1). In the low-field region of the Ye phase diagram (Fig. 1b), there are no indications of the reasons that can lead to the appearance of anomalies in the velocity at $E \approx 1$ kV/cm. On the other hand, a boundary exists between the paraelectric and glass-like phases in the low-field region of the Colla phase diagram (Fig. 1a). This boundary between phases is manifested in MBS as a jump in the velocity of the hypersonic LA phonon (Fig. 2). The velocity anomaly in the vicinity of 10 kV/cm corresponds to the phase transition to the induced ferroelectric phase with a reversible polarization. Note that the reverse run of the curve for the velocity of LA phonons in this region of the E–T phase diagram is reversible; that is, with regard to the experimental error and the hysteresis, we return, in fact, to the initial point as the applied field decreases.

At $T = 225$ K (Fig. 2c), the velocity remains unchanged as the electric field varies up to 2 kV/cm. Starting with 2 kV/cm, an increase in the field leads to a smooth step in the velocity, which terminates at $E = 4.5$ kV/cm. A further increase in the applied field (up to 12 kV/cm) does not affect the velocity of acoustic phonon. Velocity anomalies similar to the case observed are called “steplike.” This anomaly is reproduced well in the reverse run of the applied electric field. The anomaly observed in the field dependence at $T = 225$ K (Fig. 2c) corresponds to the structural phase

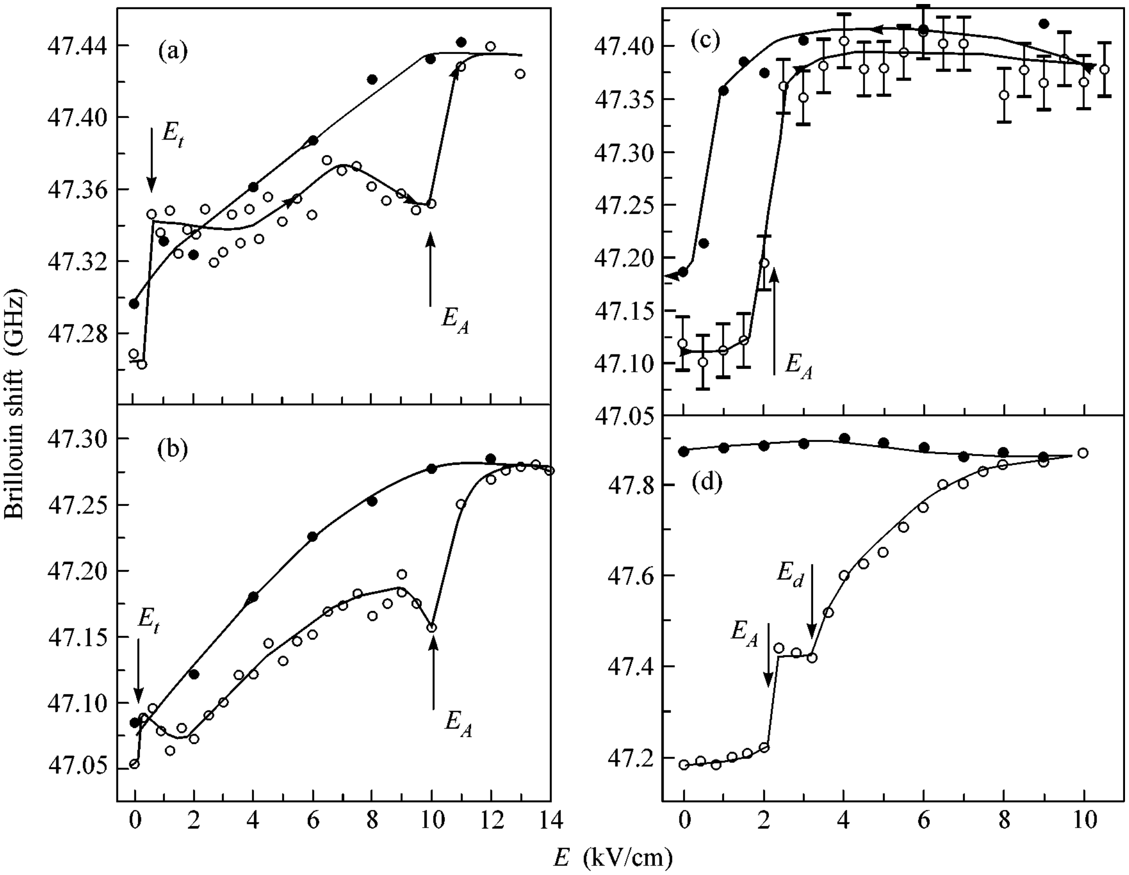


Fig. 2. Dependence of the frequency shift on the applied electric field at various temperatures: T = (a) 255, (b) 240, (c) 225, and (d) 205 K. White and black circles reflect the increase and decrease in the electric-field magnitude, respectively.

transition in the *E–T* phase diagram. Note that the Colla and Ye versions of the *E–T* phase diagram converge in the region of temperatures under consideration ($T_{c0} < T < T_{\text{tricrit}}$).

A situation considerably more eventful arises at T = 205 K (Fig. 2d). Four sections can be distinguished in the field dependence of the velocity: from 0 to 2 kV/cm and from 7 to 10 kV/cm, the velocity is independent of the magnitude of applied field; at $E_{\text{th}} \approx 2.2$ kV/cm, there is a jump in the velocity; and, starting with $E_d \approx$ 3.8 kV/cm, the velocity smoothly increases, reaching a maximum value at E = 7 kV/cm. The fact that a reverse decrease in the field magnitude does not lead to a change in the velocity deserves attention. It is conceivable that the velocity will decrease to the initial state upon changing the polarity of the applied electric field, as was observed in the case with PZN-PT [13]. Compare the dependence of the velocity of acoustic phonons at T = 205 K considered above with the *E–T* phase diagrams under consideration. It is clear that the anomaly in the velocity at E_{th} = 2.2 kV/cm is consistent with the corresponding phase transformations both from phase 1 to induced ferroelectric phase 2 in the Ye *E–T* diagram (Fig. 1b) and from the glasslike phase to the same induced ferroelectric phase in the Colla *E–T* diagram (Fig. 1a). A more complicated situation with the anomaly in the velocity arises in the case of hypersound in the vicinity of E_d = 3.8 kV. In the diagrams proposed in the literature, there is no evidence of the possible reasons for such a change in the velocity. Only recently has evidence been published that a rearrangement of the nanodomain structure occurs in the region of the *E–T* phase diagram under discussion [14]. Therefore, the increase in the velocity of LA phonons upon an increase in the field magnitude, starting at 3.8 kV/cm, can be associated with the domain dynamics. As a result of the nanodomain rearrangement, a transition from the induced ferroelectric phase with a reversible polarization (*N*) to the induced ferroelectric phase with an irreversible polarization (*D*) is accomplished. The fact that the reverse run of the field dependence of velocity at T = 205 K does not depend on the applied field corroborates this suggestion. Such a phase transformation is absent in the *E–T* phase diagram proposed by Colla but is discussed in [10].

Thus, we propose a modified version of the PMN *E–T* phase diagram that takes into account the specific features in the behavior of the velocity of hypersonic acoustic phonons (Fig. 3). Phase boundary B (Fig. 1b), which is absent in Fig. 1a, is revealed in our investiga-

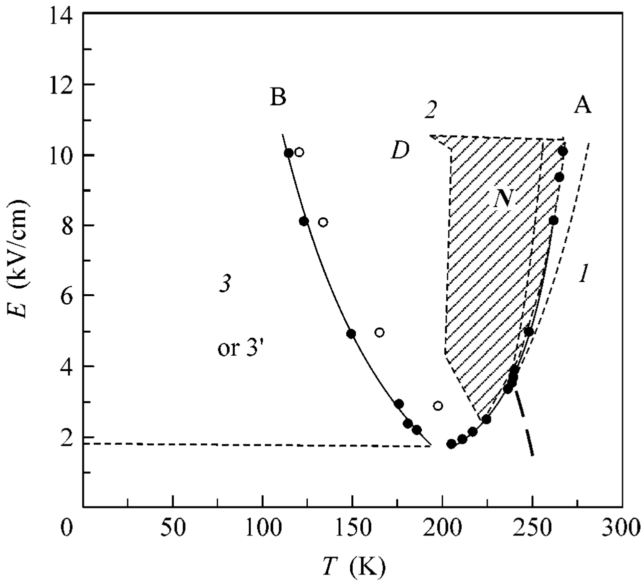


Fig. 3. Modified PMN E–T phase diagram constructed using the results of this work.

tions (Fig. 3), as well as line A. It is necessary to study the PMN crystal in the low-field region and in the temperature range from 230 to 260 K. It is conceivable that a phase boundary between the paraelectric and glassy phases, which is shown in the Colla diagram [9] and revealed in our measurements, exists in this region. Thus, the field investigations of the velocity of longitudinal acoustic phonons allowed us to propose certain refinements of the available E–T phase diagrams of PMN crystals. Clearly, Mandelstam–Brillouin light scattering cannot be considered to be a perfect technique for studying the E–T phase diagrams; however, the results of our work allow us to propose a corrected version of the PMN phase diagram and to attract the attention of experts to the observed contradictions in the experimental data.

This work was supported by the Russian Foundation for Basic Research (project no. 02-02-17678), RF President Grant no. SS-2168.2003.2, and a program of the Division of Physical Sciences of the Russian Academy of Sciences.

REFERENCES

1. G. A. Smolenskiĭ, V. A. Bokov, V. A. Isupov, N. N. Krainik, R. E. Pasynkov, and M. S. Shur, *Ferroelectrics and Related Materials* (Nauka, Leningrad, 1971; Gordon and Breach, New York, 1984).
2. L. E. Cross, Ferroelectrics **76**, 241 (1987).
3. N. De Mathan, E. Husson, G. Calvarin, *et al.*, J. Phys.: Condens. Matter **3**, 8159 (1991).
4. I. G. Siny, S. G. Lushnikov, R. S. Katiyar, and V. H. Schmidt, Ferroelectrics **226**, 191 (1999).
5. H. Fu and R. E. Cohen, Nature **403**, 281 (2000).
6. B. E. Vugmeister and H. Rabitz, Phys. Rev. B **65**, 024111 (2002).
7. S. N. Gvasaliya, S. G. Lushnikov, and B. Roessli, Phys. Rev. B **69**, 092105 (2004).
8. R. Sommer, N. K. Yushin, and J. J. van der Klink, Phys. Rev. B **48**, 13230 (1993).
9. E. V. Colla, E. Y. Koroleva, N. M. Okuneva, and S. B. Vakhrushev, Phys. Rev. Lett. **74**, 1681 (1995).
10. Z.-G. Ye, Key Eng. Mater. **155–156**, 81 (1998).
11. K. Fujishiro, Y. Uesu, Y. Yamada, *et al.*, J. Korean Phys. Soc. **32**, S964 (1998).
12. S. G. Lushnikov, I. G. Siny, and A. K. Tagantsev, in *Abstracts of the Eighth International Meeting on Ferroelectricity* (Maryland, USA, 1993), p. 400; S. G. Lushnikov and I. G. Siny, Kristallografiya **39**, 745 (1994) [Crystallogr. Rep. **39**, 675 (1994)].
13. Jae-Hyeon Ko and Seiji Kojima, Appl. Phys. Lett. **81**, 1077 (2002).
14. R. Blinc, V. Laguta, and B. Zalar, Phys. Rev. Lett. **91**, 247601 (2003).

Translated by A. Bagatur'yants

JETP Letters, Vol. 79, No. 11, 2004, pp. 559–570. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 691–704.
Original Russian Text Copyright © 2004 by Molotkov.

ON THE RESULTS OF PROJECTS OF THE RUSSIAN FOUNDATION FOR BASIC RESEARCH,
project no. 02-02-16289

Integration of Quantum Cryptography into Fiber-Optic Telecommunication Systems

S. N. Molotkov

Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow region, 142432 Russia
Faculty of Computational Mathematics and Cybernetics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
e-mail: molotkov@issp.ac.ru

Received April 6, 2004

The prototype of a new quantum cryptosystem is described. Such a cryptosystem is naturally called time-shift quantum cryptography. It realizes all basic quantum-cryptography protocols [BB84, B92, BB84(4 + 2)] in a common fiber-optic system. This scheme does not involve Mach–Zehnder interferometers, which enables one to naturally realize the multiplex mode of secure key distribution and to naturally integrate this quantum-cryptography scheme into traditional fiber-optic telecommunication systems. The proposed time coding method in quantum cryptography makes it possible to significantly simplify experimental schemes and to exclude the finest fiber-optic part, the interferometer. In essence, the fundamental difference of the time coding method from the phase coding method is that the part using phase relations in superposition between "parts" of a quantum state is eliminated from the phase-coding method, and only the part using the time division principle is retained. The time division principle is common for both methods and is minimally necessary, in contrast to the phase-coding method, which can be excluded altogether. The proposed scheme is briefly compared with the two most developed phase-coding schemes (without self-compensation and with passive self-compensation). © *2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 03.67.Dd; 42.79.Sz; 42.81.Uv; 89.70.+c

1. INTRODUCTION

Development of quantum technologies and fiber-optic communication lines resulted in the appearance of quantum cryptosystems. These systems are the limiting case of absolutely secure (protected) fiber-optic communication lines. The use of fundamental quantum-mechanical laws to protect information makes it possible to obtain results unachievable by both engineering methods of protecting fiber-optic communication lines and traditional methods of mathematical cryptography.

Cryptosystems with one-time secure keys provide the possibility of creating absolutely secure encryption systems that cannot be broken even theoretically [1–3]. In quantum cryptography, the secure key distribution between remote legitimate users is ensured by fundamental laws of nature rather than by the limited computational or technical capabilities of an eavesdropper [4]. The unconditional security of quantum cryptography in the nonrelativistic region is based, in essence, only on the Heisenberg uncertainty principle or, more formally, on the impossibility of simultaneously measuring observables described by noncommuting operators. In terms of a pair of state vectors in which classical information about the key is encoded, this means that it is impossible to gain any information about the transmitted quantum states without distorting them, provided that the latter are nonorthogonal [5]. Another fundamental quantum-mechanical exclusion closely connected to the preceding exclusion is the impossibility of copying an unknown quantum state (no-cloning theorem) [6].

Quantum cryptography and realization of various quantum cryptosystems are being developed in all developed states and in almost all leading telecommunication companies. For the past five years, quantum cryptography investigations have passed from pure fundamental studies to practical realizations and first commercial prototypes.

The available prototypes of quantum cryptosystems are mainly based on the following principles: (i) information about the key is encoded in polarization degrees of freedom [7, 8]; (ii) phase coding, for which a Mach–Zehnder interferometer is used and information is encoded in the phase shift accumulated at the receiving and transmitting arms of the interferometer [9]; (iii) quantum cryptosystems with carrier-frequency modulation [10]; and (iv) coherent-state quantum cryptography with the homodyne detection at the receiver end [11]. Certain hopes for high-rate quantum cryptography are associated with these systems. However, their cryptographic security is yet insufficiently studied, in contrast to systems (i) and (ii). The greatest progress has been achieved in the phase-coding cryptosystems with self-compensation [12–17] using Faraday reflectors. The first laboratory prototype of the quantum cryptosystem with a 1-m quantum communication channel was created at the IBM Research Center in 1989. A laboratory cryptosystem variant that was based on the time division Mach–Zehnder interferometer and used a 30-km fiber-optic communication line was realized at the British Telecom Research Laboratory in 1995 [18], as well as at the Los Alamos Laboratory, where the total length of fiber-optic lines was equal to 48 km [9]. The phase-coding principle was used in those schemes. A distance of 100 km was achieved at the NEC laboratory in 2003 [19], and a distance of 150 km was achieved in 2004 [20]. These schemes complicate and develop the ideas of phase coding with self-compensation by means of Faraday reflectors. It is difficult to realize the above cryptosystems, particularly schemes with phase coding and self-compensation. Theoretical development by the Geneva University group resulted in the practical realization of a quantum cryptosystem with the 23-km fiber-optic cable lying on the bottom of Lake Geneva between Neon and Geneva. At present, the line reaches a length of 67 km and is a complex fiber-optic interferometer with phase coding and self-compensation by means of Faraday reflectors [16] (the first so-called Plug&Play quantum cryptosystem). Investigations are actively being conducted at the IBM Laboratory (Almaden) [21, 22]. The first local quantum cryptographic network was tested in Boston for the secure key distribution between users over a distance of 10 km (this project had been ordered by the Defense Advanced Research Projects Agency) [23].

The MagiQ innovation firm recently announced the first commercial variant of a fiber-optic quantum cryptosystem that acts over a distance of 120 km and is based on the phase-coding principle. This scheme realizes the BB84 quantum cryptography protocol.

Specialists from QinetiQ and Toshiba Research Europe, United Kingdom, believe that the wide application of quantum cryptosystems will begin in next three years with government institutions and banks.

There are prototypes of quantum cryptosystems distributing a secure key through open space [24–26]. Among available data [26], the longest distance is equal to 23.4 km in both daytime and nighttime. Quantum cryptosystems are being developed for generating and distributing secure keys between ground objects and low-orbit satellites (up to an altitude of 1000 km) or between ground objects through satellites. The head of the QinetiQ project estimates that experiments on the transfer of cryptographic keys to low-orbit satellites will begin in March next year, and it will be possible to send secure keys to any point of the Earth after about seven years.

In the near future, the following parameters of quantum cryptography fiber-optic communication lines are expected:

(i) the effective information transfer rate through a fiber-optic quantum channel with an error of no more than several percent,

(ii) length of ~100–150 km of the quantum fiber-optic communication channel,

(iii) 8–16 subchannels upon division in wavelengths (multiplication).

Despite impressing advance in both understanding the cryptographic security of quantum cryptosystems and their realization, these systems contain rather complicated fiber-optic, electronic, and software components. Current work with these components is similar to the conduction of a sophisticated scientific experiment and demonstration of experimental skills rather than to work with widely used and standard scientific equipment. The next important point that as yet restricts wide dissemination of phase-coding quantum cryptosystems is as follows. These systems are as yet badly integrated into standard fiber-optic telecommunication technologies, because quantum cryptosystems contain specific components (interferometers) requiring fine adjustment. Finally, the last fundamental point is that each quantum cryptographic protocol of secure key distribution requires the creation of specific transmitting and measuring equipment. There are three basic protocols of secure key distribution: BB84 [5], B92 [5], and BB84(4 + 2) [27]. The BB84 protocol uses four quantum states: two orthogonal states for 0 and 1 in one basis and two orthogonal states for bits 0 and 1 in the other basis. The states belonging to different bases are pairwise orthogonal, which is necessary for security. The B92 protocol uses a pair of any nonorthogonal quantum states corresponding to 0 and 1. In the BB84(4 + 2) protocol derivative of the BB84 protocol, states belonging to the same bases are also nonorthogonal, in contrast to the latter protocol. It is obvious that different exchange protocols require different physical devices for the formation of quantum states at the transmitter end and, correspondingly, different instruments for quantum-mechanical measurements at the receiver end.

The cryptographic security of these protocols was analyzed in detail in [27–34]. In real systems, where a source is not strictly single-photon, avalanche photodetectors are nonideal, and attenuation is present in the fiber-optic communication channel, the above protocols ensure the security of key distribution up to a certain critical length of a fiber-optic communication line [27]. The protocol B92 is minimal in terms of the numbers of used states and measurements. However, it ensures security up to lengths of ~15–20 km [31]. The most studied BB84 protocol uses four quantum states, is more complicated in realization, and remains secure up to lengths of ~50 km [27]. Finally, four pairwise nonorthogonal states are used in the BB84(4 + 2) protocol. This protocol is even more complicated in realization and adjustment of a fiber-optic interferometer. However, it ensures security up to lengths of ~150 km [27].

For different lengths of communication channels, it is convenient to use different key-distribution protocols. In this case, one has to use substantially different transmitting and receiving physical devices. In the ideal variant, it is desirable to use universal physical (hardware) transmitting and receiving blocks for different lengths of communication channels, where different exchange protocols are realized by only a slight adjustment of software.

It is impossible to achieve the mentioned universality by realizing various protocols with the phase-coding principle. Therefore, it is necessary to develop new physical coding principles.

The proposed time-coding method is universal and, as will be shown below, provides a solution to the mentioned problems. Moreover, this method does not involve fiber-optic interferometers requiring fine adjustment and periodic alignment during the secure key distribution. This scheme is naturally integrated into standard fiber-optic telecommunication systems and allows the realization of all basic protocols of key distribution with the same transmitting and receiving hardware equipment. In addition, in contrast to other encryption methods, this method is most simply and naturally generalized to the multiplex regime of key distribution over a single fiber-optic communication channel.

2. BRIEF DESCRIPTION OF BASIC PROTOCOLS BB84, B92, AND BB84(4+2) OF SECURE KEY DISTRIBUTION

In this section, only minimally necessary description of the basic quantum cryptography protocols is given. Since we focus on the time-coding method and its realization in fiber-optic quantum cryptography schemes, we will not present detailed mathematical proofs of the security of various quantum cryptography protocols (for details, see [27–34]).

BB84 protocol. This protocol was historically the first protocol of quantum key distribution [5]. This protocol is used in the majority of experimental realizations and involves four quantum states associated with classical bits 0 and 1. A pair of quantum states corresponds to 0 ($|0(+)\rangle$) and 1 ($|1(+)\rangle$) and belongs to one basis conventionally denoted as (+). States within this basis are orthogonal ($\langle 0(+)|1(+)\rangle = 0$). The second pair of states belongs to the basis denoted (×) and also corresponds to 0 and 1. Within the second basis, the states $|0(\times)\rangle$ and $|1(\times)\rangle$ are orthogonal ($\langle 0(\times)|1(\times)\rangle = 0$). States belonging to different bases (+) and (×) are pairwise nonorthogonal. Nonorthogonality is fundamentally necessary for detecting eavesdropping attempts. The bases are turned with respect to each other by 45° ($\langle 1(+)|1, 0(\times)\rangle = 1/\sqrt{2}$, $\langle 0(+)|0(\times)\rangle = 1/\sqrt{2}$, $\langle 0(+)|1(\times)\rangle = -1/\sqrt{2}$). The physical nature of quantum states is unimportant. It is convenient to represent these states graphically (Fig. 1a).

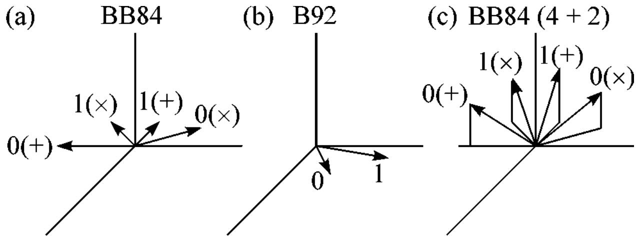


Fig. 1.

In the first realizations of the BB84 protocol, bits 0 and 1 were encoded to the photon polarization states. However, the use of polarization states was not further developed in fiber-optic systems, because an optical fiber badly holds polarization. For large distances, this is unacceptable. All recent distance records were achieved in phase-coding schemes (see below for details).

The key distribution protocol is organized as follows. One of the bases is randomly chosen at the transmitter end. Then, one of the states for 0 or 1 is randomly chosen in this basis. At the receiver end, quantum-mechanical measurements are carried out in a basis that is randomly chosen independently of the transmitter end. After quite a long series of measurements, through classical public communication channel, it is opened which basis has been used in each message, but it is not reported which states have been sent in each basis. Then, users retain only messages in which the chosen bases coincide. The number of such messages is equal to about half of the original sequence. Since states belonging to the same basis are orthogonal and, correspondingly, are certainly distinguishable (with unit probability), legitimate users have identical and random sequences of bits in the absence of external intrusion. The post selection of bases is fundamentally important for the security of the protocol.

States belonging to different bases are certainly undistinguishable, because they are pairwise nonorthogonal. Therefore, if an eavesdropper in the quantum channel does not know in which basis states are chosen, he cannot certainly distinguish transmitted states, which inevitably leads to an increase in error stream at the receiver end.

For that part of the bit sequence retained by legitimate users in which their bases coincide, bits are randomly opened through the classical public channel and are compared. Then, these open bits are rejected. In an ideal (without noise) quantum communication channel, discrepancy in one opened bit is sufficient for detecting the eavesdropper. In real systems, it is impossible to distinguish between noise-induced and eavesdropper induced errors. It is proved that, if errors are no more than ~11% [28, 30], after correction of errors through the open channel and privacy amplification, legitimate users can extract the secure key from the unopened sequence, and this key is identical for them and unknown for the eavesdropper (more precisely, with unit probability, he has exponentially low information depending on the key length).

In real experiments, the state is rather worse. In addition to noise (decoherence processes in the quantum communication channel, which lead, e.g., to the failure in polarization), not all states reach the receiver end due to the presence of attenuation. Moreover, a source of states is not strictly single-photon (although experimental advancement to single-photon sources was reported in [35]), and the efficiency of avalanche photodetectors differs from unity. Due to these circumstances, the protocol guarantees security only to a certain critical length. For real attenuation parameters in optical fiber (0.2 dB/km), efficiency of photodetectors, and mean photon number in weakened laser radiation (coherent state) $\mu \approx 0.1$, the critical length for the BB84 protocol is equal to about 50 km [27].

Since laser radiation (coherent state) has Poisson statistics in the photon number, the number of photons in the channel is equal to $n = 1\ldots\infty$ with the probability $\mu^n e^{-\mu}$. Nondestructive measurements of the number of photons in each particular message through the channel formally exist. If there is one photon, the eavesdropper blocks the channel. If $n > 1$, the eavesdropper separates some photons by means of a beam splitter for his measurements and sends the remaining photons through his quantum channel with lower attenuation. In this case, he holds separated photons in quantum memory and does not carry out measurements until the legitimate users open which basis is used in each message. After that, the eavesdropper carries out measurements in the true basis and certainly gets to know the transmitted states. This eavesdropping procedure is possible if attenuation (correspondingly, the length of the original quantum channel) is larger than a certain critical value. In this case, the eavesdropper has complete information about the distributed key and remains undetected. Though nobody can technically realize this procedure at present, this attack to the key must be taken into account, because unconditional security must be ensured by fundamental quantum mechanical exclusions rather than by technical restrictions. In this case, the security of the key is guaranteed to lengths of ~50 km (for details, see [27]). For a strictly single-photon source, this eavesdropping strategy fails.

B92 protocol. This protocol uses any pair of nonorthogonal quantum states ($|\varphi_0\rangle$ and $|\varphi_1\rangle$, $\langle\varphi_0|\varphi_1\rangle \neq 0$) corresponding to bits 0 and 1. It is convenient to represent these states graphically (Fig. 1b). These states need not be single-photon. For example, coherent states can be used. However, in this case, measurements must distinguish (though with probability less than unity) such a state as a whole; i.e., measurements must realize projections (or positive operator-valued measures) on these multiphoton states. One such procedure is homodyne detection, whose realization is rather complicated and not discussed here. Hereinafter, we treat states as single-photon states. In this case, measurements are carried out by single-photon photodetectors in the photon counting mode.

The protocol is organized as follows. One of the states is randomly chosen at the transmitter end. At the receiver end, the legitimate user carries out measurement described by the unity decomposition

$$\mathcal{P}_0^{\perp} + \mathcal{P}_1^{\perp} + \mathcal{P}_? = I, \quad \mathcal{P}_0^{\perp} = a(I - |\varphi_0\rangle\langle\varphi_0|),$$

$$\mathcal{P}_1^{\perp} = a(I - |\varphi_1\rangle\langle\varphi_1|), \quad \mathcal{P}_? = I - \mathcal{P}_0^{\perp} - \mathcal{P}_1^{\perp}, \qquad (1)$$

$$a = \frac{1}{1+\cos\zeta}, \quad \cos\zeta = \langle\varphi_1|\mathcal{P}_?|\varphi_1\rangle = \langle\varphi_0|\mathcal{P}_?|\varphi_0\rangle.$$

The space of different outcomes consists of three elements {0, 1, ?}, which arise randomly. Measurement is such that, if the state $|\varphi_0\rangle$ is sent, the outcome may be {1} or {?} and never {0}, because the outcome {0} (operator-valued measure $\mathcal{P}_0^{\perp}$) on the input state $|\varphi_0\rangle$ has zero probability. The situation is similar if the state $|\varphi_1\rangle$ is transmitted. Then, through the classical public channel, the user at the receiver end reports the numbers of messages with definite outcomes {0} or {1} (outcomes themselves are not reported). Messages with outcomes {?} are rejected. Outcomes {?} can arise both from bits 0 and 1. In the retaining sequence of outcomes {0} and {1}, legitimate users know that the state $|\varphi_0\rangle$ provides outcome {1} and vise versa. As a result, in the absence of noise in the channel and eavesdropper attack, the legitimate users have the same bit sequence. Noise in the channel, as well as the presence of the eavesdropper, increases the error stream. The study of the protocol security shows that, in the presence of attenuation and nonideal photodetectors, the protocol guarantees key security for a fiber-optic communication line with a length of about 20 km (for details, see [31]).

It is possible to qualitatively understand why the protocol cannot guarantee secure key distribution when attenuation (communication channel length) exceeds a certain critical value by the following reasoning (for details, see, e.g., [27]). In the presence of attenuation, the eavesdropper can use measurements (1) the same as a legitimate user. In this case, if outcome {0} or {1} (conclusive result) is obtained, the eavesdropper resends states for 0 or 1 in dependence of the outcome through its channel with lower attenuation to the receiver end. If inconclusive result {?} is obtained, the eavesdropper sends nothing to the receiver end. If attenuation is sufficiently large in the original channel, this eavesdropping attempt is not detected (disappearance of states in the channel is attributed to attenuation) and the eavesdropper has complete information about the distributed key and remains undetected.

BB84(4 + 2) protocol. To a certain extent, this protocol is intermediate between the two above quantum secure key distribution protocols [27]. As the BB84 protocol, this protocol uses four quantum states associated with bits 0 and 1 in two bases. The only difference is that nonorthogonal states are taken in each basis. States belonging to different bases remain pairwise nonorthogonal. It is convenient to represent these states graphically in terms of vectors (Fig. 1c). The protocol is organized as follows. The legitimate user at the transmitter end randomly chooses one of the bases. Then, one of states 0 or 1 is randomly chosen in this basis and is sent through the quantum communication channel. The second legitimate user at the receiver end independently chooses measurements of two types (in different bases[1]). Different types of measurements are described by the different unity decompositions. In basis (+), this decomposition has the form

$$\begin{gathered}
\mathcal{P}^{\perp}_{0(+)} + \mathcal{P}^{\perp}_{1(+)} + \mathcal{P}_{?(+)} = I, \\
\mathcal{P}^{\perp}_{0(+)} = a(+)(I - |0(+)\rangle\langle 0(+)|), \\
\mathcal{P}^{\perp}_{1(+)} = a(+)(I - |1(+)\rangle\langle 1(+)|), \\
\mathcal{P}_{?(+)} = I - \mathcal{P}^{\perp}_{0(+)} - \mathcal{P}^{\perp}_{1(+)}, \\
a(+) = \frac{1}{1 + \cos\zeta(+)}, \\
\cos\zeta(+) = \langle 1(+)|\mathcal{P}_{?(+)}|1(+)\rangle = \langle 0(+)|\mathcal{P}_{?(+)}|0(+)\rangle.
\end{gathered} \tag{2}$$

In basis (×), it is

$$\begin{gathered}
\mathcal{P}^{\perp}_{0(\times)} + \mathcal{P}^{\perp}_{1(\times)} + \mathcal{P}_{?(\times)} = I, \\
\mathcal{P}^{\perp}_{0(\times)} = a(\times)(I - |0(\times)\rangle\langle 0(\times)|), \\
\mathcal{P}^{\perp}_{1(\times)} = a(\times)(I - |1(\times)\rangle\langle 1(\times)|), \\
\mathcal{P}_{?(\times)} = I - \mathcal{P}^{\perp}_{0(\times)} - \mathcal{P}^{\perp}_{1(\times)}, \\
a(\times) = \frac{1}{1 + \cos\zeta(\times)}, \\
\cos\zeta(\times) = \langle 1(\times)|\mathcal{P}_{?(\times)}|1(\times)\rangle = \langle 0(\times)|\mathcal{P}_{?(\times)}|0(\times)\rangle.
\end{gathered} \tag{3}$$

Here, all states are pairwise nonorthogonal. Other measurements are also possible, but measurements described by Eqs. (2) and (3) are simple in experimental realization.

Then, after the transmission of quite a long sequence, through the classical public communication channel, it is opened which basis has been used in each message. Messages for which bases do not coincide are rejected. For retaining messages, the user at the receiver end publicly opens the numbers of messages for which he obtains inconclusive outcomes ($\mathcal{P}^{\perp}_{?(+,\times)}$). These messages are also rejected. Among the remaining messages, for which conclusive outcomes were obtained, the secure key is extracted by correction of errors through the open channel and privacy amplification. This protocol is less studied than the two protocols discussed above. However, the available analysis of its security shows that it is most firm in terms of distance and remains secure for quantum fiber-optic communication channels with lengths up to about 150 km [27].

3. UNIVERSAL REALIZATION OF THE BASIC QUANTUM-CRYPTOGRAPHY PROTOCOLS IN THE FIBER-OPTIC SCHEME WITHOUT INTERFEROMETERS (TIME-CODING METHOD)

Nonorthogonal states are necessary for security and the detection of eavesdropping attempts. The time-coding concept also uses the property of nonorthogonal states to be reliably undistinguishable. For the encryption of bits 0 and 1, only a state with one space–time form (wave function) is used, but it is shifted by different time intervals in each message, which ensures nonorthogonality. This idea makes it possible to substantially simplify the fiber-optic part of the quantum cryptography scheme and completely remove the Mach–Zehnder interferometer, which is the finest tunable part of phase-coding systems. The absence of the interferometer makes it possible to realize all basic quantum cryptographic protocols in a common fiber-optic system without its change, adjustment, and change of controlling electronics for a specific exchange protocol. Moreover, in the framework of available fiber-optic technologies, this scheme is naturally generalized to the multichannel (multiplex) case, where several secure keys are transmitted though a single fiber-optic communication channel.

We first describe time diagrams of various protocols and then their realization.

BB84 protocol. The basic idea of the initial BB84 protocol is the use of two bases. States 0 and 1 in each basis are orthogonal. States from different bases are pairwise nonorthogonal.

States at the receiver end. The same state shifted with respect to the clock pulse in each message by a certain value is used as a single-photon state (Fig. 2). The clock pulse is a short (duration of about 1 ns) multiphoton (classical) state with a carrier wavelength of 1310 nm. In Fig. 2, the clock pulse is shown by the dark background. Two bases—{+(1), ×(1)} and {+(2), ×(2)}—are used. In each subbasis {+(1) and ×(1)} within the first basis (index 1), states $0 - |0_1(+)\rangle$ and $1 - |1_1(+)\rangle$ for 0 and 1, respectively, in subbasis $|0_1(\times)\rangle$ and $1 - |1_1(\times)\rangle$ for 1 are orthogonal. States belonging to different subbases {+(1) and ×(1)} are pairwise nonor-

---

[1] Since the states are nonorthogonal, the term “basis” is likely not very good, but it does not lead to ambiguity in this context.

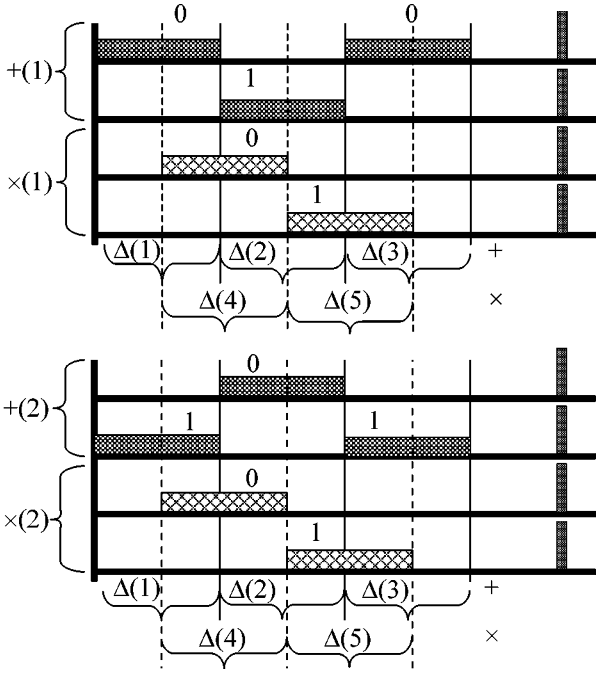


Fig. 2.

thogonal, which is achieved by corresponding time shifts.

The properties of basis 2 ({+(2), ×(2)}) are similar. States in each basis and subbases are shown in Fig. 2.

In comparison with the original BB84 protocol, the introduction of two bases with two subbases in each of them is necessary for rejecting the interferometer. In essence, this realization is equivalent to the original BB84 protocol.

The protocol is organized as follows. One of the two bases {+(1), ×(1)} or {+(2), ×(2)} is randomly chosen. Within the chosen basis, one of the subbases + or × is randomly chosen. Then, one of states 0 or 1 is randomly chosen in this subbasis. For example, as is seen in Fig. 2, if basis 1 and, then, subbasis +(1) are chosen, two variants are possible for choosing 0. One of the possible variants for 0 is randomly chosen.

For the second basis, two variants for 1 exist in the subbasis 2(+), which eliminates asymmetry.

Only one of the states (shown by shelves in Fig. 2) is always sent into the communication channel. The shape of the state is always the same. The difference in shifts with respect to the clock pulse is controlled by software (see below).

Measurement of states at the receiver end. States are detected at the receiver end with respect to the clock pulse in each message in time window Δ(1), …, Δ(5), which is chosen randomly and independently of the choice at the transmitter end (details will be given below).

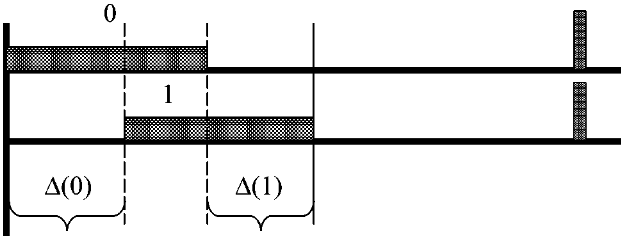


Fig. 3.

After long series of measurements, through the public channel, the user at the receiver end reports the numbers of messages where the photodetector is triggered. Through the public channel, the user at the transmitter end reports which basis, {+(1), ×(1)} or {+(2), ×(2)}, and which subbasis are used for each of these messages but does not report which bit, 0 or 1, was chosen in the subbasis.

Since Alice knows what she has sent, and Bob knows in which time window a count has been detected, after the announcement of the basis and subbasis, Bob knows what bit, 0 or 1, has been sent by Alice (see Fig. 2).

In particular, if Alice has sent bit 0 in the basis +(1) (the left shelf in the upper line in Fig. 2) and Bob detects a count in the window Δ(1) (information about the window is not reported), after the report of basis 1 or 2 and subbasis + or ×, both users know what has been sent.

If Bob carried out measurements in the window, e.g., Δ(4), this message is rejected.

This protocol requires two bits of classical information rather than one bit as in the standard BB84 protocol.

B92 protocol. A state with the same amplitude shape (wave function) is used in this protocol. The states corresponding to bits 0 and 1 are nonorthogonal due to the time shift with respect to the clock pulse in each message.

Alice randomly chooses one of the states and sends it to Bob. Bob randomly and independently of Alice chooses the detection time window, Δ(0) or Δ(1). After long series of measurements, through the classical public channel, Bob reports in which messages he detected triggering. After that, both users know the transmitted bit (see Fig. 3).

BB84(4 + 2) protocol. This protocol is similar to the BB84 protocol. The only difference is that states in different subbases are nonorthogonal (see Fig. 4). The nonorthogonality degree (overlapping) can be controlled. The choice of states is similar to the BB84 protocol. The choice of time windows for measurements at the receiver end is slightly different. Time windows are chosen randomly and independently of the transmitter end. After the measurement series, Alice reports which basis and subbasis have been used in each specific mes-

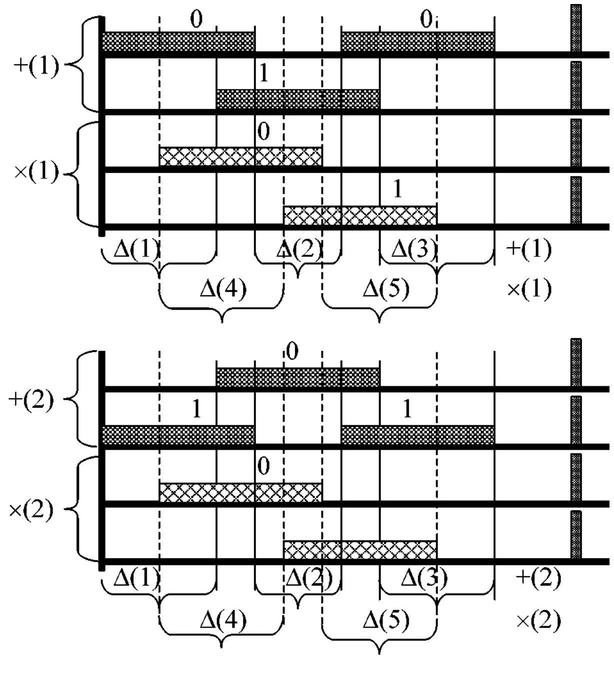


Fig. 4.

sage. Knowing the announced basis and the time window chosen for measurement, Bob reports (through the public channel) the numbers of messages where the choice of the time window correlates with the basis (similarly to the BB84 protocol, see Fig. 4).

In essence, measurement in finite time windows realizes measurements (1)–(3) yielding the distinctive or indistinctive outcomes. Measurements in a finite time window are formally described by the unity decomposition, when a positive operator-valued measure is attributed to each time region (see, e.g., [36]).

Description of scheme operation. The prototype block diagram is shown in Fig. 5. Before a key distribution session, a random number generator generates a random sequence of bits 0 and 1, which is written into temporal memory. Then, the random sequence in memory is used as an internal code for choosing the basis and informative states 0 or 1 in dependence of the used protocol. A pulse generator activates the scheme of the formation of the clock pulse for a 1310-nm laser and the scheme of the formation of informative states (1550-nm CW DFB laser). In our case, the CW DFB laser is a source that is almost always turned on. Upon direct modulation, the application of an additional cur-

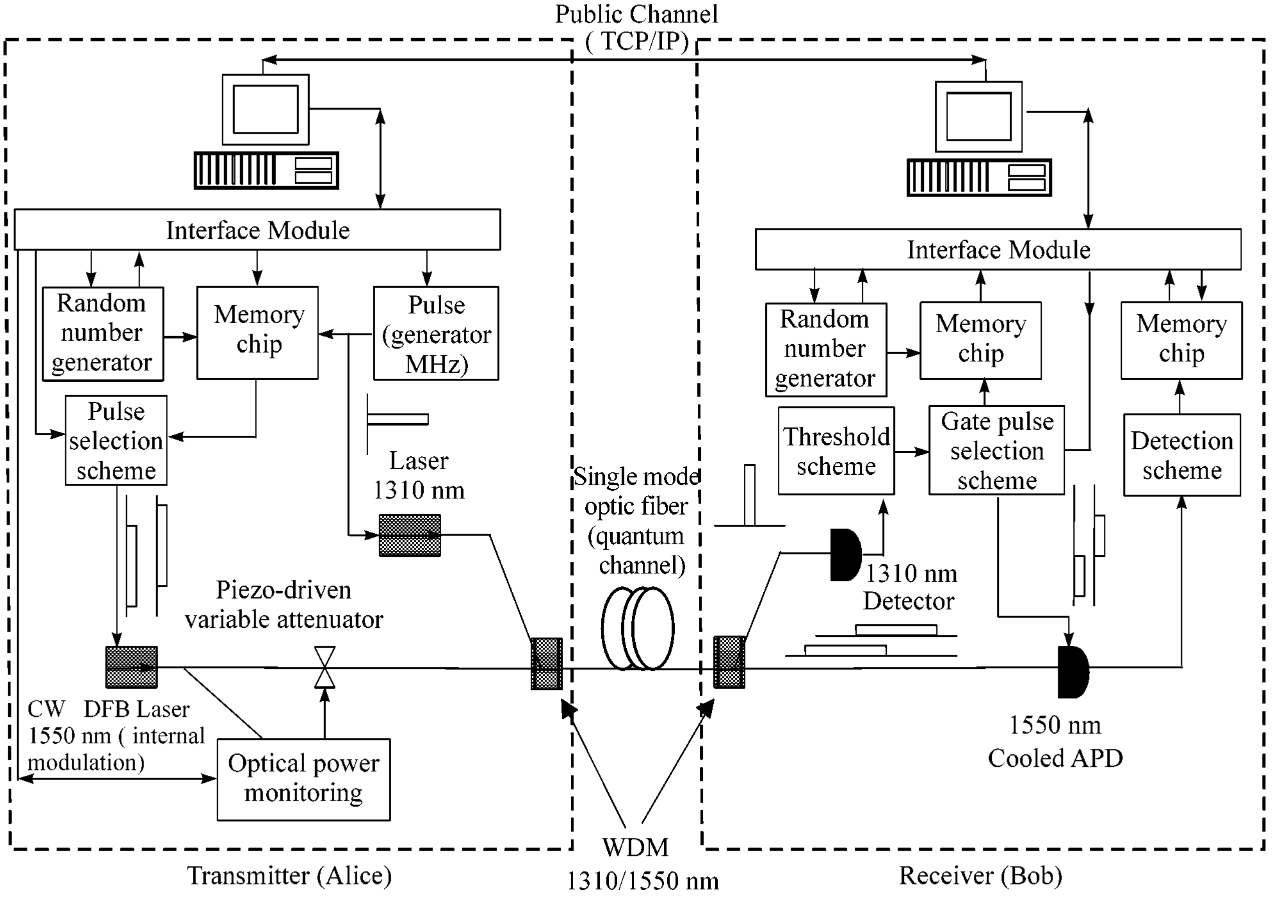


Fig. 5.

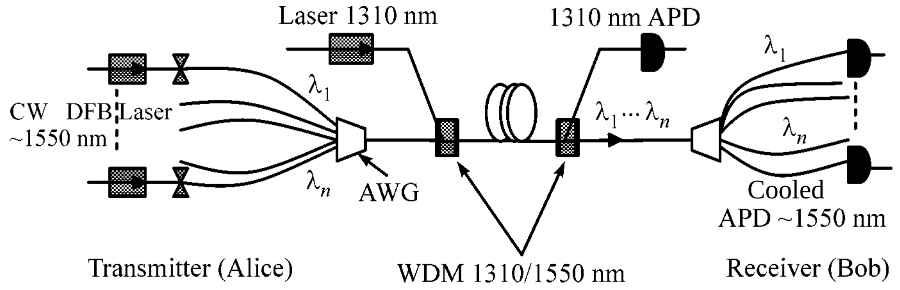


Fig. 6.

rent pulse step to the CW DBF laser induces a signal in the fiber-optic communication line. An external modulator also can be used (fiber-optic shutter). In this case, the laser is turned on and is in the generation mode. In good CW DFB lasers with multiple quantum wells, the line width can reach hundreds of kilohertz. For our goals, an informative state with a duration of, e.g., 20 ns is sufficient (correspondingly, the line width 10^7 Hz of CW DFB laser radiation is sufficient). It is senseless to take too long informative states, because inverse bias must be applied to the avalanche photodetector for this time, which increases the probability of parasite dark counts.

At the receiver end, there is also an independent random number generator that generates a random sequence and writes it to temporary memory. Then, this sequence is used to choose time windows for superposition of pulses of inverse bias on the avalanche photodetector (conditionally activates the avalanche photodetector). The activation of the avalanche photodetector is initiated by the clock pulse from the 1310-nm laser. When the avalanche photodetector is triggered, the detection scheme yields a signal to another temporal memory for the accumulation of the primary key. The secure key is obtained from this primary key after the transmission through the quantum channel, exchange through the classical public channel (according to the TCP/IP protocol with identification), correction of errors, and privacy amplification.

Multiplex case. This scheme allows natural generalization to the multiplex case with the division of channels in wavelengths. Figure 6 shows the block diagram of a multiplex cryptosystem whose operation principle is similar to the preceding scheme. The only change is the introduction of a router (diffraction grating based on a waveguide set) into the scheme. This router is made by the AWG technology, which makes it possible to bring together states with different wavelengths into one fiber-optic cable at the transmitter end and separate them at the receiver end. Figure 6 shows the scheme of the fiber-optic part.

4. BRIEF COMPARISON WITH QUANTUM CRYPTOGRAPHY SCHEMES INVOLVING THE PHASE-CODING METHOD (WITH AND WITHOUT SELF-COMPENSATION)

Let us briefly analyze our scheme in comparison with well-developed schemes based on the phase-coding method.

Phase-coding scheme without self-compensation. This fiber-optic scheme is used in the first local quantum cryptographic network in Boston [23] (this project had been ordered by the Defense Advanced Research Projects Agency).[2]

Figure 7 shows the block diagram of the prototype of the phase-coding quantum-cryptography scheme realizing the BB84 protocol without passive self-compensation. The scheme is an unbalanced Mach–Zehnder interferometer with time division. Interference is absent between the long and short arms inside small arms. Interference is present between arms at the transmitter and receiver ends.

Encryption is organized as follows. Alice chooses one of two bases, + or ×, and the phase delay $(0, 1) \longrightarrow \varphi_A = (0, \pi)$ and $(\pi/2, 3\pi/2)$ is established for sending 0 or 1 in the + and × bases, respectively. Bit 0 or 1 is randomly chosen by introducing the corresponding phase delay. Independently of Alice, Bob randomly chooses one of the bases. The delay φ_B is established in dependence of the chosen basis. Sent bit 0 or 1 is distinguished by triggering detectors (upper or lower). In dependence of varying the difference of phase delays at the transmitter and receiver ends, constructive or destructive interferences occur at the exit; i.e., the probabilities of triggering the upper and lower photodetectors (see Fig. 7) are $\propto \cos^2(\varphi_A - \varphi_B) = \cos^2(2\pi(\Delta l_A - \Delta l_B)/\lambda)$ and $\propto \sin^2(\varphi_A - \varphi_B) = \sin^2(2\pi(\Delta l_A - \Delta l_B)/\lambda)$.

After a long series, Alice opens which basis has been chosen in each message but does not report which bit, 0 or 1, has been chosen in this basis. Messages where bases do not coincide are rejected. In the remaining messages, knowing which detector has been triggered, both users know the transmitted bit. In particular, if the both users take the + basis and Alice sends bit 1

---

[2] Figure 7 is taken from [23] with the kind sanction of the authors.

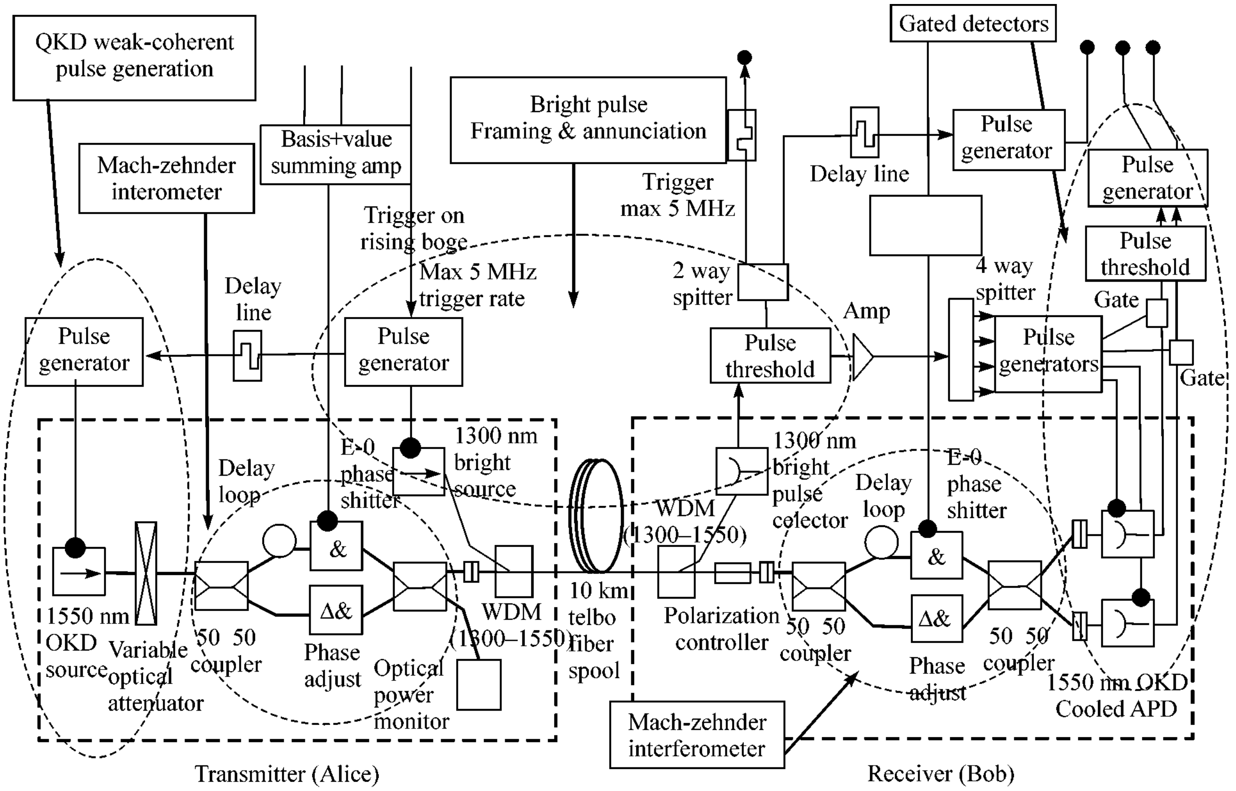


Fig. 7.

($\varphi_A = \pi$) and Bob correspondingly takes the + basis ($\varphi_B = 0$), the probabilities of triggering the upper and lower photodetectors are $\propto\cos^2(\varphi_A - \varphi_B) = \cos^2(\pi) = 1$ and $\propto\sin^2(\varphi_A - \varphi_B) = \sin(\pi) = 0$, respectively. If Alice takes bit 0 ($\varphi_A = 0$) in the same basis, the lower photodetector is triggered with the probability $\propto\sin^2(\varphi_A - \varphi_B) = \sin^2(0) = 1$.

The state at the exit of the transmitting arm of the interferometer is a superposition of two halves shifted in phase by the path difference $\exp ik(\Delta L + \Delta l_A)$ ($k = 2\pi/\lambda$, $\lambda = 1550$ nm). Here, ΔL is the constant path difference due to the difference between the lengths of the upper and lower arms in the small interferometer at the transmitter end and $2\pi\Delta l_A/\lambda = \varphi_A$ is the variable path difference between the upper and lower arms. The path difference ΔL ensures the shift between halves in time superposition (time division principle).

A state at the exit in front of the photodetectors at the receiver end is a superposition of three peaks shifted in time by $\Delta L/c$. The central peak is twice as high as side peaks. The superposition of three peaks arises, because each of state halves coming from the communication channel passes through both the long and short arms of the interferometer at the receiver end. The state is detected by the application of negative bias on avalanche photodetectors at the time of the arrival of the central peak.

For scheme operation, the difference $\Delta l_A - \Delta l_B$ between differences between long and short paths at the transmitter and receiver ends of the interferometer must be no more than fractions of λ, i.e., fractions of a micron. The transmitting and receiving arms of the interferometer are spaced by tens of kilometers. Any deviation from ideal values leads to errors. In particular, bit 0 has been sent, but outcome 1 can be obtained due to the departure of the phase difference and vice versa. For this reason, thermostabilization of arms and periodic adjustment of arms is inevitably required for stable operation. Key distribution is naturally interrupted for these operations. Exact periodic adjustment is carried out by special piezo-driven modulators in the short arms of the interferometer. This adjustment requires additional controlling electronics and software and strongly complicates the entire scheme. Real work with such a device requires experimental skill.

The phase-coding principle uses, in addition to time division, phase relations between the individual "parts" of states shifted in time. The use of phase relations for encryption inevitably leads to the need for interferometer schemes requiring high accuracy.

In the present time-coding scheme, finest and complicated phase relations are excluded, and only time division remains.

It is worth noting that phase-coding schemes cannot be naturally and easily generalized to the multiplex

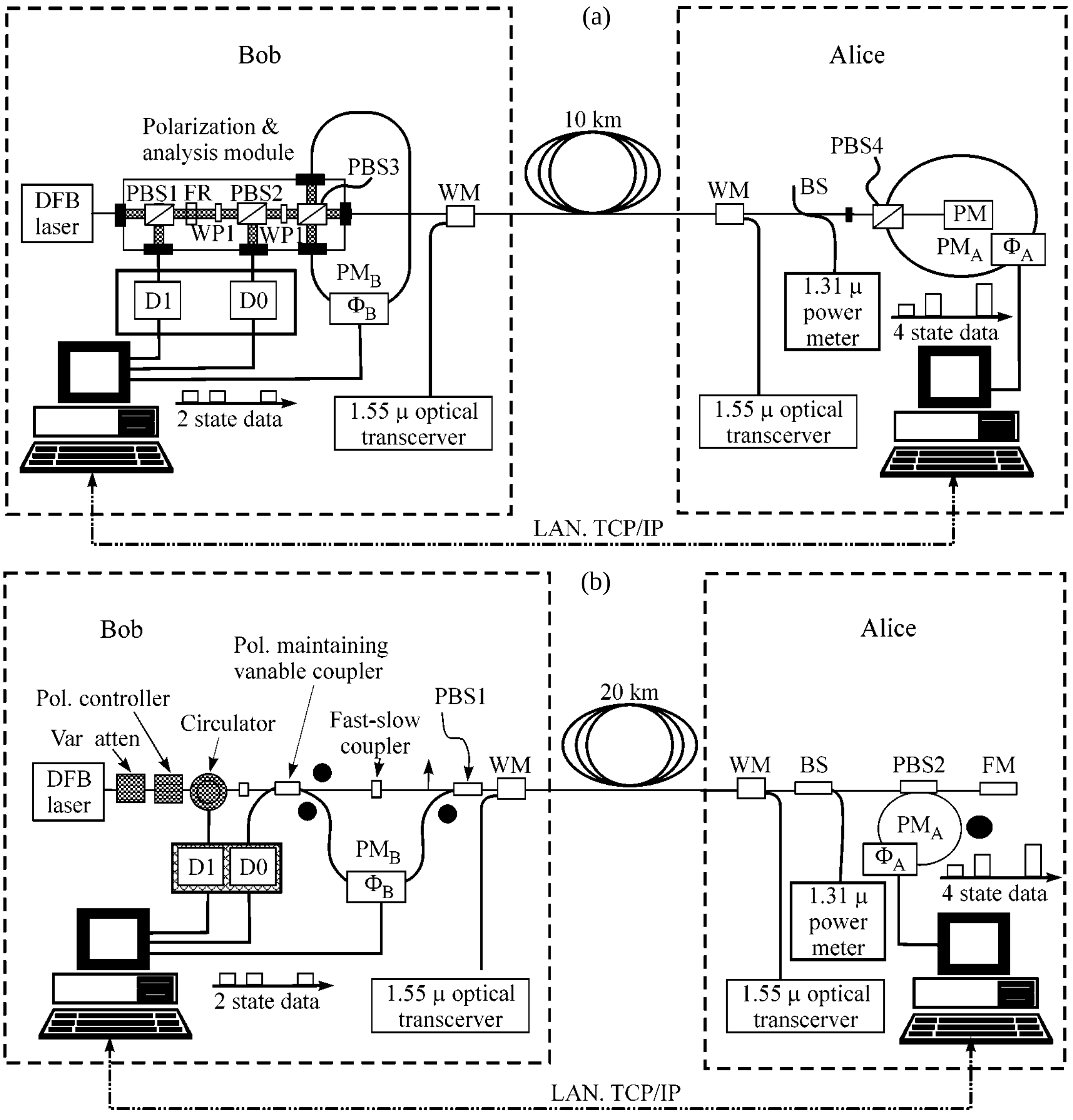


Fig. 8.

case, because the phase differences are different for different wavelengths.

Phase-coding schemes with passive self-compensation. Let us give brief comparative analysis with the quantum cryptography scheme with phase coding and passive self-compensation. This scheme was studied in detail at the IBM Research Laboratory (Almaden, California).[3]

Phase shifts are sensitive to birefringence inevitably present in fiber-optic communication lines. Different polarization components in states gain different phase shifts due to birefringence. Such parasitic phase shifts are eliminated by passive self-compensation, which further complicates the fiber-optic part of the quantum cryptosystem. Moreover, the scheme becomes double-pass, which effectively doubles the length of the fiber-optic communication line. Self-compensation is organized as follows. Two orthogonal polarization components, horizontal and vertical, acquire different phase shifts in the direct pass. Then, the horizontal component switches over to the vertical component and vice versa due to reflection from a Faraday mirror, and the difference between phase shifts acquired by different components is compensated in the inverse pass.

Quantum cryptosystems are intensely being developed by several research groups. The creation and adjustment of such systems also require fine experimental skill. Let us describe the work of this scheme by two examples [21].

Figure 8a shows the layout of the prototype of such a system [21]. Laser radiation forms signal states, which pass through the system of polarization beam

---

[3] Figure 8 is taken from [21] with the kind sanction of the authors.

splitters after weakening. A state with two polarization components (vertical and horizontal) arises at the exit of a PBS3 filter. The horizontal component enters the communication channel, and the vertical component enters the channel with a certain time delay, which is ensured by a fiber-optic loop maintaining the polarization state. Here, the time division of the parts of the quantum state is also used. To the time of arrival at the receiver end (Fig. 8a, Alice), the horizontal component is oriented arbitrarily due to uncontrollable rotation in the optical fiber. The polarization beam splitter (PBS4) separates two orthogonal components. Through the optical fiber, which conserves polarization, these components enter a delay loop, where they propagate oppositely. By means of a phase modulator in the middle of the loop, Alice sets phase delay (similarly to coding in the above scheme without self-compensation). Since the modulator is at the middle of the loop, different polarization components propagating oppositely acquire the same phase shift. Upon passing the modulator, the same additional phase shift corresponding to the choice of 0 or 1arises. Reflecting from the Faraday mirror, orthogonal polarization components mutually change places. Upon passing the loop, both components propagate backward to the transmitter end (Bob). The difference between the phase shifts for orthogonal polarization components, which arises at the direct pass, is passively compensated due to the mutual substitution of the components on the Faraday mirror. Similar transformations occur with the component with initially vertical polarization and shifted (delayed) in time.

To the time of returning the parts of the state (delayed and undelayed) to the transmitter end, Alice sets her phase delay in the lower arm of the interferometer by the modulator as was described for the above scheme. Depending on the phase difference at the Alice and Bob ends, constructive or destructive interference occurs between the undelayed and delayed parts of the state. For coincident bases (choosing phase delays), the detector is triggered in channel 0 or 1 (see Fig. 8a).

Figure 8b shows the improved variant of the preceding scheme with a similar operation principle. The number of polarization beam splitters in the scheme part sensitive to phase changes is decreased due to the overall use of the polarization-maintaining optical fiber.

Even this brief description of the schemes shows that phase-coding systems are very fine optical schemes, and their use and work with them require high experimental skill, in contrast to work with standard physical measuring instruments.

5. CONCLUSIONS

The time-coding method in quantum cryptography makes it possible to significantly simplify experimental schemes and to exclude the finest fiber-optic part, the interferometer. In essence, the fundamental difference of the time coding method from the phase-coding method is that the part using phase relations in superposition between parts of a quantum state is eliminated from the phase-coding method, and only the part using the time division principle is retained. The time division principle is common for both methods and is minimally necessary, in contrast to the phase-coding method, which can be excluded altogether.

Thus, the above prototype schemes for quantum cryptography that involve the time-coding method have a number of the following fundamental advantages over the available phase-coding schemes.

A fiber-optic interferometer and, therefore, fine (to fractions of a micron) balancing of arms and their constant adjustment upon operation are not required. Electronics for controlling optical phase modulators in the interferometer arms is unnecessary. Polarization-sensitive fiber-optic components are not necessary.

These schemes provide the realization of all basic quantum cryptography protocols.

Time-coding schemes can be expanded to the multiplex mode, which enables one to increase the rate of key distribution due to an increase in the number of frequency channels and to use the cryptosystem in local networks, where each receiver has his own wavelength.

These schemes are single-pass and involve the minimum number of fiber-optic components and connectors, which makes it possible to increase the length of the fiber-optic communication channel due to lower losses in the optical fiber.

Finally, the time-coding concept is naturally integrated into the structure of standard fiber-optic telecommunication systems in the structure of transmitted signals (see operation time) and is, in essence, almost classical with only the difference that the signal level is as low as single-photon.

I am grateful to A.N. Klimov, K.N. El'tsov, M.I. Belovolov, S.S. Nazin, A.V. Korol'kov, A.N. Penin, and S.P. Kulik. This work was supported by the Academy of Cryptography of the Russian Federation and the Russian Foundation for Basic Research (project no. 02-02-16289).

REFERENCES

1. G. S. Vernam, J. Am. Inst. Electr. Eng. **55**, 109 (1926).
2. V. A. Kotel'nikov, Report (1941).
3. C. E. Shannon, Bell Syst. Tech. J. **28**, 658 (1949).
4. S. Wiesner, SIGACT News **15**, 78 (1983).
5. C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computer Systems and Signal Processes* (Bangalore, 1984), p. 175; C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
6. W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
7. C. H. Bennett, F. Bessette, G. Brassard, *et al.*, J. Cryptology **5**, 3 (1992).
8. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, quant-ph/0101098; Rev. Mod. Phys. **74**, 145 (2002).

9. R. J. Hughes, G. L. Morgan, and C. G. Peterson, J. Mod. Opt. **47**, 533 (1998).
10. P. C. Sun, Y. Mazurenko, and Y. Fainman, Opt. Lett. **20**, 1062 (1995); Y. Mazurenko, R. Giust, and J. P. Goedgebuer, Opt. Commun. **133**, 87 (1997).
11. F. Grosshans, G. Van Assche, J. Wenger, *et al.*, Nature **421**, 238 (2003).
12. H. Zbinden, H. B. Pasquinucci, N. Gisin, and G. Ribordy, Appl. Phys. B **67**, 743 (1998).
13. A. Muller, T. Herzog, B. Huttner, *et al.*, Appl. Phys. Lett. **70**, 793 (1997).
14. H. Zbinden, J. D. Gautier, N. Gisin, *et al.*, Electron. Lett. **33**, 586 (1998).
15. D. S. Bethune and W. P. Risk, IEEE J. Quantum Electron. **36**, 340 (2000).
16. G. Ribordy, J. D. Gautier, N. Gisin, *et al.*, Electron. Lett. **34**, 2116 (1998); D. Stucki, N. Gisin, O. Guinnard, *et al.*, quant-ph/0203118.
17. M. Bourennane, F. Gibson, A. Karlsson, *et al.*, Opt. Express **4**, 383 (1999).
18. C. Marand and P. D. Townsend, Opt. Lett. **20**, 1695 (1995).
19. H. Kosaka, A. Tomita, Y. Nambu, *et al.*, quant-ph/0306066.
20. T. Kimura, Y. Nambu, T. Hatanaka, *et al.*, Preprint (2004).
21. D. S. Bethune and W. P. Risk, New J. Phys. **4**, 42.1 (2002).
22. D. S. Bethune, M. Navarro, and W. P. Risk, quant-ph/0104089.
23. C. Elliot, D. Pearson, and G. Troxel, quant-ph/0307049.
24. J. G. Rarity, P. R. Tapster, P. M. Gorman, and P. Knight, New J. Phys. **4**, 82.1 (2002).
25. R. J. Huges, J. E. Nordholt, D. Derkas, and C. G. Peterson, quant-ph/0206092.
26. C. Kurtsiefera, P. Zarda, M. Halder, *et al.*, Preprint (2002).
27. A. Acin, N. Gisin, and V. Scarani, quant-ph/0302037.
28. D. Mayers, quant-ph/9802025.
29. E. Biham, M. Boyer, P. O. Boykin, *et al.*, quant-ph/9912053.
30. P. W. Shor and J. Preskill, quant-ph/0003004.
31. K. Tamaki, M. Koashi, and N. Imoto, quant-ph/0212161 (2002).
32. N. Lutkenhaus, Phys. Rev. A **61**, 052304 (2000).
33. G. Brassard, N. Lutkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000).
34. G. Gilbert and M. Hamrick, *Practical Quantum Cryptography: A Comprehensive Analysis*, Mitre Technical Report, MTR00W0000052 (Mitre Corporation, McLean, VA, 2000), Part 1.
35. A. Beveratos, R. Brouri, T. Gacoin, *et al.*, quant-ph/0206136.
36. S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 1156 (2003) [JETP Lett. **78**, 659 (2003)].

Translated by R. Tyapaev

JETP Letters, Vol. 79, No. 11, 2004, pp. 571–581. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 79, No. 11, 2004, pp. 705–716.
Original Russian Text Copyright © 2004 by Kadomtseva, Zvezdin, Popov, Pyatakov, Vorob'ev.

ON THE RESULTS OF PROJECTS OF THE RUSSIAN FOUNDATION FOR BASIC RESEARCH,
project no. 01-02-16468

Space–Time Parity Violation and Magnetoelectric Interactions in Antiferromagnets

A. M. Kadomtseva[1], A. K. Zvezdin[2], Yu. F. Popov[1, *], A. P. Pyatakov[1], and G. P. Vorob'ev[1]

[1] *Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
**e-mail: Popov@plms.phys.msu.ru*
[2] *Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia*

Received April 29, 2004

The properties of antiferromagnetic materials with violated space–time parity are considered. Particular attention is given to the bismuth ferrite $BiFeO_3$ ferroelectric magnet. This material is distinguished from other antiferromagnets in that the inversion center is absent in its crystal and magnetic structures. This circumstance gives rise to diversified and unusual properties, namely, to the appearance of a spatially modulated spin structure and to the unique possibility of the linear magnetoelectric effect coexisting with a weak ferromagnetic moment. The magnetic-induced phase transitions accompanied by the suppression of the modulated spin structure and appearance of a number of new and unusual effects are considered. These are the linear magnetoelectric effect and the appearance of a toroidal moment and a weak ferromagnetic moment of the magnetoelectric nature. *© 2004 MAIK "Nauka/Interperiodica".*

PACS numbers: 75.80.+q; 75.50.Ee

1. INTRODUCTION

The magnetic symmetry method is an elegant and efficient tool for studying the physical properties of crystals with complex magnetic structure, notably, antiferromagnets. On the basis of this approach, the Dzyaloshinski–Moriya interaction and weak ferromagnetism [1–3], the linear magnetoelectric (ME) effect [4–6], piezomagnetism [7, 8], linear magnetostriction [9], and a variety of unusual optical effects associated with the antiferromagnetic vector, e.g., quadratic Faraday effect [10] or linear birefringence [11, 12], were discovered in the past century. Similar effects also occur in kinetics [13] (see also [14–16]).

The occurrence of one or other of these effects in an antiferromagnetic crystal with the centrosymmetric crystallographic structure depends on the space parity I of its magnetic structure. For example, weak ferromagnetism arises in crystals with the even magnetic structure, whereas the linear magnetoelectric effect is forbidden in them. Quite the reverse, the linear magnetoelectric effect is allowed and a weak ferromagnetism is forbidden in crystals with the odd magnetic structure. In this respect, crystals with trigonal symmetry—rhombohedral $MnCo_3$, $FeBO_3$, and α-Fe_2O_3 antiferromagnets with the even magnetic structure and Cr_2O_3 antiferromagnet with the odd magnetic structure—have received the most study. The crystal structures of all these materials possess the inversion center (they belong to the space group $R3\bar{c} = D_{3d}^6$). In this work, the physical properties of bismuth ferrite $BiFeO_3$ (and other materials on its base) are considered. This material is related to the aforementioned rhombohedral antiferromagnets but is different from them in that the inversion center is absent in both its crystal and magnetic structures. It will be shown that quite diversified and uncommon properties of this material are due precisely to this fact.

2. CRYSTAL AND MAGNETIC STRUCTURES OF $BiFeO_3$

The crystal structure of bismuth ferrite is characterized by a rhombohedrally distorted perovskite unit cell (Fig. 1a) that is very close to a cube with the edge a_c = 3.96 Å and the almost direct angle α = 89.4° at its face. However, when analyzing the properties of bismuth ferrite, it is more suitable to consider a hexagonal cell (Fig. 1b) with the parameters a_{hex} = 5.58 Å and c_{hex} = 13.9 Å. The iron and bismuth ions are offset from their centrosymmetric positions, resulting in the spontaneous polarization P_s along the $[111]_c$ ($[001]_{\text{hex}}$) direction [17]. Early neutron diffraction studies [18] have shown that bismuth ferrite possesses the antiferromagnetic ordering of the G type, where every atom is surrounded by six atoms with oppositely oriented spins. This signi-

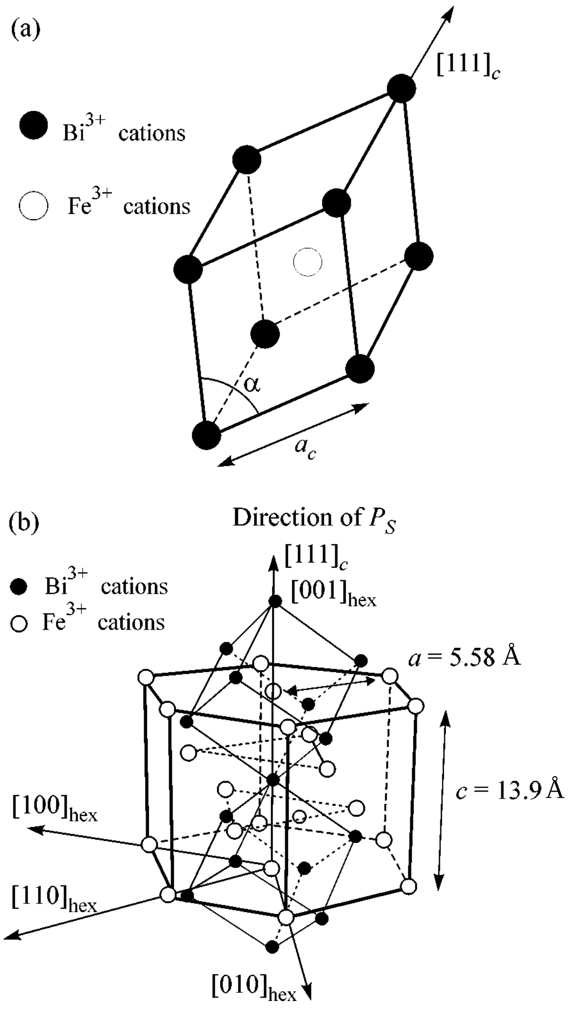


Fig. 1. Perovskite crystal structure: (a) rhombohedral and (b) hexagonal unit cells.

fies that bismuth ferrite is a ferroelectric magnet [19], i.e., a material with the coexisting magnetic and electric order parameters.

More precise measurements performed on a time-of-flight neutron diffractometer [20] revealed a more complicated spatially modulated magnetic structure with a large period $\lambda = (620 \pm 20$ Å) incommensurate with the lattice parameter. The magnetic moments of iron ions retain their local mutually antiferromagnetic *G*-type orientation and rotate along the propagation direction of the modulated wave in the plane perpendicular to the hexagonal basal plane. The spin distribution along the spatially modulated structure was determined from the NMR spectra obtained in the experiments of Zalesskiĭ *et al.* in [21–23].

3. SYMMETRY AND MAGNETOELECTRIC INTERACTIONS

To analyze the magnetoelectric properties of bismuth ferrite, we choose the space group $R3\bar{c}$ as a “parent” symmetry for the $R3c$ symmetry under study. The latter differs from $R3\bar{c}$ only by the presence of a polar vector $\mathbf{P} = (0, 0, P_s)$. Although, in reality, the phase transition in BiFeO$_3$ at the Curie point T_C differs from the assumed $R3c \longrightarrow R3\bar{c}$ transition, this is of no importance for our purposes, i.e., for determining the adequate invariants that are responsible for the properties of the system. Indeed, the parent symmetry $R3\bar{c}$ can be used to develop the perturbation theory for determining the thermodynamic potential and other physical quantities of the system at the sacrifice of the smallness of **P**; i.e., we develop the perturbation theory with respect to **P**. Correspondingly, $\xi = \Delta a/a$ is a small parameter, where a is the lattice constant and Δa is the characteristic atomic deviation from the positions symmetric about the space inversion in $R3c$.

The exchange magnetic structure, i.e., the mutual directions of magnetic moments in crystal, is determined by the following code (Turov nomenclature [24]): I^-, 3_Z^+, 2_X^+, where the space-inversion element I, the threefold axis 3_Z (aligned with the c axis), and the twofold axis x are the group generators and the indices ± of these elements specify their parity about the transposition of magnetic sublattices; i.e., the sign “+” indicates that the symmetry element transposes the ions within the same magnetic sublattice of an antiferromagnet, and the sign “–” indicates that the sublattice is transposed into the sublattice with the opposite spin direction upon the symmetry operation. With these symmetry operators, the antiferromagnetic vector obeys the following transformation rules:

$$I^{\pm}\mathbf{L} = \pm\mathbf{L}, \quad 2_X^{\pm}L_X = \pm L_X, \quad 2_X^{\pm}L_{Y(Z)} = \mp L_{Y(Z)}.$$

For the other vectors, the action of elements with different indices is the same:

$$I^{\pm}\mathbf{m} = \mathbf{m}, \quad 2_X^{\pm}m_X = m_X, \quad 2z^{\pm}m_{Y(Z)} = -m_{Y(Z)},$$

$$I^{\pm}\mathbf{P} = \mathbf{P}, \quad 2_X^{\pm}P_X = P_X, \quad 2_X^{\pm}P_{Y(Z)} = -P_{Y(Z)}.$$

The layout of the group elements relative to the magnetic ions in bismuth ferrite is shown in Fig. 2a. Of interest is to compare the code of this structure (BiFeO$_3$) with the codes of other antiferromagnetic compounds belonging to the space group $R3\bar{c}$: I^+, 3_Z^+, 2_X^- for hematite (α-Fe$_2$O$_3$) and I^-, 3_Z^+, 2_X^- for chromite (Cr$_2$O$_3$), whose exchange structures are shown in Figs. 2b and 2c, respectively. As far as we know, materials with the I^-, 3_Z^+, 2_X^+ code have not been considered as yet.

The space group $R3\bar{c}$ contains eight irreducible representations: four one-dimensional (Γ_1, Γ_2, Γ_4, Γ_5) and four two-dimensional (Γ_3, Γ_3', Γ_6, Γ_6') representations (Table 1). Their matrix representations are given in the columns corresponding to the generating symmetry elements. The vector components of electric field **E**, magnetic field **H**, electric polarization **P**, magnetization **m**, and antiferromagnetic vectors $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{L}_3$ for the exchange structures of bismuth ferrite, hematite, and chromite, respectively (Fig. 2), form irreducible representations and are placed in Table 1 according to their transformation properties. For instance, it follows from this table that the $(L_z)_1$ component changes sign and the vector $L_\perp = (L_x, L_y)_1$ transforms into $(L_x, -L_y)_1$ under the symmetry operation 2_X^+. The transformation properties of the products M_iL_i, H_iE_i, and L_iE_i for bismuth ferrite are given in Table 2.

One can readily see in Table 1 that the $(H_xl_y - H_yl_x)$ combination in hematite, where $\mathbf{l} = \mathbf{L}_2/2M_0$ is the unit antiferromagnetic vector, corresponds to the first irreducible representation, i.e., is invariant. This invariant is responsible for the formation of weak magnetization $(M_x, M_y) \sim (l_y, -l_x)$ in hematite; i.e., the latter is a weak ferromagnet with the magnetization vector perpendicular to the antiferromagnetic vector.

It can also be shown that the spontaneous magnetization is forbidden in chromite Cr_2O_3. The combinations H_jl_k, where $\mathbf{l} = \mathbf{L}_3/2M_0$, are not invariant. At the same time, the chromite symmetry allows the magnetoelectric effect because of the presence of the $(E_xH_x + E_yH_y)l_z$, $E_zH_zl_z$, $H_z(E_xl_x + E_yl_y)$, and $E_z(H_xl_x + H_yl_y)$ invariants.

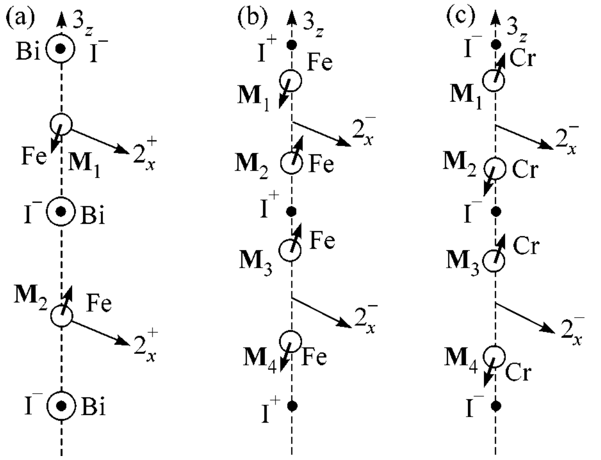


Fig. 2. Exchange structures of (a) bismuth ferrite $BiFeO_3$, $\mathbf{L}_1 = \mathbf{M}_1 - \mathbf{M}_2$; (b) hematite α-$Fe_2O_3$, $\mathbf{L}_2 = \mathbf{M}_1 - \mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4$; and (c) chromite Cr_2O_5, $\mathbf{L}_3 = \mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_3 - \mathbf{M}_4$.

From Tables 1 and 2, one can easily find the following invariant for bismuth ferrite:

$$2M_0P_z(m_yl_x - m_xl_y), \tag{1}$$

where P_z is the component of spontaneous polarization $\mathbf{P} = (0, 0, P_z)$ along the c axis, M_0 is the magnitude of the magnetization vector of sublattices, and $\mathbf{m} = (\mathbf{M}_1 + \mathbf{M}_2)/2M_0$ and $\mathbf{l} = \mathbf{L}_1/2M_0$ are the unit magnetization and antiferromagnetic vectors, respectively. This interac-

Table 1. Irreducible representations of the space group $R3\bar{c}$

	E^+	$I^{\pm}$	3_z^+	$2_x^{\pm}$	E_i; P_i	H_i; M_i	L_i
Γ_1	1	1	1	1			$(L_z)_2$
Γ_2	1	1	1	–1		H_z, m_z	
Γ_3	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	R	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$		$\begin{pmatrix} H_x \\ H_y \end{pmatrix} \begin{pmatrix} m_x \\ m_y \end{pmatrix}$	
Γ'_3	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	R	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$			$\begin{pmatrix} L_x \\ L_y \end{pmatrix}_2$
Γ_4	1	–1	1	1			$(L_z)_3$
Γ_5	1	–1	1	–1	E_z, P_z, ∇_z		$(L_z)_1$
Γ_6	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	R	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} E_x \\ E_y \end{pmatrix}$; $\begin{pmatrix} P_x \\ P_y \end{pmatrix}$; $\begin{pmatrix} \nabla_x \\ \nabla_y \end{pmatrix}$		$\begin{pmatrix} L_x \\ L_y \end{pmatrix}_1$
Γ'_6	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	R	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$			$\begin{pmatrix} L_x \\ L_y \end{pmatrix}_3$

Note: R is the matrix of rotation through an angle of 120° about the z axis (c axis in the hexagonal system).

Table 2. Irreducible representations for bismuth ferrite

	E^+	I^-	3_z^+	$2_x^\pm$	M_iL_i; H_iL_i; H_iE_i; L_iE_i
Γ_1	1	1	1	1	$(E_xL_x + E_yL_y)$
Γ_2	1	1	1	–1	$(E_xL_y - E_yL_x)$
Γ_3	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	R	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	
Γ_3'	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	R	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	
Γ_4	1	–1	1	1	$(m_xL_x + m_yL_y)$
Γ_5	1	–1	1	–1	$(m_yL_x - m_xL_y)$; $(H_yL_x - H_xL_y)$; $(H_yE_x - H_xE_y)$
Γ_6	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	R	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} H_yL_y - H_xL_x \\ H_xL_y + H_yL_x \end{pmatrix}$
Γ_6'	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	R	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	

Note: R is the matrix of rotation through an angle of 120° about the z axis (c axis in the hexagonal system).

tion has the same form as the Dzyaloshinski–Moriya interaction and, hence, also gives rise to a weak ferromagnetism with the magnetization

$$\mathbf{m} \sim (P_zl_y - P_zl_x, 0). \tag{2}$$

This is by no means contradictory to the well-known theorem in the theory of antiferromagnetism about the impossibility of a weak ferromagnetism coexisting with the magnetoelectric effect [24]. This theorem is related to those antiferromagnets whose space group contains space-inversion element (i.e., whose crystal structure is odd about the space inversion). In our case ($BiFeO_3$), we deal with ferroelectric magnets, where this requirement is not fulfilled. It is worth noting that the physical nature of a weak ferromagnetism in ferroelectric magnets is basically different from the Dzyaloshinski–Moriya case. A weak ferromagnetic moment in ferroelectric magnets results from the magnetoelectric interaction. In other words, this magnetic moment arises due to the magnetoelectric interaction in the internal effective electric field.

The expression for the free-energy density also includes the magnetoelectric terms proportional to the invariants of the $H_iE_il_i$ type:

$$\begin{aligned} f = \ldots - a_1[E_x(H_yl_y - H_xl_x) \\ + E_y(H_xl_y + H_yl_x)] - a_2H_z(E_xl_y - E_yl_x) \\ - a_3E_z(H_yl_x - H_xl_y) - a_4l_z(H_yE_x - H_xE_y) + \ldots. \end{aligned} \tag{3}$$

For the tensor relating the magnetic-field-induced polarization vector $P_i = \alpha_{ij}H_j$ to the magnetic vector in the linear magnetoelectric effect, one obtains

$$\alpha_{ij} = \begin{vmatrix} -a_1l_x & a_4l_z + a_1l_y & a_2l_y \\ a_1l_y - a_4l_z & a_1l_x & -a_2l_x \\ -a_3l_y & a_3l_x & 0 \end{vmatrix}. \tag{4}$$

Apart from the magnetoelectric (ME) effect and spontaneous magnetization, the magnetic symmetry of bismuth ferrite also allows the magnetic ordering of special (toroidal) type [25, 26]. The presence of toroidal moment is due to the term proportional to the

$$(\mathbf{T}[\mathbf{E} \times \mathbf{H}]) \tag{5}$$

invariant in the expression for the free energy. Hence, it follows that the vector components of the toroidal moment are proportional to the antisymmetric part of the tensor of the linear ME effect:

$$T_i = \varepsilon_{ijk}\alpha_{jk}, \tag{6}$$

i.e.,

$$T_1 \sim \alpha_{23} - \alpha_{32}, \quad T_2 \sim \alpha_{31} - \alpha_{13}, \quad T_3 \sim \alpha_{12} - \alpha_{21}. \tag{6a}$$

By using Eqs. (6a) and (4), one can readily verify that the vector components of the toroidal moment in bismuth ferrite are proportional to the components of antiferromagnetic vector:

$$\begin{pmatrix} T_x \\ T_y \end{pmatrix} \sim \begin{pmatrix} l_x \\ l_y \end{pmatrix}; \quad T_z \sim l_z. \qquad (7)$$

The analysis of the irreducible representations of crystal space groups allows one to predict not only the properties of magnetic materials but also their microscopic magnetic structure. Taking into account that the differential operator ∇ transforms as a polar vector, we can write the new term in the expression for the free-energy density (Table 1):

$$f_L = \gamma P_z(l_x \nabla_x l_z + l_y \nabla_y l_z - l_z \nabla_x l_x - l_z \nabla_y l_y), \qquad (8)$$

where γ is the coefficient of the ME origin. Note that, by analogy with the invariant of the form $l_i(\partial l_j/\partial x) - l_j(\partial l_i/\partial x)$, combination (8) was named the Lifshitz invariant. However, the distinctive feature of combination (8) is that it is nonzero only in the presence of spontaneous polarization. Lifshitz invariant (8) has the ME origin, but, contrary to the homogeneous ME interaction of the Dzyaloshinski–Moriya type, it is nonzero only in the presence of an inhomogeneous antiferromagnetic structure, i.e., only if the spatial derivatives of the antiferromagnetic vector are nonzero.

Therefore, two types of ME interaction are present in bismuth ferrite: a homogeneous interaction characterized by the tensor α_{ij} of the linear ME effect and an inhomogeneous interaction characterized by the γ constant. A distinctive feature of bismuth ferrite is the presence of spontaneous polarization in it. This implies that the space-inversion element is absent in its crystallographic space group and manifests itself in the coexistence of weak ferromagnetism (2), ME effect (4), and inhomogeneous ME interaction (8).

Interestingly, the inversion symmetry of chromite Cr_2O_3 is violated upon applying the external electric field E^{ext} along the c axis, and the spontaneous magnetization arises:

$$\mathbf{m} \sim (E^{ext} l_x, E^{ext} l_y, 0), \qquad (9)$$

where $\mathbf{l} = \mathbf{L}_3/2M_0$ is the unit antiferromagnetic vector.

4. SPATIALLY MODULATED STRUCTURE AND MAGNETIC-INDUCED PHASE TRANSITIONS

The total expression for the free-energy density has the form

$$f = f_L + f_{exch} + f_{an}, \qquad (10)$$

where

$$f_{exch} = A \sum_{i = x, y, z} (\nabla l_i)^2 = A((\nabla\theta)^2 + \sin^2\theta(\nabla\varphi)^2) \qquad (11)$$

is the exchange energy; $A = 3 \times 10^7$ erg/cm is the constant of inhomogeneous exchange (exchange stiffness);

$$f_{an} = K_u \sin^2\theta \qquad (12)$$

is the anisotropy energy; θ and φ are the polar and azimuthal angles of the unit antiferromagnetic vector; and $\mathbf{l} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ in the spherical coordinate system with the polar axis aligned with the principal axis c (hexagonal system).

The minimization of the free-energy functional $F = \int f\, dV$ by the Lagrange–Euler method in the approximation ignoring anisotropy gives for the functions $\theta(x, y, z)$ and $\varphi(x, y, z)$ [27, 28]

$$\varphi_0 = \text{const} = \arctan(q_y/q_x), \quad \theta_0 = q_x x + q_y y, \qquad (13)$$

where $\mathbf{q}$ is the spiral wave vector. Solution (13) describes a cycloid whose plane is perpendicular to the basal plane and oriented along the propagation direction of the modulation wave.

A more precise solution allowing for the anisotropy K_u of the material is given by the formula [28, 29]

$$\cos\theta = \text{sn}(q_x x, m = -K_u/E), \qquad (14)$$

which corresponds to an anharmonic cycloid. For an anisotropy constant much smaller than the exchange energy, $K_u \ll E \sim Aq^2$, the module m of elliptic sine tends to zero, and solution (14) turns into harmonic solution (13).

It is worth noting that the presence of spatially modulated structure was confirmed not only by neutron diffraction but also by nuclear magnetic resonance [21–23]. Instead of a single peak corresponding to a homogeneous structure, the NMR line was more complicated and showed two maxima corresponding to the spin orientations perpendicular and parallel to the principal axis (Fig. 3). The analysis of the NMR lineshape not only showed the presence of a cycloid but also allowed the spin distribution to be reconstructed along the cycloid. At low temperatures (T = 77 K), this distribution proved to be essentially anharmonic: for most of the cycloid period, spins form a small angle with the axis, as was seen from the higher intensity of the high- frequency peak. With an increase in temperature, the lineshape becomes more symmetric, the anharmonicity decreases, and, at room temperature, the coordinate dependence of the angle becomes close to linear [21] (Figs. 3a, 3b). By substituting Eq. (13) into Eq. (10), one obtains for the volume-averaged free-energy density in the approximation of harmonic cycloid

$$\langle F \rangle = Aq^2 - (\gamma P_z)q + K_u/2. \qquad (15)$$

The wave vector corresponding to the energy minimum is

$$q = \frac{2\pi}{\lambda} = \frac{\gamma P_z}{2A}. \qquad (16)$$

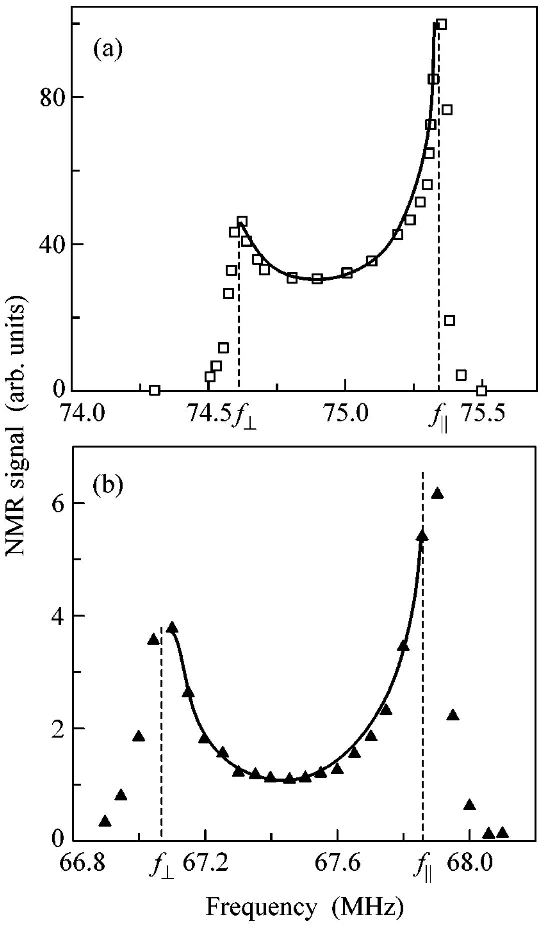


Fig. 3. ^{57}Fe NMR spectra of $BiFeO_3$ at (a) 77 and (b) 304 K [21].

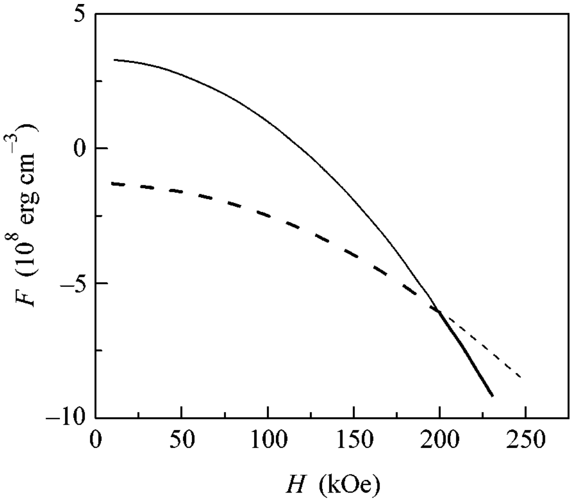


Fig. 4. Dependence of the free-energy density on the magnetic field. The solid line is for the homogeneous state and the dashed line is for the spatially modulated state.

Knowing the structure period ($\lambda = 620$ Å) and assuming that the polarization is $P_z = 6 \times 10^{-6}$ C/cm^2 and the exchange constant is $A = 3 \times 10^{-7}$ erg/cm, one can estimate the inhomogeneous ME coefficient $\gamma = 10^5$ erg/C = 10^{-2} V.

Because of the presence of a cycloid, volume-averaged ME effect (4), spontaneous magnetization (2), and toroidal moment (7) are zero. All three effects appear only upon the suppression of the spatially modulated structure.

The modulated structure can be suppressed, e.g., by applying a strong magnetic field. Magnetic field adds the following term to the expression for the anisotropy constant (field is applied along the threefold axis c):

$$K_u = K_u^0 - \chi_\perp \frac{H_z^2}{2}, \quad (17)$$

where $\chi_\perp$ is the magnetic susceptibility in the direction perpendicular to the antiferromagnetic vector and K_u^0 is the uniaxial anisotropy constant in zero field. For fields higher than a certain critical value, the spatially modulated structure may become energetically unfavorable, as compared to the homogeneous state whose energy is given by the formula

$$F_{\text{hom}} = K_u, \quad (18)$$

where K_u is effective anisotropy (17).

The field dependences of the free-energy density are presented in Fig. 4 for spatially modulated structure (15) and homogeneous state (18). Although the spatially modulated structure at low fields is energetically more favorable than the homogeneous state, the situation changes in the range of high fields. The critical field H_c, at which the phase transition occurs, is determined from Eqs. (15) and (18):

$$H_c = \sqrt{\frac{2(K_u^0 + 2Aq^2)}{\chi_\perp}}. \quad (19)$$

The estimate with $K_u^0 \ll Aq^2$, $A = 3 \times 10^{-7}$ erg/cm, and $\chi_\perp = 4.7 \times 10^{-5}$ gives for the critical field ~200 kOe. The phase transitions in bismuth ferrite with an anharmonic cycloid were considered theoretically in [29].

The suppression of the spatially modulated structure in bismuth ferrite and appearance of the linear ME effect and the toroidal moment were confirmed experimentally by measuring the field dependence of polarization in pulsed fields [27, 30, 31]. At $H < H_c$, the polarization depends on field almost quadratically, whereas, at the critical field, the polarization undergoes an abrupt jump that is accompanied by the appearance of the linear ME effect and the renormalization of the quadratic ME tensor (Fig. 5). The values of the critical field and the linear ME effect were found to be ~200 kOe and ~10^{-10} C/(m^2 Oe), respectively.

The orientation of the cycloid plane in crystal was determined from the measurements of electric a and b polarizations as functions of a magnetic field applied along the crystal c axis [31]. The $P_a(H_c)$ and $P_b(H_c)$ dependences are shown in Fig. 6. One can see that, after the cycloid disappears in the field $H_c \approx 200$ kOe, the electric polarization increases jumpwise and varies at $H > 250$ kOe almost linearly with field. It would be natural to assume that the spins of Fe^{3+} ions do not offset from the cycloid plane in the resulting structure but “lie down” in the basal plane so that $L_z = 0$ and $L_y/L_x = \tan\varphi$, where φ, according to the neutron diffraction data, is the angle between the cycloid plane and the crystal a axis. Then, in the linear approximation (Eq. (4)), one obtains from the slopes of the linear portions of experimental curves:

$$\frac{dP_a/dH}{dP_b/dH} = \frac{\alpha_{13}}{\alpha_{23}} = -\frac{l_y}{l_x} = -1.8. \qquad (20)$$

The resulting value $\varphi = -60°$ corresponds to the angle obtained in [20] from the neutron diffraction data ($[110]_{hex}$ direction in Fig. 1b).

In [27, 30, 31], measurements were carried out in pulsed fields. Recent experiments [32] on the observation of antiferromagnetic resonance in a static magnetic field have shown that the eigenfrequency spectrum of bismuth ferrite strongly changes near the point $H_c = 180$ kOe (Fig. 7). The authors of that work assigned this change to the phase transition from the spatially modulated structure to the homogeneous state [32]. It was also shown that the phase transition is accompanied by an appreciable hysteresis of the resonance lines in the increasing and decreasing fields (Fig. 8). A value of 3.3 V/(cm Oe) found for the ME coefficient from the curve-fitting procedure in the above-critical fields is in good agreement with the giant ME effect in thin bismuth ferrite films.

5. WEAK FERROMAGNETISM

The symmetry analysis of crystal properties suggests that a weak ferromagnetism caused by the canted sublattice magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$ may occur in bismuth ferrite.

Spontaneous magnetization (2) depends on the mutual arrangement of the polarization and antiferromagnetic vectors and is proportional to the sine of the angle between the polarization and antiferromagnetic vectors:

$$m \sim PL\sin\theta. \qquad (21)$$

It follows from Eq. (21) that the canting angle φ between the sublattice magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$ is determined as

$$\varphi = m/L \sim \sin\theta, \qquad (22)$$

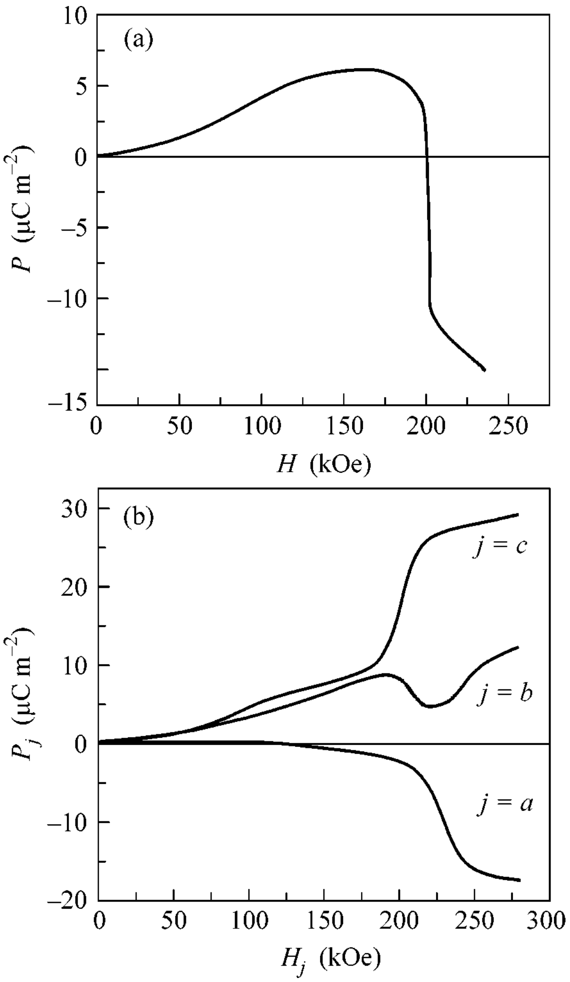


Fig. 5. Dependence of the longitudinal electric polarization on the magnetic field for different crystallographic directions: (a) along the cubic axis and (b) along the a, b, and c axes in the hexagonal system [27, 30, 31].

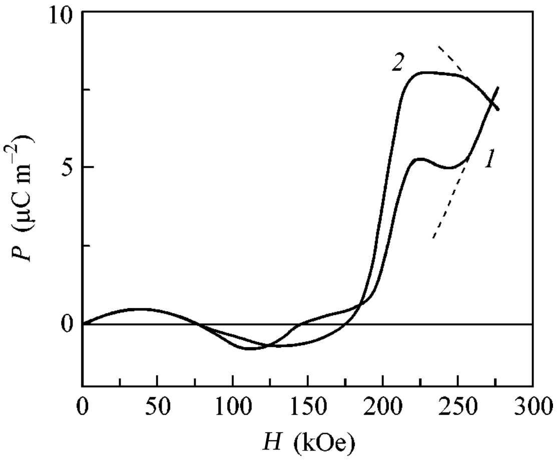


Fig. 6. Experimental curves obtained for the polarizations along different directions in a magnetic field $H \parallel c$ at a temperature of 18 K: (*1*) along the b axis and (*2*) along the a axis. The dashed lines show the slopes of the linear portions of the curves at $H > 250$ kOe, which were used to calculate the derivatives dP_a/dH and dP_b/dH by formula (20).

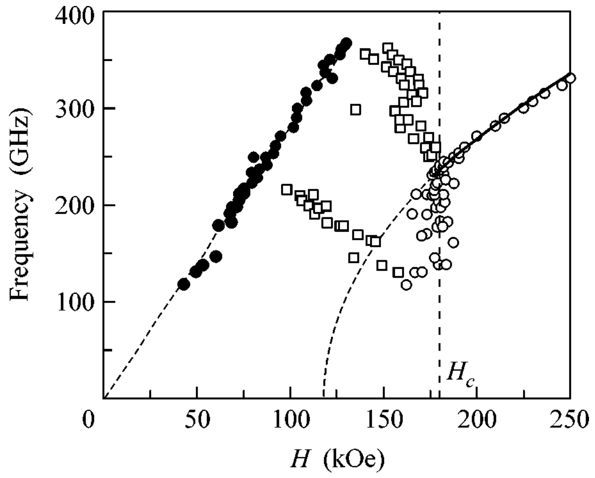


Fig. 7. Antiferromagnetic resonance frequencies as functions of magnetic field H ($T = 4.2$ K) [32].

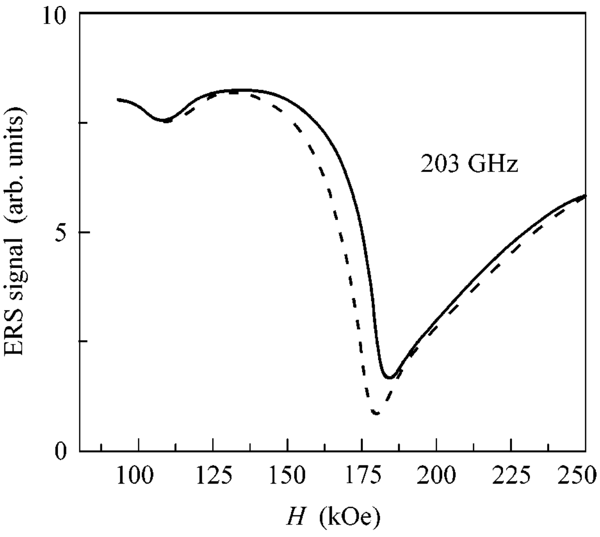


Fig. 8. Magnetic hysteresis of the absorption peak. Solid line is for the increasing field, and dashes are for the decreasing field [32].

i.e., it is maximal if the antiferromagnetic vector is perpendicular to the crystal principal axis and is zero if it is parallel to the latter.

The presence of the spatially modulated structure results in a periodic variation of the canting angle between the sublattices, so that magnetization (21) averaged over the cycloid period is zero.

It is natural to expect that spontaneous magnetization can arise in fields higher than the critical field H_c, in which the spatially modulated structure disappears and a homogeneous antiferromagnetic ordering is established.

In the very recent past [34], the magnetization of $BiFeO_3$ was measured as a function of magnetic field (Fig. 9). Measurements were made at $T = 10$ K for the $[001]_c$ direction (along the edge of a crystal with natural faces shaped almost like cube faces). Identical magnetization curves were also obtained for the $[100]_c$ and $[010]_c$ directions. Measurements for the $\mathbf{H} \parallel \mathbf{c}$ geometry were not carried out because of the poor sensitivity of the setup (this could be done only with small-sized oriented samples).

It is seen that the run of magnetization curve changes sharply at fields close to critical $H_c \sim 200$ kOe because of the suppression of cycloid by the magnetic field and transition to a homogeneous antiferromagnetic phase.

The experimental magnetization curves at low and high fields are well described by the linear dependences

$$M = \chi H, \quad H < 100 \text{ Oe}, \tag{23a}$$

$$M = M^{\text{spont}}_{[001]} + \chi_\perp H, \quad H > H_c, \tag{23b}$$

i.e., the spontaneous magnetization $M^{\text{spont}}_{[001]} = 0.25 \pm 0.02$ (G cm^3)/g (1.95 ± 0.15 G) arises in bismuth ferrite above the critical field H_c, while the field dependence in this range is described by a linear function with the slope equal to the susceptibility $\chi_\perp = 0.6 \times 10^{-5}$ (G cm^3)/(g Oe) of the material in the direction perpendicular to the antiferromagnetic vector [35]. In low fields ($H < 100$ kOe), the field dependence of magnetization is nicely described by the direct proportionality, where the susceptibility χ comprises approximately 5/6 of the total susceptibility $\chi_\perp$. In the range of intermediate fields ($100 \text{ kOe} < H < 200$ kOe), the dependence is noticeably nonlinear.

The observed experimental dependence can be understood if the spontaneous magnetization $\langle M^{sp}\rangle$ averaged over the cycloid period and the magnetic-induced magnetization M^H depend on the external field $\mathbf{H}$:

$$\mathbf{M}(H) = \langle \mathbf{M}^{\text{spont}}\rangle(H) + \mathbf{M}^H(H). \tag{24}$$

Considering that the field in the experiment is aligned with the $[001]_c$ axis, i.e., $\mathbf{H} = (H\sqrt{2}/\sqrt{3}, 0, H/\sqrt{3})$, one has for the total magnetization along the $[001]_c$ direction

$$\begin{aligned} M_{[001]}(H) &= M^{\text{spont}}_{[001]}\langle \sin\theta\rangle_\lambda \\ &+ \underbrace{\frac{2}{3}\chi_\perp H}_{M^H_{x[001]}} + \underbrace{\frac{1}{3}\chi_\perp H\langle \sin^2\theta\rangle_\lambda}_{M^H_{z[001]}}, \end{aligned} \tag{25}$$

where θ is the angle between the spontaneous magnetization P_z (c axis) and the antiferromagnetic vector and $M^H_{x[001]}$ and $M^H_{z[001]}$ are, respectively, the projections of

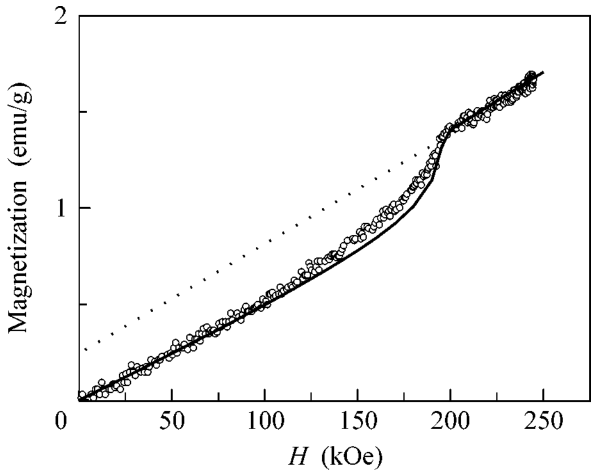


Fig. 9. Magnetization of bismuth ferrite as a function of magnetic field at a temperature of 10 K [34]. Dots are the experimental data obtained in a field oriented along the $[001]_c$ direction, and solid line is the theoretical dependence (Eq. (28)).

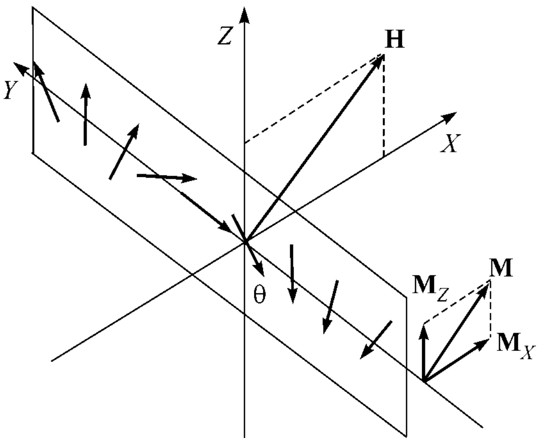


Fig. 10. Cycloidal spin structure in an external magnetic field. The directions of the field **H** and magnetization **M** are different because the spatially modulated spin structure is anisotropic.

the M_x and M_z components onto the $[001]_c$ direction (Fig. 10). Equation (25) can be used to explain the experimentally observed dependences (23). The values of sine $\langle\sin\theta(y)\rangle_\lambda$ and the square of sine $\langle\sin^2\theta(y)\rangle_\lambda$ averaged over the spatial period are different for low $H \approx 0$ and high $H \longrightarrow H_c$ fields and vary within (0, 1) and (1/2, 1), respectively. Dependence (25) turns into (23a) at $H \longrightarrow 0$ and to (23b) at $H \longrightarrow H_c$.

6. TOROIDAL ORDERING

Measurements of the field dependence of electric polarization showed that a toroidal moment appears upon the magnetic phase transition in bismuth ferrite [31].

To determine the toroidal moment in $BiFeO_3$, the antisymmetric part (proportional to the toroidal moment) of the tensor of the linear ME effect was examined at fields H exceeding the phase-transition field H_c and oriented at an angle of 45° to the crystal a and b axes in the basal plane. Such a field orientation made it possible to measure the polarization components $P_a(H)$ and $P_b(H)$ in the same state of the sample.

For the indicated field orientation and $H > H_c$, the antiferromagnetic vector is aligned with the threefold axis $L \parallel c$ ($L_x = L_y = 0$) and the crystal belongs to the 3m magnetic class, so that α_{12} and α_{21} (see Eq. (4)) are the only nonzero components, and, in addition,

$$\alpha_{12} = -\alpha_{21}. \qquad (26)$$

The matrix β_{ijk} of the ME effect quadratic in magnetic field is more cumbersome, and the nonzero matrix elements in this case are

$$\beta_{111} = \gamma_1, \quad \beta_{122} = \beta_{212} = \beta_{221} = -\gamma. \qquad (27)$$

Thus, taking into account that $H_a = H_b = H/\sqrt{2}$, one finds

$$P_a = \frac{\alpha_{12}}{\sqrt{2}}H, \quad P_b = \frac{\alpha_{21}}{\sqrt{2}}H - \gamma_1 H^2. \qquad (28)$$

By fitting the tails of experimental curves (Fig. 11) to the linear (for $P_a(H)$) and quadratic (for $P_b(H)$) approximations (the approximating straight line and parabola are shown by dashes), the following ME coefficients were obtained:

$$\alpha_{12} = -(0.029 \pm 0.003) \times 10^{-6}\ \mathrm{C/(m^2\ kOe)},$$

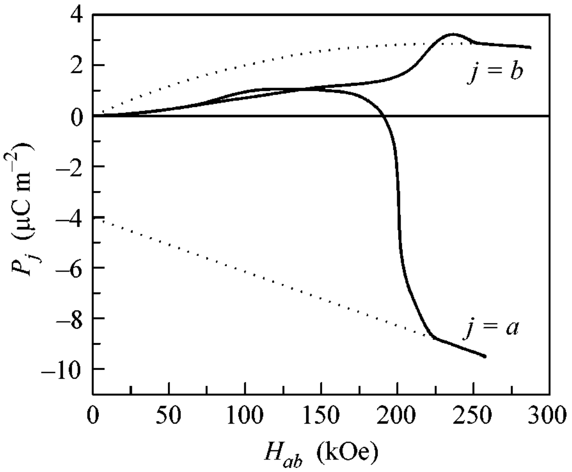


Fig. 11. The polarizations along the a and b directions as functions of a field applied in the basal plane at an angle of 45° to the a axis [31]. Dotted curves are the theoretical dependences calculated by Eqs. (28).

$$\alpha_{21} = +(0.032 \pm 0.003) \times 10^{-6}\ \mathrm{C/(m^2\ kOe)}, \quad (29)$$

$$\gamma_1 = 5 \times 10^{-11}\ \mathrm{C/(m^2\ kOe)}.$$

Thus, the obtained α_{12} and α_{21} components differ only in sign and, to within the error, are equal in magnitude, in full agreement with Eq. (29). The fact that the nondiagonal tensor components of the linear ME effect are antisymmetric is evidence that the toroidal moment $T_z \sim (\alpha_{12} - \alpha_{21})$ will appear after the suppression of cycloid.

7. NEW MATERIALS WITH GIANT MAGNETOELECTRIC EFFECT

Apart from the strong magnetic field, other methods exist of suppressing the spatially modulated structure. One such method is the substitution of rare-earth ions for the bismuth ions. The compounds with formula $RFeO_3$ (rare-earth orthoferrites) also have the perovskite structure, though orthorhombically distorted. The introduction of rare-earth impurities can increase the anisotropy constant to the extent that the presence of spatially modulated structure would be energetically unfavorable. Measurements of the magnetic-field dependences of polarization in the $Bi_{1-x}R_xFeO_3$ (R = La, Gd, Dy; $0.4 < x < 0.5$) compounds [35–38] showed the presence of the linear ME effect up to the liquid nitrogen temperature, evidencing the suppression of the spatially modulated spin structure in the compounds of this type. The high-field polarization measurements showed that the spatially modulated structure is retained in bismuth ferrite with a smaller lanthanum content ($x < 0.3$), but the presence of lanthanum additives reduces the transition field from the spatially modulated to the homogeneous state [37, 38].

Interesting results were obtained in thin bismuth ferrite films of size 50–500 nm [33], where the observed ME effect $dE/dH = 3$ V/(cm Oe) was an order of magnitude greater than in the classical chromite magnetoelectric. According to the presently accepted nomenclature, such effects are referred to as giant effects. The modulated structure in these compounds is suppressed likely due to the stresses that arise during the epitaxial growth and can produce, through the magnetostriction or piezoelectric mechanism, magnetic and electric fields that are critical to the phase transitions.

In a search for materials exhibiting the giant ME effect, attention of researchers has also be drawn to the bismuth ferrite-based solid solutions $BiFeO_3$–$xPbTiO_3$ with the substitution of lanthanum for the bismuth ions [39]. In these compounds, an anomalously high electric polarization and the appearance of spontaneous magnetization are observed near the point of phase transition ($PbTiO_3$ content $x \sim 30\%$ [40]) from the orthorhombic to the tetragonal phase [39]. These two phenomena can serve as indirect evidence of the presence of a rather strong ME effect in these materials.

8. CONCLUSIONS

Symmetry analysis has been applied to the magnetic and magnetoelectric properties of antiferromagnets. It has been shown that the violation of the inversion symmetry in both crystal and magnetic structures of the bismuth ferrite antiferromagnet $BiFeO_3$ gives rise to the unique possibility of the linear ME effect coexisting with a weak ferromagnetic moment in $BiFeO_3$, which is basically impossible for conventional antiferromagnets. The polarization breaks the crystal central symmetry and manifests itself also in the appearance of an inhomogeneous ME interaction (Lifshitz invariant), leading to the formation of a spatially modulated spin structure in the material. The suppression of the spatially modulated spin structure is the necessary condition for the presence of the linear ME effect, the spontaneous magnetization, and the toroidal moment. One of the ways of suppressing the spin cycloid consists in applying a strong magnetic field, which induces the phase transition from the spatially modulated state to the homogeneous state, after which all three above effects appear. Other methods of suppressing the spin cycloid (the substitution of rare-earth ions for the bismuth ions and the fabrication of bismuth ferrite epitaxial films) made it possible to obtain a giant ME effect that is an order of magnitude greater than the previously observed effect.

REFERENCES

1. L. Neel and R. Pauthenet, C. R. Acad. Sci., Paris **234**, 2172 (1952).
2. A. S. Borovik-Romanov and M. P. Orlova, Zh. Éksp. Teor. Fiz. **31**, 579 (1956) [Sov. Phys. JETP **4**, 531 (1956)].
3. I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **32**, 1547 (1957) [Sov. Phys. JETP **5**, 1259 (1957)]; F. Moriya, Phys. Rev. **120**, 91 (1960).
4. I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **37**, 881 (1959) [Sov. Phys. JETP **10**, 628 (1960)].
5. D. N. Astrov, Zh. Éksp. Teor. Fiz. **38**, 984 (1960) [Sov. Phys. JETP **11**, 708 (1960)].
6. V. J. Folen, G. T. Rado, and E. W. Stalder, Phys. Rev. Lett. **6**, 607 (1961).
7. I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **33**, 807 (1957) [Sov. Phys. JETP **6**, 621 (1958)].
8. A. S. Borovik-Romanov, Zh. Éksp. Teor. Fiz. **38**, 1088 (1960) [Sov. Phys. JETP **11**, 786 (1960)].
9. A. S. Borovik-Romanov and B. E. Javelov, in *Proceedings of 3rd Regional Conference on Magnetics* (Prague, 1963), p. 81.
10. N. F. Kharchenko, A. V. Bibik, and V. V. Eremenko, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 533 (1985) [JETP Lett. **42**, 553 (1985)].
11. V. S. Ostrovskiĭ and V. M. Loktev, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 134 (1977) [JETP Lett. **26**, 130 (1977)].
12. S. Leykuras, H. Legal, D. Minella, *et al.*, Physica B (Amsterdam) **89**, 43 (1977).

13. E. A. Turov and V. G. Shavrov, Izv. Akad. Nauk SSSR, Ser. Fiz. **27**, 1487 (1963).
14. E. A. Turov and V. G. Shavrov, Zh. Éksp. Teor. Fiz. **45**, 349 (1963) [Sov. Phys. JETP **18**, 242 (1964)].
15. K. B. Vlasov *et al.*, in *Dynamic and Kinetic Properties of Magnetics*, Ed. by S. V. Vonsovskiĭ and E. A. Turov (Nauka, Moscow, 1986) [in Russian].
16. E. A. Turov, *Kinetic, Optical, Acoustic Properties of Antiferromagnetics* (Akad. Nauk SSSR, Sverdlovsk, 1990).
17. J. R. Teague, R. Gerson, and W. J. James, Solid State Commun. **8**, 1073 (1970).
18. S. V. Kiselev, R. P. Ozerov, and G. S. Zhdanov, Dokl. Akad. Nauk SSSR **145**, 1255 (1962) [Sov. Phys. Dokl. **7**, 742 (1962)].
19. G. A. Smolenskiĭ and I. E. Chupis, Usp. Fiz. Nauk **137**, 415 (1982) [Sov. Phys. Usp. **25**, 475 (1982)].
20. I. Sosnowska, T. Peterlin-Neumaier, and E. Steichele, J. Phys. C **15**, 4835 (1982).
21. A. V. Zalesskiĭ, A. K. Zvezdin, A. A. Frolov, and A. A. Bush, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 682 (2000) [JETP Lett. **71**, 465 (2000)].
22. A. V. Zalesskiĭ, A. A. Frolov, A. K. Zvezdin, *et al.*, Zh. Éksp. Teor. Fiz. **122**, 116 (2002) [JETP **95**, 101 (2002)].
23. A. V. Zalesskiĭ, A. A. Frolov, T. A. Khimich, and A. A. Bush, Fiz. Tverd. Tela (St. Petersburg) **45**, 134 (2003) [Phys. Solid State **45**, 141 (2003)].
24. E. A. Turov, Usp. Fiz. Nauk **164**, 325 (1994) [Phys. Usp. **37**, 303 (1994)].
25. V. L. Ginzburg, A. A. Gorbatsevich, Yu. V. Kopaev, and B. A. Volkov, Solid State Commun. **50**, 339 (1984).
26. H. Schmid, Ferroelectrics **252**, 41 (2001).
27. Yu. F. Popov, A. K. Zvezdin, G. P. Vorob'ev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 65 (1993) [JETP Lett. **57**, 69 (1993)].
28. I. Sosnowska and A. K. Zvezdin, J. Magn. Magn. Mater. **140–144**, 167 (1995).
29. M. M. Tehranchi, N. F. Kubrakov, and A. K. Zvezdin, Ferroelectrics **204**, 181 (1997).
30. Yu. F. Popov, A. M. Kadomtseva, G. P. Vorob'ev, and A. K. Zvezdin, Ferroelectrics **162**, 135 (1994).
31. Yu. F. Popov, A. M. Kadomtseva, S. S. Krotov, *et al.*, Fiz. Nizk. Temp. **27**, 649 (2001) [Low Temp. Phys. **27**, 478 (2001)].
32. B. Ruette, S. Zvyagin, A. P. Pyatakov, *et al.*, Phys. Rev. B **69**, 064114 (2004).
33. J. Wang, H. Zheng, V. Nagarajan, *et al.*, Science **299**, 1719 (2003).
34. Yu. F. Popov, A. M. Kadomtseva, A. K. Zvezdin, *et al.*, in *Magnetoelectronic Phenomena in Crystals*, Ed. by Manfred Fiebig (Kluwer Academic, Boston, 2004).
35. Z. V. Gabbasova, M. D. Kuz'min, A. K. Zvezdin, *et al.*, Phys. Lett. A **158**, 491 (1991).
36. V. A. Murashev, D. N. Rakov, I. S. Dubenko, *et al.*, Kristallografiya **35**, 912 (1990) [Sov. Phys. Crystallogr. **35**, 538 (1990)].
37. A. M. Kadomtseva, Yu. F. Popov, G. P. Vorob'ev, and A. K. Zvezdin, Physica B (Amsterdam) **211**, 327 (1995).
38. G. P. Vorob'ev, A. K. Zvezdin, A. M. Kadomtseva, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **37**, 2428 (1995) [Phys. Solid State **37**, 1329 (1995)].
39. Naigang Wang, J. Cheng, A. Pyatakov, *et al.*, J. Appl. Phys. (in press).
40. S. A. Fedulov, Yu. N. Venevtsev, and G. S. Zhdanov, Kristallografiya **7**, 77 (1962) [Sov. Phys. Crystallogr. **7**, 62 (1962)].

Translated by V. Sakun

JETP Letters, Vol. 79, No. 11, 2004, p. 582.

Erratum: "Entangled Photon States in Consecutive Nonlinear Optical Interactions" [JETP Letters 79, 253 (2004)]

A. V. Rodionov and A. S. Chirkin

PACS numbers: 42.50.Dv; 42.65.Ky; 03.67.Mn; 03.65.Ud; 99.10.Cd

The last term in Eq. (3) contains a misprint: operator a_2 should be replaced by a_3. The proper form of Eq. (3) is

$$H_{\text{int}} = i\hbar\{\beta_1(a_1^+a_2^+ - a_1a_2) + \beta_2(a_3^+a_1 - a_3a_1^+)\}.$$