Temporal structure of the muon disk at large distances from the axis of extensive air showers with $E_0 \geqslant 6 \times 10^{16}$ eV

A. V. Glushkov, V. B. Kosarev, I. T. Makarov, I. E. Sleptsov, and S. A. Filippov

Institute of Cosmophysical Studies and Aeronomy, Yakutsk Science Center, Siberian Branch of the Russian Academy of Sciences, 677891 Yakutsk, Russia

(Submitted 19 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 361–366 (25 March 1998)

Preliminary results are reported from an investigation of the temporal structure of the muon disk in extensive air showers (EASs) with primary energy $E_0 \geqslant 6 \times 10^{16}$ eV at distances 100–1500 m from the axis. The investigation is performed at the Yakutsk array using the large muon detector, which commenced operation in November 1995, with a planned area 184 m^2 and a detection threshold $E_\mu \approx 0.5 \cdot \sec\theta$ GeV. For $E_0 > 10^{18}$ eV it is found that the thickness of the muon disk tends to decrease. This requires substantial changes in our notions of the development of EASs. © *1998 American Institute of Physics.* [S0021-3640(98)00106-6]

PACS numbers: 96.40.Pq, 95.85.Ry

Muons with threshold energy $E_\mu \approx 1.0 \cdot \sec\theta$ GeV have been investigated at the Yakutsk array continually since 1974. A large amount of experimental data has been accumulated over this period of time, making it possible to investigate in detail the spatial distribution function (SDF) of muons in extensive air showers (EASs) with primary energy $E_0 \approx 10^{17} - 3 \times 10^{19}$ eV and zenith angles $\theta \leqslant 60°$ over a wide range of distances R from the EAS axis. In Refs. 1 and 2 it is shown that the form of the SDF for $E_0 \geqslant (3 - 5) \times 10^{18}$ eV is different from that at lower energies. Specifically, it becomes much steeper at distances $R > 400$ m.

To determine the reasons for such a difference in the SDF and to perform further investigations, a large muon detector (LMD) consisting of 92 scintillation counters each with an area of 2 m^2, arranged in six rows over an area of 26×12 m^2, was built at the Yakutsk array.[3] The detector is located 180 m from the center of the installation. The earthen shield gives a muon detection threshold of 0.5 sec θ GeV. Each counter operates independently and is equipped with a separate amplitude–time channel for measuring the number of particles and the arrival time of the first particle with accuracy of ~ 6 ns. Test observations on the LMD have been conducted since November 1995. Work on extending the detector over the entire detection area is now being completed.

We report the results of an analysis of the data obtained up to the end of 1997 using 30 counters with a total area of 60 m^2. The experimental data that have now been

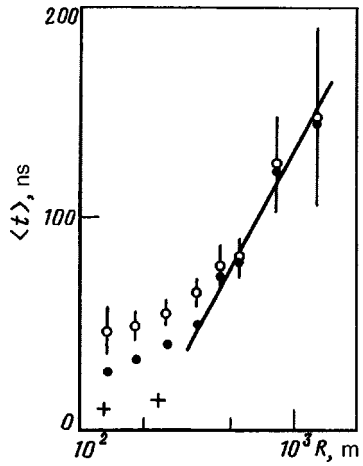

FIG. 1. Average delays $\langle t\rangle$ in the arrival of muons with energy $E_\mu \approx 0.5 \sec\theta$ GeV in showers with $E_0 = 10^{17}$–10^{18} eV and $\cos\theta \geqslant 0.95$ at different distances from the axis relative to the fastest muons in each shower: ○ — muon density $\rho_\mu \leqslant 1$ m^{-2}, + — $\rho_\mu \geqslant 4$ m^{-2}, ● — all densities.

accumulated not only confirm, to a high degree of accuracy, the results obtained in Ref. 3 in the energy range 10^{17}–10^{18} eV but also make it possible to present preliminary data on the structure of the muon disk for $E_0 > 10^{18}$ eV.

Figure 1 shows the average delay time $\langle t\rangle$ in the arrival of muons in EASs with $E_0 = 10^{17}$–10^{18} eV and $\cos\theta \geqslant 0.95$ at distances $R = 100$–1500 m relative to the very first muons in each shower. The open circles correspond to showers with muon densities $\rho_\mu \leqslant 1$ m^{-2}, the crosses correspond to $\rho_\mu \geqslant 4$ m^{-2}, and the filled circles correspond to all showers. One can see that when $\geqslant 8$ muons strike each detector the relative delays are very short and comparable to the accuracy of our time measurements. Events with $\rho_\mu \leqslant 1$ m^{-2} reflect the ''looser'' trailing edge.

The delay distributions all have an exponential form $\exp(-t/\lambda)$. The parameters λ for $\rho_\mu \leqslant 1$ m^{-2} are equal to $\langle t\rangle$ to within the limits of the experimental error. On this basis it is easy to obtain a relation for estimating the time T required to detect a relative fraction η of all muons:

$$T \approx -\langle t\rangle \ln(1-\eta) \text{ ns}. \tag{1}$$

It follows from Fig. 1 and Eq. (1) that 95% of all muons in EASs with $E_0 \leqslant 10^{18}$ eV at distances $R \leqslant 1000$ m arrive no later than 400 ns after the first-arriving muons.

Figure 2 shows $\langle t\rangle$ versus E_0 in showers with $\cos\theta \geqslant 0.8$ for $R = 630$ m, while Fig. 3 shows the variation of $\langle t\rangle$ at the same distance versus $\sec\theta$ in EASs with $E_0 = 10^{17}$–10^{18} eV. All values were obtained without detector selection according to muon density.

The data presented in Figs. 1–3 satisfy the relation

$$\langle t\rangle = a_0 + a_1 \cdot \log(E_0/10^{18}) + a_2 \cdot (1 - \sec\theta) + a_3 \cdot \log(R/600), \tag{2}$$

where $a_0 = 95 \pm 2$ ns, $a_1 = 7 \pm 1$ ns, $a_2 = 110 \pm 4$ ns, and $a_3 = 170 \pm 9$ ns. It is applicable in showers with $E_0 \approx 6 \times 10^{16}$–$10^{18}$ eV and $\theta \leqslant 45°$ at distances $R \approx 400 - 1500$ m.

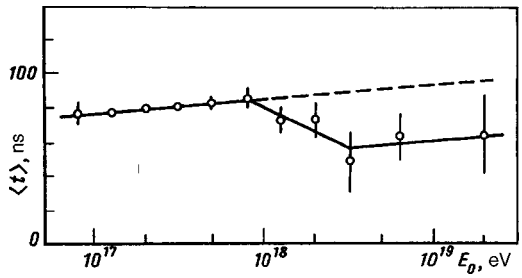

FIG. 2. $\langle t\rangle$ versus E_0 in showers with $\cos\theta \geqslant 0.8$ at $R=630$ m without detector selection according to muon density.

Data analysis showed that rate of increase of $\langle t\rangle$ with increasing E_0, $\partial\langle t\rangle/\partial\log(E_0)=a_1$ is the same for events with different θ and R (within the range of variation indicated above for these parameters). The increase in $\langle t\rangle$ is due to a displacement of the depth X_m of the maximum development of the shower toward the observation level X (for Yakutsk $X=1020\cdot\sec\theta$), i.e., a decrease in the distance $X-X_m$ up to the maximum development of the shower. This is clearly seen in Figs. 4a and 4b, which show the relative delays $t_a<t_b$ due to the degradation of the geometric muon-collection factor. For fixed E_0 the distance $X-X_m$ increases with the zenith angle, as result of which the difference of the muon delays decreases.

The rate of displacement of the maximum of the shower $\mathrm{ER}=\partial X_m/\partial\log(E_0)$ (ER — elongation rate) can be estimated from Eq. (2) as

$$\mathrm{ER}\approx a_1\times 1020/a_2\approx 65\ \mathrm{g\cdot cm^{-2}}. \tag{3}$$

Let us now examine the data for $E_0>10^{18}$ eV. According to Fig. 2 the measured values of $\langle t\rangle$ are all less than the expected values (dashed line) obtained by extrapolating the measured values from the region $E_0<10^{18}$ eV. In our view, two sections can be

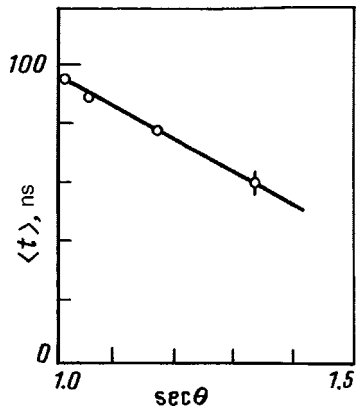

FIG. 3. $\langle t\rangle$ versus sec θ in showers with $E_0=10^{17}$–10^{18} eV at $R=630$ m without detector selection according to muon density.

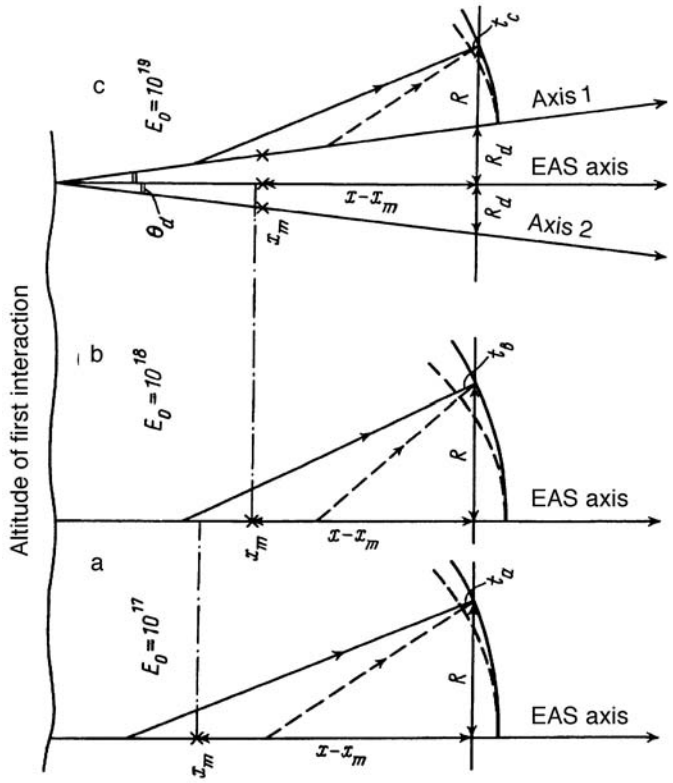

FIG. 4. Formation of relative delays in different EAS development schemes.

distinguished within the limits of the measurement error: The first section with $E_0 \approx (1-4)\times 10^{18}$ eV is a transitional section and the second section with $E_0 > 4\times 10^{18}$ eV satisfies

$$\langle t \rangle \approx 62 + a_1 \cdot \log(E_0/10^{19}) \text{ ns}. \tag{4}$$

At first glance it is difficult to understand this result from the physical point of view. If one attempts to explain it by rapid recession of the shower maximum, i.e., by a decrease in X_m, then the total number of electrons in these showers should decrease just as rapidly, which according to Refs. 1 and 2 does not happen.

We note that for $E_0 > 10^{18}$ eV many parameters of EASs show anomalies.[1,2,4,5] We believe that these anomalies all appear for the same reason. The anomalies are due not to the experimental procedure used at the Yakutsk array but rather to some new processes occurring in the development of an EAS.

Analysis of the data examined above confirms the following picture of the development of EASs, which we proposed earlier as one possibility.[1] For $E_0 > 10^{18}$ eV separate multicore showers appear. The relative fraction of these showers gradually increases and reaches 100% for $E_0 > (5-6)\times 10^{18}$ eV. Despite the fact that the overall particle balance and the dynamics of the longitudinal development of showers for $E_0 > (5-6)\times 10^{18}$ eV

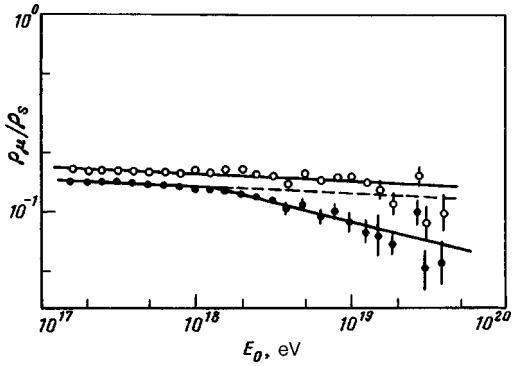

FIG. 5. Relative fraction of muons ($E_\mu \approx 1.0$ sec θ GeV) versus the total number of charged particles at distances $R=300$ m (●) and $R=600$ m (○) in showers with $\langle \cos\theta \rangle = 0.95$ as a function of primary energy.

most likely do not change appreciably, a substantial rearrangement of their transverse structure occurs. Figure 4c shows schematically how the thickness of the disk will change in this case. Since one of the axes of the EAS lies closer to the detector by the amount of its deflection R_d and makes the main contribution to the indications of the detector, the relative delays t_c will be less than t_a and t_b, though the shower maximum once again approaches the observation level.

Let us estimate R_d. According to Eqs. (2) and (4), a constant shift by ~ 30 ns occurs before and after our proposed change in the transverse structure of the EAS. From Fig. 1 we can see that $\langle t \rangle \approx 62$ ns correspond to $R \approx 430$ m. Hence we easily find $R_d \approx 630 - 430 = 200$ m.

We shall now present an additional experimental result which is not associated with the LMD data but which confirms the hypothesis stated above. Figure 5 (filled circles) shows as a function of E_0 the relative fraction ρ_μ/ρ_s of muons with threshold $E_\mu \approx 1.0 \cdot \sec\theta$ GeV among the total number of delayed particles in an EAS with $\langle \cos\theta \rangle = 0.95$ at a distance $R=300$ m from the axis. The open circles show the same data for $R=600$ m. We employed the shower sample used in Refs. 1 and 2. One can see that in the entire range of variation of E_0 the ratio $\rho_\mu(600)/\rho_s(600)$ decreases slightly with increasing primary energy without any appreciable deviations from a linear law. But the fraction $\rho_\mu(300)/\rho_s(300)$ behaves differently: For $E_0 > (1-2)\times 10^{18}$ eV it decreases more rapidly than it varied up to this point.

The reason can be easily understood with the aid of Fig. 6, which shows the SDFs of charged particles (filled circles) and muons (open circles) in an EAS with $E_0 = 10^{18}$ eV and $\langle \cos\theta \rangle = 0.95$. For $E_0 > 10^{18}$ eV it should be expected on the basis of what we have said above that all densities should gradually shift toward one of the closest axes of a multicore shower (roughly by ~ 200 m). Such a displacement will be almost unnoticeable at the periphery of the shower and will be large for $R < 300$ m, where the muon fraction decreases rapidly.

In our view, the multicore nature of EASs with giant transverse momenta most likely arises in the first nuclear-interaction event. As a result of this interaction, $m=2$, 3, 4 ... nucleons with energies $\sim E_0/m$ emerge from the same point at a angle θ_d with

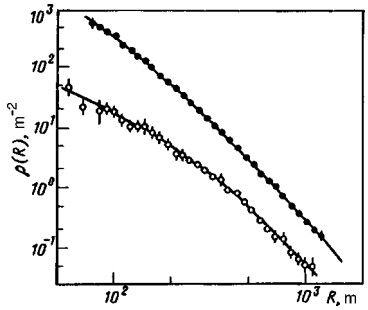

FIG. 6. Spatial distribution functions of charged particles (●) and muons (○) in showers with $\langle E_0\rangle = 10^{18}$ eV and $\langle\cos\theta\rangle = 0.95$.

respect to the direction of motion of the primary particle and these nucleons then form conventional EASs whose axes lie along a circle with radius $\sim R_d$.

The value of θ_d can be estimated as follows. Assume that the first nuclear interaction occurs at a depth ~ 50 g·cm^{-2} and using the relation

$$P = 1020 \cdot \sec\theta \cdot \exp(-H \cdot \cos\theta/6.85)\, g \cdot \mathrm{cm}^{-2} \tag{5}$$

between the distance from the level of observation H (km) for Yakutsk and the pressure P at this point (with $\cos\theta = 0.91$ for the showers studied), we find $H \approx 22$ km. The angle $\theta_d = \tan^{-1}(R_d/H) \approx 0.6°$ corresponds to this altitude.

Further observations will enable us to determine the structure of the muon disk for $E_0 > 10^{18}$ eV more accurately and report the results in subsequent publications. We also plan to make a direct search for multicore showers on the basis of the data which we now have, but this will be difficult to do at the Yakutsk installation with a 500–1000 m detector spacing. However, we now have an entire series of experimental facts that cannot be explained on the basis of the standard ideas about the development of EASs at extremely high energies. We offer as a possible hypothesis biaxial EASs and call upon other investigators to participate in solving this question by looking at the primary elementary nuclear-interaction event from new points of view.

The large muon detector was built as a prototype for the SHAL-1000 installation, designed as part of the State Scientific and Technical High-Energy Physics Program. This work was performed as part of the program of the Russian Ministry of Science supporting the Yakutsk EAS array (control No. 01-30), which is included in the ''List of unique scientific-research and experimental installations of national importance.''

[1] A. V. Glushkov, I. T. Makarov, E. S. Nikiforova *et al.*, Yad. Fiz. **58**, 1265 (1995) [Phys. At. Nucl. **58**, 1186 (1995)].
[2] A. V. Glushkov, I. T. Makarov, E. S. Nikiforova *et al.*, Astropart. Phys. **4**, 1274 (1995).
[3] B. N. Afanasiev, S. A. Filipov, A. V. Glushkov *et al.*, *Proceedings of the 25th International Cosmic Ray Conference (ICRC)*, Durban, South Africa, 28 July–8 August, 1997, Vol. 7, p. 261.
[4] A. V. Glushkov, M. I. Pravdin, and I. E. Sleptsov, Izv. Ross. Akad. Nauk, Ser. Fiz. **61**, 516 (1997).
[5] A. V. Glushkov, M. I. Pravdin, and I. Ye. Sleptsov, *Proceedings of the 25th International Cosmic Ray Conference (ICRC)*, Durban, South Africa, 28 July–8 August, 1997, Vol. 6, p. 233.

Translated by M. E. Alferieff

Dyon condensation and the Aharonov–Bohm effect

E. T. Akhmedov,[a] M. N. Chernodub, and M. I. Polikarpov
Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia

(Submitted 6 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 367–371 (25 March 1998)

We derive the string representation of the Abelian Higgs theory in which dyons are condensed. It happens that in such a representation the topological interaction exists in the expectation value of the Wilson loop. Due to this interaction the dynamics of the string spanned on the Wilson loop is nontrivial. © *1998 American Institute of Physics.* [S0021-3640(98)00206-0]

PACS numbers: 14.80.Mz, 14.80.Bn, 11.25.Db, 12.38.Bx, 03.65.Bz

The method of Abelian projections[1] is one of the popular approaches to the confinement problem[2] in non-Abelian gauge theories. Numerous computer simulations of the lattice gluodynamics in the Abelian projection (see, e.g., Refs. 3) show that the vacuum of gluodynamics behaves as a dual superconductor.[4] The key role in the dual superconductor model of the quantum chromodynamics (QCD) vacuum is played by Abelian monopoles.[1] In the Abelian projection quarks are electrically charged particles, and if monopoles are condensed, the dual Abrikosov string carrying the electric flux is formed between quark and antiquark. Because of a nonzero string tension the quarks are confined by the linear potential.

The Abelian monopole currents in gluodynamics are correlated[5] with (anti)instantons. For (anti-)self-dual fields the Abelian monopoles become Abelian dyons.[6] Moreover, in the vacuum of lattice gluodynamics the local correlator of the topological charge density and the product of the electric and magnetic currents is positive.[7] This means that the Abelian monopoles have electric charge. The sign of this electric charge coincides with the sign of the product of the magnetic charge and the topological charge density. Thus the infrared properties of QCD in the Abelian projection can be described by the Abelian Higgs model (AHM) in which dyons are condensed. The electric charge of the dyons fluctuates.[b)]

Note that there exists a model of the QCD vacuum[2] in which the *non-Abelian* dyons are responsible for the confinement. The non-Abelian dyons (as instantons) give rise to the Abelian dyons in the Abelian projection.

Below we study the properties of the Abrikosov–Nielsen–Olesen (ANO) strings in the Abelian model in which dyons are condensed. We consider Abelian dyons which have a constant electric charge. This model can be a zeroth approximation for the realistic effective model of the QCD vacuum in which the electric charge of the condensed dyons fluctuates.

We start with the following expression for the partition function in Euclidian space–time:[c)]

$$Z_{\rm dyon}=\int \mathcal{D}A_\mu \mathcal{D}B_\mu \mathcal{D}\Phi\ \exp\left\{-\int d^4x \mathcal{L}_{\rm dyon}(A,B,\Phi)\right\}, \tag{1}$$

where the dyon Lagrangian is:

$$\mathcal{L}_{\rm dyon}(A,B,\Phi)=\mathcal{L}_{\rm gauge}(A,B)+1/2\,|(\partial_\mu - ieA_\mu - igB_\mu)\Phi|^2+\lambda(|\Phi|^2-\eta^2)^2. \tag{2}$$

The field B_μ is the magnetic gauge potential, which is dual to the electric gauge potential A_μ, and Φ is the dyon field with the electric charge e and magnetic charge g. It was shown in Ref. 9 that it is possible to write a Lagrangian in which both fields A_μ and B_μ are regular:

$$\mathcal{L}_{\rm gauge}(A,B)=\frac{1}{2}[n\cdot(\partial\wedge A)]^2+1/2\,[n\cdot(\partial\wedge B)]^2+\frac{i}{2}[n\cdot(\partial\wedge A)]^\nu[n\cdot{}^*(\partial\wedge B)]_\nu$$

$$-\frac{i}{2}[n\cdot(\partial\wedge B)]^\nu[n\cdot{}^*(\partial\wedge A)]_\nu,$$

where $[a\cdot(b\wedge c)]^\nu\equiv a_\mu(b^\mu c^\nu - b^\nu c^\mu)$, $[a\cdot{}^*(b\wedge c)]^\nu\equiv a_\mu \epsilon^{\mu\nu\alpha\beta}(b_\alpha c_\beta)$ and n_μ is an arbitrary unit four-vector, $n^2=1$.

The partition function (1) can be represented as the partition function of the AHM. The Lagrangian $\mathcal{L}_{\rm gauge}$ is invariant under the linear transformation of the fields A and B (Ref. 9):

$$\begin{pmatrix}A\\B\end{pmatrix}\to\begin{pmatrix}A'\\B'\end{pmatrix}=\begin{pmatrix}\cos\upsilon & -\sin\upsilon\\ \sin\upsilon & \cos\upsilon\end{pmatrix}\begin{pmatrix}A\\B\end{pmatrix}, \tag{3}$$

where υ is an arbitrary constant. Applying this transformation with the parameter

$$\upsilon=-\arctan\frac{g}{e}, \tag{4}$$

to Eqs. (1) and (2) and integrating over the field A', we get the partition function of the AHM:[9]

$$Z_{\rm dyon}\propto Z_{\rm AHM}=\int \mathcal{D}B'_\mu \mathcal{D}\Phi\ \exp\left\{-\int d^4x \mathcal{L}_{\rm AHM}(B',\Phi)\right\},$$

$$\mathcal{L}_{\rm AHM}(B',\Phi)=\frac{1}{4}(\partial_{[\mu}B'_{\nu]})^2+\frac{1}{2}|(\partial_\mu - i\tilde{g}B'_\mu)\Phi|^2+\lambda(|\Phi|^2-\eta^2)^2, \tag{5}$$

the Higgs field Φ has the magnetic charge[d)] $\tilde{g}=\sqrt{e^2+g^2}$.

Consider the quantum average of the Wilson loop in the dyon theory (1):

$$\langle W_e^C\rangle_{\rm dyon}=\frac{1}{Z_{\rm dyon}}\int \mathcal{D}A_\mu \mathcal{D}B_\mu \mathcal{D}\Phi\ \exp\left\{-\int d^4x \mathcal{L}_{\rm dyon}(A,B,\Phi)\right\} W_e^C(A), \tag{6}$$

$$W_e^C(A)=\exp\left\{ie_0\int d^4x j_\mu A^\mu\right\},\quad j_\mu(x)=\oint_C d\tilde{x}_\mu \delta^{(4)}(x-\tilde{x}(\tau)),$$

which creates the particle with the electric charge e_0 on the world trajectory[e)] C.

Applying the transformation (3), (4) to the quantum average (6) and integrating over the field A'_μ we get:

$$\langle W_e^C \rangle_{\rm dyon} = \langle K^C_{(q_e,q_m)} \rangle_{\rm AHM}, \tag{7}$$

where the expectation value on the right-hand side of this equation is calculated in the AHM with the Lagrangian (5). The operator K is the product of the t'Hooft loop[10] H^C and the Wilson loop W^C:

$$K^C_{(q_e,q_m)}(B') = H^C_{q_e}(B') W^C_{q_m}(B'), \quad q_e = e_0 g/\tilde{g}, \quad q_m = e_0 e/\tilde{g}. \tag{8}$$

The operator $H^C_{q_e}$ is defined as follows:

$$H^C_{q_e}(B') = \exp\left\{ -\frac{1}{4} \int d^4x \left[\left(\partial_{[\mu} B'_{\nu]} - q_e \cdot \frac{1}{2} \varepsilon_{\mu\nu\alpha\beta} G^C_{\alpha\beta} \right)^2 - (\partial_{[\mu} B'_{\nu]})^2 \right] \right\}, \tag{9}$$

where the tensor $G^C_{\mu\nu} = (n\cdot\partial)^{-1} j_{[\mu} n_{\nu]}$ satisfies the relation $\partial_\nu G^C_{\mu\nu} = j_\mu$. The tensor $F^d_{\mu\nu} = q_e G^C_{\mu\nu}$ plays the role of the dual field strength tensor: $\partial_\nu F^d_{\mu\nu} = q_e j_\mu$. In the string representation of the AHM[11] the operator $H^C_{q_e}$ creates the string spanned on the loop C; this string carries the flux q_e.

The product K^C of the operators H^C and W^C creates the dyon loop with electric charge q_e and magnetic charge q_m on the world trajectory C in the AHM (5).

Now we discuss the string representation for the AHM (5).[11,12] At the center of the ANO strings the field $\Phi = |\Phi| e^{i\theta}$ vanishes, Im Φ = Re Φ = 0, and the phase θ is singular on the two-dimensional surfaces which are world sheets of the ANO strings. The measure of the integration over the fields Φ can be rewritten as follows: $\mathcal{D}\Phi = \text{const} \cdot \mathcal{D}|\Phi|^2 \mathcal{D}\theta$. The integral $\int \mathcal{D}\theta$ contains the integration over functions which are singular on two-dimensional manifolds, and we subdivide θ into a regular part θ^r and a singular part θ^s: $\theta = \theta^r + \theta^s$; here θ^s is defined by:

$$\partial_{[\mu,}\partial_{\nu]} \theta^s(x,\tilde{x}) = 2\pi \epsilon_{\mu\nu\alpha\beta} \Sigma_{\alpha\beta}(x,\tilde{x}),$$

$$\Sigma_{\alpha\beta}(x,\tilde{x}) = \int_\Sigma d^2\sigma \epsilon^{ab} \partial_a \tilde{x}_\alpha \partial_b \tilde{x}_\beta \delta^{(4)}[x - \tilde{x}(\sigma)], \quad \partial_a = \frac{\partial}{\partial \sigma^a}, \tag{10}$$

where the vector function $\tilde{x}_\mu$ is the position of the string, Σ is the collection of all closed surfaces, $\sigma = (\sigma_1, \sigma_2)$ is the parametrization of the string surface; the measure $\mathcal{D}\theta$ can be decomposed as follows: $\mathcal{D}\theta = \mathcal{D}\theta^r \mathcal{D}\theta^s$.

For simplicity we consider below the London limit of the AHM ($\lambda \to \infty$). In this limit the radial part of the field Φ is fixed everywhere except for the centers of the ANO strings. All the expressions below can be generalized to the case of arbitrary λ; this leads to an additional functional integral over the radial part of $|\Phi|$.

Performing the transformations as in Refs. 12 and 11, we get the following string theory for the quantum average (6) of the Wilson loop:

$$\langle W_e^C \rangle_{\rm dyon} = \frac{1}{Z_{str}} \int [\mathcal{D}\tilde{x}] \cdot J(\tilde{x}) \cdot \exp\left\{ -\int d^4x \int d^4y \left[\frac{q_m^2}{2} j_\mu(x) D_m(x-y) j_\mu(y) \right. \right.$$

$$+\pi i\zeta\cdot j_\mu(x)D_m(x-y)\partial_\nu\epsilon_{\mu\nu\alpha\beta}(\Sigma_{\alpha\beta}(y)+\mathcal{N}G^C_{\mu\nu}(y))+\pi^2\eta^2(\Sigma_{\mu\nu}(x)$$

$$+\mathcal{N}G^C_{\mu\nu}(x))D_m(x-y)(\Sigma_{\mu\nu}(y)+\mathcal{N}G^C_{\mu\nu}(y))\Big]+2\pi i\zeta\mathbb{L}(\Sigma,C)\Big\}, \tag{11}$$

where

$$\mathcal{N}=\frac{e_0 g}{2\pi}, \quad \zeta=\frac{e_0 e}{\tilde{g}^2}=\frac{e_0 e}{e^2+g^2}, \tag{12}$$

$D_m(x)$ is the scalar Yukawa propagator, $(\Delta+m^2)D_m(x)=\delta^{(4)}(x)$, and $m^2=2\tilde{g}^2\eta^2$ is the mass of the dual gauge boson (B').

The measure $[\mathcal{D}\tilde{x}_\mu]$ assumes both integration over all possible positions and summation over all topologies of the string's world sheets Σ; $J(\tilde{x})$ is the Jacobian of the transformation from the field θ^s to the string position $\tilde{x}_\mu$. The Jacobian $J(\tilde{x})$ was estimated in Ref. 11 for string with spherical or disc topology.

The first three terms in the exponent in Eq. (11) describe the short-range interaction and the self-interaction of the ANO strings and dyon–antidyon pair through the exchange of the massive gauge boson. The constant $\mathcal{N}$ which appears in these terms has a physical meaning. It is equal to the number of the elementary fluxes in the string which connects the dyon–antidyon pair introduced by the operator K of Eq. (8). By definition, $\mathcal{N}=q_e/\Psi_0$, where q_e is equal to the total electric flux from the dyon and $\Psi_0=2\pi/\tilde{g}$ is the flux carried by the elementary string in the AHM (5). Since this number of the elementary fluxes $\mathcal{N}$ must be an integer, we get the charge quantization rule: $e_0 g\in 2\pi\mathcal{N}$, $\mathcal{N}\in\mathbb{Z}$ (Ref. 9).

The last term in Eq. (11),

$$\mathbb{L}(\Sigma,C)=\frac{1}{4\pi^2}\int d^4x\int d^4y\,\epsilon_{\mu\nu\alpha\beta}\Sigma_{\mu\nu}(x)j_\alpha(y)\frac{(x-y)_\beta}{|x-y|^4}$$

is the linking number of the string world sheet Σ and the trajectory C of the dyon. This formula represents the long-range interaction which describes the *dual* four-dimensional analogue[13] of the dual Aharonov–Bohm effect: strings correspond to electric solenoids which scatter the magnetic charges of Abelian dyons. This linking number term is important for the infrared properties of the theory, since it may induce an additional long-range potential between quark and antiquark.[14] It also leads to nontrivial commutation relations between different operators in the theory.[11]

The authors are grateful to F. V. Gubarev, Yu. A. Simonov, T. Suzuki and A. V. Zaharov for helpful discussions. This work was supported by the grants INTAS-96-370, INTAS-RFBR-95-0681, RFBR-96-02-17230a and RFBR-96-15-96740.

[a)]e–mail: akhmedov@vxitep.itep.ru

[b)]Note that according to the Schwinger quantization rule the electric charge e of the dyon is not fixed while magnetic charge g is quantized: $e_0 g\in 2\pi\mathcal{N}$, e_0 is an elementary electric charge of an external electric particle, see Eq. (6).

[c)]The theory with $e=0$ (monopoles are condensed) has been investigated as an effective Abelian theory of QCD in Refs. 8.

[d)]We call B'_μ the dual gauge field (thus Φ carries magnetic charge) since we consider (5) as the Abelian effective model of the QCD vacuum. Actually, after the transformation (3) this is a matter of convention.
[e)]This average corresponds to the quark Wilson loop if we consider (1) as an effective theory of QCD.

[1] G. 't Hooft, Nucl. Phys. B **190**[FS3], 455 (1981).
[2] Yu. A. Simonov, Phys. Usp. **39**, 313 (1996).
[3] T. Suzuki, Nucl. Phys. B (Proc. Suppl.) **30**, 176 (1993);M. I. Polikarpov, Nucl. Phys. B (Proc. Suppl.) **53**, 134 (1997); M. N. Chernodub and M. I. Polikarpov, preprint ITEP-TH-55/97, http://xxx.lanl.gov/abs/hep-th/9710205.
[4] G. 't Hooft, in *High Energy Physics*, edited by M. Zichichi (Editrice Compositori, Bolognia, 1976);S. Mandelstam, Phys. Rep. **23C**, 245 (1976).
[5] O. Miyamura and S. Origuchi, *RCNP Confinement 1995*, Osaka, Japan, March 22–26, 1995, p. 137; M. N. Chernodub, F. V. Gubarev, JETP Lett. **62**, 100 (1995); A. Hart and M. Teper, Phys. Lett. **371**, 261 (1996); S. Thurner, M. Feurstein, H. Markum *et al.*, Phys. Rev. D **54**, 3457 (1996); R. C. Brower, K. N. Orginos, Chung-I Tan, Phys. Rev. D **55**, 6313 (1997); M. Fukushima, S. Sasaki, H. Suganuma *et al.*, Phys. Lett. B **399**, 141 (1997).
[6] G. Schierholz, *RCNP Confinement 1995*, Osaka, Japan, March 22–26, 1995, p. 96, http://xxx.lanl.gov/abs/hep-lat/9506033; V. Bornyakov, G. Schierholz, Phys. Lett. **384**, 190 (1996).
[7] M. N. Chernodub, F. V. Gubarev, and M. I. Polikarpov, preprint ITEP-TH-44/97, http://xxx.lanl.gov/abs/hep-lat/9709039; preprint ITEP-TH-70/97, http://xxx.lanl.gov/abs/hep-lat/9801010.
[8] S. Maedan and T. Suzuki, Prog. Theor. Phys. **81**, 229 (1989); H. Suganuma, S. Sasaki, and H. Toki, Nucl. Phys. B **435**, 207 (1995).
[9] D. Zwanziger, Phys. Rev. D **3**, 880 (1971).
[10] G. 't Hooft, Nucl. Phys. B **138**, 1 (1978); Nucl. Phys. B **153**, 141 (1979).
[11] E. Akhmedov, M. Chernodub, M. Polikarpov *et al.*, Phys. Rev. D **53**, 2087 (1996).
[12] M. I. Polikarpov, U.-J. Wiese and M. A. Zubkov, Phys. Lett. B **309**, 133 (1993); P. Orland, Nucl. Phys. B **428**, 221 (1994); M. Sato and S. Yahikozawa, Nucl. Phys. B **436**, 100 (1995); E. T. Akhmedov, JETP Lett. **64**, 82 (1996).
[13] M. G. Alford and F. Wilczek, Phys. Rev. Lett. **62**, 1071 (1989); M. G. Alford, J. March-Russel, and F. Wilczek, Nucl. Phys. B **337**, 695 (1990); J. Preskill and L. M. Krauss, Nucl. Phys. B **341**, 50 (1990); L. M. Kraus and F. Wilczek, Phys. Rev. Lett. **62**, 1221 (1989).
[14] F. A. Bais, A. Morozov, M. de Wild Propitius, Phys. Rev. Lett. **71**, 2383 (1993); M. N. Chernodub, F. V. Gubarev, and M. I. Polikarpov, http://xxx.lanl.gov/abs/hep-lat/9704021, to be published in Phys. Lett. B; Nucl. Phys. Proc. Suppl. **53**, 581 (1997); http://xxx.lanl.gov/abs/hep-lat/9607045.

Published in English in the original Russian journal. Edited by Steve Torstveit.

Large interference-induced increase in the intensity of quantum transitions in triple-barrier structures

E. I. Golant and A. B. Pashkovskiĭ
State Research and Production Corporation ''Istok,'' 141120 Fryazino, Moscow Region, Russia

(Submitted 23 December 1997; resubmitted 30 January 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 372–377 (25 March 1998)

A solution of the nonstationary Schrödinger equation describing the resonance interaction of electrons with a weak rf field is found for asymmetric triple-barrier resonance-tunneling structures with thin, high barriers, and an expression is obtained for the active weak-signal conductivity of such structures. It is found that in a number of cases the probability of quantum resonance transitions from an upper to a lower level can increase sharply when the lower levels are shifted relative to one another. © *1998 American Institute of Physics.*
[S0021-3640(98)00306-5]

PACS numbers: 42.55.Px, 73.20.Dx

In Ref. 1 it was shown that in principle triple-barrier quantum-size semiconductor structures can be used to convert electron energy into rf energy. In recent years, however, investigations of electron transitions between the levels of such structures has become especially urgent as result of the progress made in building THz-range unipolar quantum cascade lasers[2,3] in which radiative transitions occur in each of serially coupled triple-barrier heterostructures separated by superlattice sections which act as electron injector and thermalizer. It is ordinarily assumed that the electrons incident on the structures (see Fig. 1) occupy the upper level of the first double-barrier structure, relax to the lower level, and escape from the interaction region through a resonance level in the second double-barrier structure. It should be noted that an absolute majority of theoretical results and all experimental results concerning lasers operating on intersubband transitions have been obtained for the successive tunneling regime, where electron in a quantum well undergoes intense collisions with (mainly optical) phonons, destroying the coherence of the electron wave function. The possibility of obtaining a population inversion on the working levels and lasing in a quantum cascade laser can be explained by the characteristic features of intra- and intersubband quantum transitions with the participation of optical phonons.[4] In such a regime the position of the levels in the second well for which electrons escape as fast as possible from the lower working level is very important for obtaining a population inversion on the levels in the first quantum well. Two variants were investigated: a) The levels are coincident and rapid escape is accomplished by resonance tunneling[1] and b) the lower level in the first quantum well is raised relative to the lower level of the second well by an amount equal to the energy of an optical phonon

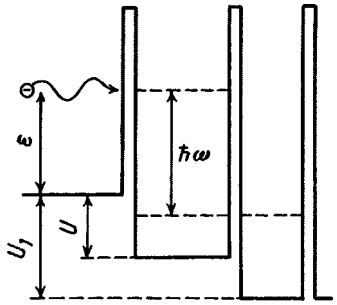

FIG. 1. Schematic band diagram of the triple-barrier structure studied.

and rapid escape is accomplished by a nonradiative diagonal resonance transition of the electrons.[3] Only the b version has been implemented experimentally.

In Ref. 2 it is estimated that the number of radiative transitions in a quantum cascade laser with successive tunneling of electrons is of the order of 10^{-3}–10^{-4} times the total number of intersubband transitions. This makes it necessary to pass very high currents through the structure in order to obtain a negative rf conductivity sufficient for lasing. At the same time, the idea of a laser with purely ballistic (coherent) electron transport was advanced some time ago.[4] The practical realization of this idea would open up fundamentally new possibilities both for investigating electronic processes accompanying resonance tunneling[5,6] and for improving lasers.[7,8] In the present letter we wish to show that in a triple-barrier working structure of a quantum cascade laser with coherent electron tunneling interference effects can be used to increase the intensity and quantum efficiency of the transitions substantially. The optimal relative arrangement of the levels in neighboring quantum wells turns out to be completely different from that used in successive-tunneling structures.

Let us consider an asymmetric triple-barrier structure with thin (δ-like) barriers to which a uniform rf electric field varying in time as $\mathcal{E}\cos\omega t = E(e^{i\omega t}+e^{-i\omega t})$, $\mathcal{E}=2E$, is applied. For definiteness we shall assume that a monoenergetic electron flux is incident from the left onto the Kth resonance level in the first double-barrier structure, the frequency of the rf field corresponds to transitions to the Lth level of the same structure (see Fig. 1), and one of the resonance levels in the second double-barrier structure lies near the Lth level of the first structure. Then the nonstationary Schrödinger equation is

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m^*}\frac{\partial^2\psi}{\partial x^2} + H(x)\psi + H(x,t)\psi, \tag{1}$$

$$H(x) = -U(\theta(x)-\theta(x-a)) - U_1\theta(x-a) + \alpha\delta(x)\psi + \alpha\rho\delta(x-a) + \alpha\gamma\delta(x-a-l),$$

$$H(x,t) = -qE\cdot[x(\theta(x)-\theta(x-a-l)) + a\theta(x-a-l)](e^{i\omega t}+e^{-i\omega t}).$$

Here q and m^* are the electron charge and mass, $\alpha=\varphi_b b$, φ_b and b are the height and width of the first barrier, $\theta(x)$ is a unit step function, γ and ρ are numerical coefficients, U and U_1 are the offsets of the conduction-band bottom at the barriers (see Fig. 1), and a and l are the distances between the barriers. The unperturbed electron wave function ψ_0, normalized to one electron, is

$$\psi_0(x)=\begin{cases}\exp[ik_0x]+D_0\ \exp[-ik_0x], & x<0,\\ A_0\ \sin(kx)+B_0\ \cos(kx), & 0<x<a,\\ Z_0\ \sin(k_1(x-a))+W_0\ \cos(k_1(x-a)), & a<x<a+l,\\ C_0\ \exp[ik_1(x-a-l)], & x>a+l,\end{cases} \tag{2}$$

where the energy of the electrons incident on the structure is ϵ and their wave numbers are

$$k_0=(2m^*\epsilon/\hbar^2)^{1/2},\quad k=(2m^*(\epsilon+U)/\hbar^2)^{1/2},\quad k_1=(2m^*(\epsilon+U_1)/\hbar^2)^{1/2}.$$

For sufficiently strong barriers ($y\gg k$)

$$D_0\approx\frac{i\rho^2k_0\ \sin(k_1l)+k_1\ \cos(k_1l)}{i\rho^2k_0\ \sin(k_1l)-k_1\ \cos(k_1l)},\quad B_0\approx\frac{2i\rho^2k_0\ \sin(k_1l)}{i\rho^2k_0\ \sin(k_1l)-k_1\ \cos(k_1l)}, \tag{3}$$

$$A_0\approx\frac{y}{k}B_0,\quad y=\frac{2m^*\alpha}{\hbar^2}.$$

In the weak-signal approximation the correction ψ_1 to the ground-state wave function[9] is $\psi_1=\psi_+(x)e^{-i(\omega_0+\omega)t}+\psi_-(x)e^{-i(\omega_0-\omega)t}$ ($\omega_0=\epsilon/\hbar$). In our case ψ_- is (ψ_+ is small)

$$\psi_-(x)=\begin{cases}D_-\ \exp(\kappa x), & x<0,\\ A_-\ \sin(k_-x)+B_-\cos(k_-x)+\varphi_-(x), & 0<x<a,\\ Z_-\ \sin[k_{1-}(x-a)]+W_-\cos[k_{1-}(x-a)]+\chi_-(x), & a<x<a+l,\\ C_-\ \exp[ik_{1-}(x-a)]+P_-\exp[ik_1(x-a)], & x>a+l,\end{cases} \tag{4}$$

where

$$\kappa=(2m^*(\hbar\omega-\epsilon)/\hbar^2)^{1/2},\quad k_-=(2m^*(\epsilon+U-\hbar\omega)/\hbar^2)^{1/2},$$

$$k_-=(2m^*(\epsilon+U_1-\hbar\omega)/\hbar^2)^{1/2},\quad P_-=+\frac{qEa}{\hbar\omega}\psi_0(a+l),$$

and $\varphi_-(x)$ and $\chi_-(x)$ are particular solutions of the equation

$$\hbar(\omega_0-\omega)\psi_-(x)=-\frac{\hbar^2}{2m^*}\frac{\partial^2\psi_-}{\partial x^2}-qEx\psi_0(x), \tag{5}$$

and in the present case (see Ref. 9)

$$\varphi_-(x),\quad \chi_-(x)=+\frac{qEx}{\hbar\omega}\psi_0(x)+\frac{qE}{m^*\omega^2}\psi_0'(x). \tag{6}$$

The system of equations for the coefficients A_-, B_-, C_-, D_-, Z_-, and W_- has the following matrix form:

$$\begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ -\kappa-y & k_- & 0 & 0 & 0 & 0 \\ 0 & \sin(k_-a) & \cos(k_-a) & 0 & -1 & 0 \\ 0 & -k_-\cos(k_-a) & k_-\sin(k_-a) & k_{1-} & -\rho y & 0 \\ 0 & 0 & 0 & \sin(k_{1-}a) & \cos(k_{1-}a) & -1 \\ 0 & 0 & 0 & -k_{1-}\cos(k_{1-}a) & k_{1-}\sin(k_{1-}a) & ik_{1-}\gamma y \end{pmatrix}$$

$$\times \begin{pmatrix} D_- \\ A_- \\ B_- \\ Z_- \\ W_- \\ C_- \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{pmatrix} = \begin{pmatrix} \varphi_-(0) \\ -\varphi'_-(0) \\ \chi_-(a)-\varphi_-(a) \\ \rho y\chi_-(a)-\chi'_-(a)+\varphi'_-(a) \\ P_- -\chi_-(a+l) \\ (\gamma y - ik_1)P_- +\chi'_-(a+l) \end{pmatrix}. \tag{7}$$

For $y \gg k_-$, k_{1-}, by analogy with a double-barrier structure,[6] we seek the conditions for resonance transitions from the upper to the lower level when

$$|\cos k_-a| \approx 1, \quad |\cos k_{1-}a| \approx 1, \quad |\sin k_-a| \ll 1, \quad |\sin k_{1-}a| \ll 1. \tag{8}$$

For definiteness we shall study two levels with the same parity, and we seek the resonance conditions (the conditions under which the determinant (7) assumes its minimum value) in the form

$$\cos k_-a \approx 1, \quad \cos k_{1-}a \approx 1,$$

$$\sin k_-a \approx -\frac{k_-}{y}\cdot\left(\beta+\frac{\Delta\beta}{y}\right), \quad \sin k_{1-}a \approx -\frac{k_{1-}}{y}\cdot\left(\xi+\frac{\Delta\xi}{y}\right). \tag{9}$$

Substituting expressions (9) into the system of equations (7) it is easy to show that the modulus of the determinant of this system becomes small (it does not contain the large parameter y) when

$$\beta = \frac{1+\rho+\gamma-\gamma\xi(1+\rho)}{\rho+\gamma-\rho\gamma\xi}, \quad \Delta\beta+\Delta\xi\frac{\gamma^2}{(\rho+\gamma-\rho\gamma\xi)^2} = -\kappa. \tag{10}$$

The determinant itself then equals

$$\Delta = -\frac{ik_-k_{1-}^2}{\rho+\gamma-\rho\gamma\xi}, \tag{11}$$

and we obtain for the coefficient C_- determining the probability of transitions to the lower level

$$C_- \approx \frac{4qEy^2}{m^*\omega^2 k_{1-}}\frac{\rho^2 k_0\sin^2(k_1l)}{i\rho^2 k_0\sin(k_1l)-k_1\cos(k_1l)}(\rho+\gamma-\rho\gamma\xi), \tag{12}$$

while the conductivity of the structure on which the monoenergetic electron flux with density n is incident is given by

$$\sigma = 8\sigma_s \frac{k_-}{k_{1-}} \frac{\rho^4 k_0^2 \sin^2(k_1 l)}{\rho^4 k_0^2 \sin^2(k_1 l) + k_1^2 \cos^2(k_1 l)} (\rho + \gamma - \rho\gamma\xi)^2, \tag{13}$$

where

$$\sigma_s \approx -\frac{8q^2 m^* \alpha^4 n}{\pi L \hbar^6 \omega^3} [1 - (-1)^{K-L}] \tag{14}$$

is the conductivity of a symmetric double-barrier structure with the barrier strength equal to α, where electrons with density n move from the Kth level to the Lth level.[9]

Two fundamental differences from the case of double-barrier structures are immediately evident:

1) The resonance levels can lie both above and below the corresponding levels in a well with infinite walls (always below in double-barrier structures).

2) As expected, the probability of transitions from the upper to the lower level (proportional to $k_{1-}|C_-|^2$ and σ) depends on the relative arrangement of the lower levels. But, for the relative arrangement of the barriers determined by the condition (10), instead of decreasing when the levels are shifted relative to one another (and, correspondingly, the parameters β and ξ change) the transition probability, on the contrary, increases!

The increase in the transition probability can be very large. After all, for the conditions of applicability of this calculation to be satisfied it is sufficient that $|\xi| \ll y/k_{1-}$. In real structures the condition $|\xi| \gg 1$ can be easily satisfied. Thus, in a GaAs-based structure (the electron effective mass $m^* \approx 0.067 m_0$) with AlAs barriers ($\varphi_B \approx 1.04$ eV) of thickness $b \approx 20$ Å and interbarrier separation $a \approx 100$ Å, y for the first resonance level is more than an order of magnitude greater than k_- and the condition (8) holds even when $|\xi| \approx 5$, which in turn results in a more than order of magnitude higher transition probability than in cases $\xi = \beta$ with identical barriers.

Thus, a new and quite unexpected effect appears in triple-barrier structures — when the lower resonance levels in neighboring wells are shifted relative to one another, interference effects can sharply increase the probability of resonance transitions from an upper to a lower level as compared with the transition probability which obtains when these levels are coincident.

It should be noted, however, that the transition probability depends strongly on how the shift of the levels relative to one another occurs. Thus, as $\xi \to (\rho + \gamma)/\rho\gamma$, $|\beta| \to \infty$ while the transition probability decreases appreciably (of course, if the condition (8) still holds).

These effects can be qualitatively explained as follows. It is known that the resonance conductivity σ (probability of transitions between levels) of a double-barrier structure similar to that studied above but without the third barrier ($\gamma = 0$) depends strongly on the strength of the second barrier.[10] In our notation it is given by

$$\sigma \approx \sigma_s \frac{8\rho^6 k_0^2 k_-}{(\rho^2 k_0 + k_1)^2 k_{1-}}. \tag{15}$$

One can see that as the parameter ρ increases, when the structure essentially becomes simply a quantum well, the conductivity increases without bound. By analogy, our triple-barrier structure can be regarded as a double-barrier structure in which the second double-barrier structure plays the role of the second barrier. When the levels in two neighboring wells are coincident, the transmittance of the second structure is maximum (the strength of the second barrier in the double-barrier analog is minimum) and therefore the probability of transitions between levels is also low. As the parameter ξ varies (the position of the third barrier or conduction-band bottom in the second well varies), the resonance level in the second well shifts relative to the level in the first well and its transmittance drops (the strength of the second barrier in the double-barrier analog increases), and therefore the probability of transitions between levels also increases. It is interesting to note that $\beta \to (1+\rho)/\rho$ as $|\xi| \to \infty$, i.e., the resonance conditions are satisfied in the double-barrier structure.

For structural parameters such that $\xi \to (\rho+\gamma)/\rho\gamma$ resonance transitions occur at energies which are substantially shifted from the resonance level in the double-barrier structure ($|\beta|$ approaches infinity and not $(1+\rho)/\rho$), which, correspondingly, causes the the probability of these transitions to decrease.

It should be noted that in the limiting case of purely coherent tunneling studied here the proposed structure with sufficiently strong barriers is completely nontransmitting for elastic (no energy change) electron tunneling. For this reason, every electron transmitted through the structure must emit a photon and therefore the quantum efficiency of the process is close to one, and in addition, in contrast to the process studied in Ref. 6, for any amplitude of the field. Of course, under real experimental conditions only some electrons tunnel coherently, and the question of whether or not this part can be made sufficiently large requires additional investigation. In any case, the construction of active structures of quantum cascade lasers taking account of the effect described above will make it possible not only to determine the relative fraction of electrons that tunnel coherently but also to increase the quantum efficiency and correspondingly decrease the threshold current of these devices.

This work was supported by the Russian Fund for Fundamental Research, Project 97-02-16652, and the Scientific Council as part (Project 97-1094) of the program ''Physics of Solid-State Nanostructures.''

[1] A. Kastalsky, V. J. Goldman, and J. H. Abeles, Appl. Phys. Lett. **59**, 2636 (1991).
[2] J. Faist, F. Capasso, C. Sirtori *et al.*, Appl. Phys. Lett. **64**, 1144 (1994).
[3] J. Faist, F. Capasso, C. Sirtori *et al.*, Phys. Rev. Lett. **76**, 411 (1996).
[4] E. I. Golant, A. B. Pashkovskiĭ, and A. S. Tager, Pis'ma Zh. Tekh. Fiz. **20**(21), 74 (1994) [Tech. Phys. Lett. **20**, 886 (1994)].
[5] M. Yu. Sumetskiĭ and M. L. Fel'tshin, JETP Lett. **53**, 24 (1991).
[6] E. I. Golant and A. B. Pashkovskiĭ, JETP Lett. **63**, 590 (1996).
[7] E. I. Golant and A. B. Pashkovskiĭ, Zh. Éksp. Teor. Fiz. **112**, 237 (1997) [JETP **85**, 130 (1997)].
[8] V. F. Elesin, Zh. Éksp. Teor. Fiz. **112**, 483 (1997) [JETP **85**, 264 (1997)].
[9] A. B. Pashkovskiĭ, Zh. Éksp. Teor. Fiz. **109**, 1779 (1996) [JETP **82**, 959 (1996)].
[10] E. I. Golant and A. B. Pashkovskiĭ, Fiz. Tekh. Poluprovodn. **31**, 1077 (1997) [Semiconductors **31**, 921 (1997)].

Translated by M. E. Alferieff

Interaction of a modulated electron beam with a magnetoactive plasma

A. V. Kostrov[a)] and M. V. Starodubtsev
Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

C. Krafft[b)] and G. Matthieussent
Laboratory of Physics of Gases and Plasmas, Université Paris–Sud, CNRS, 91405 Orsay, France

A. S. Volokitin
Institute of Terrestrial Magnetism, the Ionosphere, and Propagation of Radio Waves, Russian Academy of Sciences, 142092 Moscow, Russia

(Submitted 6 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 378–382 (25 March 1998)

Experimental results concerning the interaction of a modulated electron beam with a magnetoactive plasma in the whistler frequency range are reported. It was shown experimentally that when a beam is injected into the plasma, waves can be generated by two possible mechanisms: Cherenkov emission of whistlers by the modulated beam, and transition radiation from the beam injection point. In the case of weak beam currents ($N_b/N_0 \ll 10^{-4}$) the Cherenkov resonance radiation is more than an order of magnitude stronger than the transition radiation; the Cherenkov emission efficiency decreases at high beam currents. The transformation of the distribution function of the beam is investigated for the case of weak beam currents. It is shown that in the case of the Cherenkov interaction with whistlers the beam is retarded and the beam distribution function becomes wider and acquires a plateau region. [S0021-3640(98)00406-X]

PACS numbers: 52.40.Mj, 41.60.Bq

The possibility of using modulated electron beams as an emitter of whistler-range waves has been discussed in recent years in application to active experiments in space.[1] The theoretical works[2,3] concerning this problem focus mainly on the analysis of the Cherenkov emission of electromagnetic waves in an infinite plasma and the first laboratory experiment[4] demonstrated the possibility of such emission of whistlers. In the present work it was shown experimentally that besides Cherenkov radiation there exists in the entire volume a nonresonance radiation from the point where the modulated beam is injected into the plasma (transition radiation).

The experiments were performed on the apparatus shown schematically in Fig. 1. The plasma source consisted of a heated oxide cathode and grid between which an accelerating voltage pulse was applied with repetition frequency 5 Hz. The resulting

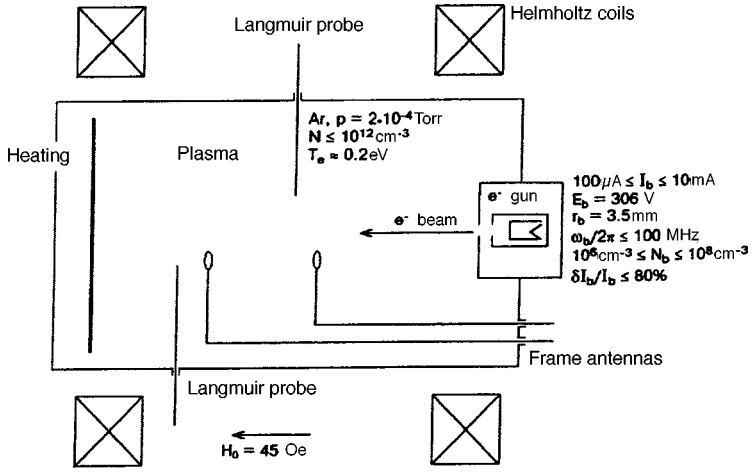

FIG. 1. Diagram of the experimental arrangement.

accelerated-electron flux ionized the neutral gas (argon at pressure 5×10^{-4} Torr). As a result, a 70 cm long and 50 cm in diameter quasi-one-dimensional plasma column was produced in the vacuum chamber. The initial plasma density N_e was of the order of 10^{12} cm^{-3} and then decreased with a characteristic time of 1 ms (see Fig. 3a below). The experiments were performed in the decomposing plasma regime. An electron temperature $T_c\approx0.2$ eV was established at the plasma decomposition stage. Two Helmholtz coils (coil diameter 2 mm, coil separation -1.5 m) produced a uniform magnetic field. The magnetic field was equal to $H_0\approx45$ Oe.

The electron gun used to produce a density-modulated beam consisted of a triode with a grid anode. The accelerating voltage was equal to 300 V and the beam current could be varied from 10 μA to 10 mA, which corresponds to density N_b from 10^6 cm^{-3} to 10^8 cm^{-3}. The beam diameter was equal to 7 mm. The electron beam density was modulated with an rf voltage applied to the grid of the electron gun. The modulation frequency was $f\sim100$ MHz, which corresponded to the whistler frequency range; the degree of modulation was of the order of 80%. The modulated beam was injected continually throughout the entire period of plasma decomposition. Only longitudinal beam injection into the plasma was used in the experiments, i.e., the pitch angle was always equal to zero.

The energy distribution function of the electron beam was investigated with a multigrid analyzer. Two electrostatically shielded frame antennas, each of which could be moved in two directions — along the axis of the apparatus and in a radial direction — were used to study the structure of the fields excited in the plasma volume.

Analysis of the conditions for Cherenkov emission $\omega=k_\parallel V_b$ shows that excitation of the characteristic waves of the system with longitudinal wavelength $\lambda_\parallel=2\pi/k_\parallel$ is possible if the plasma density is less than a critical value (for $\omega_{pe}\gg\omega_{He}>\omega_0>\sqrt{(\omega_{He}\omega_{Hi})}$), determined from the condition $\omega_{pe}^2<\omega_{pekp}^2=k_\parallel^2c^2(\omega_{He}-\omega)/\omega$, where ω_{pe} is the electron plasma frequency, ω_{He} and ω_{Hi} are, respectively, the electron and ion gyrofrequencies, $\lambda_\parallel$ and $k_\parallel$ are, respectively, the wavelength and longitudinal wave num-

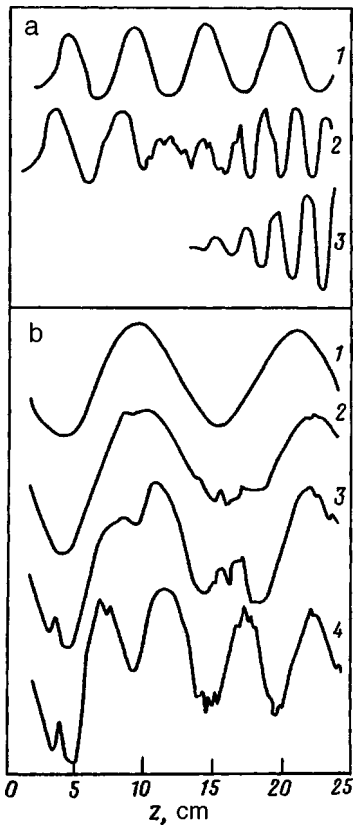

FIG. 2. Interferometric measurements of waves in plasma (Cherenkov resonance conditions are not satisfied): a) *1* — surface wave of a modulated electron beam, *2* — emission from the beam absorption point (a surface wave of the beam is also visible), *3* — wave emitted by the frame antenna under the same conditions; b) *4* — wave emitted by the frame antenna; *2*, *3*, and *4* — emission from the beam injection point for different values of the beam current. A surface wave of the beam can be see as the current increases.

ber of the propagating wave, and c is the speed of light in vacuum. Outside the resonance region of the plasma this wave is a surface wave and is localized near the beam at a distance of the order of c/ω_{pe}.

The interferometric investigations performed attest to the fact that for plasma density above a critical value a traveling surface mode exists near the beam (Fig. 2a, curve *1*), and a quasi-longitudinal whistler, which can be excited both near the point of injection of the modulated beam into the plasma and near the point of absorption of the beam, is also detected (Fig. 2a, curve *2*; Fig. 2b, curve *4*). The observed nonresonance radiation near the beam injection point and the beam absorption point should be attributed to transition-type radiation, while the effective emitter responsible for this emission should be represented, to a first approximation, as an electric dipole with moment directed along the axis of the apparatus.

The Cherenkov resonance radiation of the modulated electron beam (Fig. 3b) was reliably detected at low beam currents $J_0 < 100$ μA ($N_b/N_0 < 10^{-4}$). In this case the

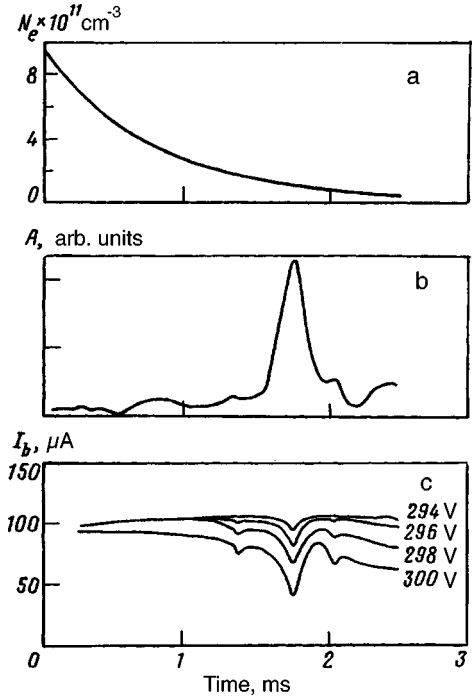

FIG. 3. a — Time dependence of the plasma density; b — amplitude of the H_r component of the rf field in the plasma (the maximum corresponds to Cherenkov resonance conditions being satisfied). The beam current equals 100 μA ($N_b/N_0 \approx 10^{-5}$); c — signal from the beam-particle analyzer for different values of the cutoff voltage on the analyzing grid.

amplitude of the Cherenkov radiation was more than an order of magnitude greater than the amplitude of the whistler transition radiation. As the beam current increased, the efficiency of the Cherenkov emission decreased compared with that of whistler emission from the injection and absorption points.

For plasma density $N_0 \approx 10^{11}$ cm^{-3} the longitudinal wavelength of the whistler equals the spatial beam-modulation length $2\pi V_b/\omega_0$. Under these conditions the frame antenna excites an electromagnetic wave with the same longitudinal wavelength $\lambda_\parallel \cong 2\pi V_B/\omega_0$.

Analysis of the energy distribution function of the electron beam shows that efficient retardation of the beam was observed when the Cherenkov resonance conditions were satisfied. The characteristic oscillograms of the beam current at the multigrid analyzer with different cutoff voltages are displayed in Fig. 3c. One can see that a decrease of the flux of electrons reaching the analyzer collector signifies retardation of the electron beam as it interacts with the plasma. The beam distribution functions in resonance and nonresonance situations are shown in Fig. 4a and 4b, respectively. The characteristic width of the beam distribution function in the absence of resonance is of the order of 6 eV. When the resonance conditions are satisfied, the beam is retarded on the average, its width increases to 12 eV, and a plateau forms in the distribution function.

In summary, our experiments established that there exist two different mechanisms

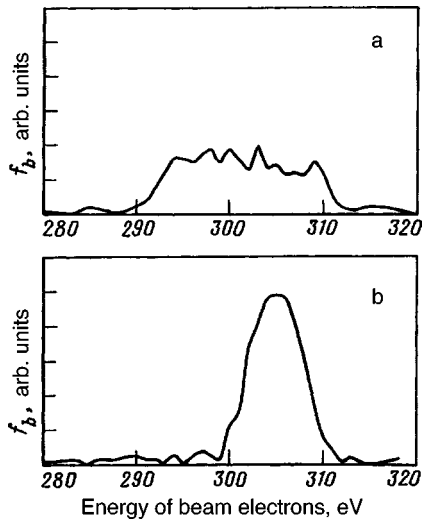

FIG. 4. Distribution function of the beam (with current $I_b=100$ μA ($N_b/N_0\approx10^{-5}$)): a — under Cherenkov resonance conditions (the average energy of the beam particles is approximately 300 V, the width is 17 V, and a plateau region is present); b — off the Cherenkov resonance (the average energy of the beam particles is 306 V, and the width is 7 V).

leading to whistler generation when a modulated electron beam is injected into a plasma: Cherenkov resonance radiation $\omega_0=k_\parallel V_b$ and nonresonance transition radiation, which exists in a wide range of plasma densities. As the beam current increases ($N_b/N_0>10^{-4}$), the Cherenkov radiation efficiency decreases as a result of broadening of the beam distribution function accompanying transition radiation of electromagnetic waves at the beam injection point.

[a)]e-mail: kstr@appl.sci-nnov.ru
[b)]e-mail: catherine.krafft@lpgp.u-psud.fr

[1]V. N. Oraevski and Y. Y. Rouzin, *Project APEX. Scientific Aims Modelling and Technics*, Science, Moscow, 1992.
[2]K. J. Harker and P. M. Banks, Planet. Space Sci. **33**, 953 (1985).
[3]K. J. Harker and P. M. Banks, Planet. Space Sci. **35**, 11 (1987).
[4]C. Krafft, P. Thevenet, G. Matthieussent *et al.*, Phys. Rev. Lett. **2**, 649 (1994).

Translated by M. E. Alferieff

Short-wavelength asymptotic behavior of the Burgers turbulence spectrum

S. L. Shalimov
Joint Institute of Earth Physics, 123810 Moscow, Russia

(Submitted 15 December 1997; resubmitted 13 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 383–386 (25 March 1998)

Using a third-order structure function, an asymptotic expression is obtained for the stationary energy spectrum in the dissipative region of Burgers turbulence excited by a random external force. It is shown that in contrast to the case of turbulence described by a homogeneous Burgers equation, the spectrum contains a parameter characterizing energy transfer into the small-scale region. © *1998 American Institute of Physics.* [S0021-3640(98)00506-4]

PACS numbers: 47.27.−i

An important example of stationary turbulence is Burgers turbulence (BT) excited by a random external force. Burgers turbulence is described by the inhomogeneous one-dimensional Burgers equation

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=\nu\frac{\partial^2 u}{\partial x^2}+f(x,t), \tag{1}$$

where $f(x,t)$ is a Gaussian field with correlation function

$$\langle f(x,t)f(x+r,t+s)\rangle = F(r)\,\delta(s).$$

In a comparatively recent numerical experiment[1] the spectrum in the dissipative region of BT was investigated by solving Eq. (1). The spectrum obtained was compared with the well-known (see, for example, Ref. 2) asymptotic expression for the BT spectrum in this region

$$E(k)=\frac{2\pi\nu^2}{L}\exp\left(-\frac{\pi\nu}{V}k\right), \tag{2}$$

where L is the outer scale, ν is the kinematic viscosity, and V is the velocity jump at the shock wave. It was shown[1] that as the wave number in the dissipative region increases, the numerical solution decreases more rapidly than the asymptotic expression (2).

We note that the latter formula was obtained from the homogeneous Burgers equation. The small-scale spectrum (2) in this case is characterized by large-scale parameters but does not contain any information about energy transfer along the spectrum. In contrast to the shape of the spectrum (2), in the present work an asymptotic expression is obtained analytically for the stationary spectrum in the dissipative region of BT, de-

scribed by Eq. (1) containing a parameter characterizing energy transfer to small scales. The presence of this parameter makes it possible to interpret the more rapid decrease of the spectrum in the dissipative region.

From Eq. (1) follows a relation, similar to the Kármán–Howarth equation from the theory of hydrodynamic turbulence,[3,4] for the correlation function of statistically homogeneous stationary turbulence:

$$\frac{1}{6}\frac{\partial S_3(r)}{\partial r}-\nu\frac{\partial^2 S_2(r)}{\partial r^2}=-(\langle f(x)u(x+r)\rangle+\langle f(x+r)u(x)\rangle), \tag{3}$$

where $S_n(r)=\langle[u(x+r)-u(x)]^n\rangle$. The right-hand side of Eq. (3) can be expressed in terms of the correlation function of the external force using the Furutsu–Novikov formula[5]

$$\langle fu[f]\rangle=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\langle f(x,t)f(x_1,t_1)\rangle\left\langle\frac{\delta u[f]}{\delta f(x_1,t_1)dx_1dt_1}\right\rangle dx_1dt_1,$$

and the fact that

$$\frac{\delta u(x,t)}{\delta f(x_1,t)dx_1dt}=\frac{1}{2}\delta(x_1-x).$$

As a result, we obtain after integrating Eq. (3)

$$S_3=6\nu\frac{\partial S_2}{\partial r}-6\int_0^r F(r')dr', \tag{4}$$

where the correlation function of the external force can be written in the form[5]

$$F(r)=2\epsilon\phi(r/L),$$

where L is the outer scale of turbulence, $\phi(r/L)$ is a dimensionless function such that in the limit $L\to\infty$: $\phi(r/L)\to\phi(0)=1$, $\epsilon\equiv\nu\langle(\partial u/\partial r)^2\rangle=\langle fu\rangle$ — the average rate of energy dissipation, determining the statistical properties of the random external force f.

Expanding the function $\phi(r/L)$ in a series and taking into account the evenness, we obtain from Eq. (4) for $r\ll L$

$$S_3=6\nu\frac{\partial S_2}{\partial r}-12\epsilon r\left(1-\frac{1}{3}\left(\frac{r}{L}\right)^2+O\left(\frac{r}{L}\right)^4\right). \tag{5}$$

Using the parameter $S=\langle(\partial u/\partial r)^3\rangle/\langle(\partial u/\partial r)^2\rangle^{3/2}$ — the asymmetry of the velocity field in the limit $r\to 0$, we have

$$S_3=-S\left(\frac{\epsilon}{\nu}\right)^{3/2}r^3+O(r^5). \tag{6}$$

We find from Eqs. (5) and (6)

$$S_2=-\frac{S}{24\nu}\left(\frac{\epsilon}{\nu}\right)^{3/2}r^4-\frac{\epsilon}{6\nu L^2}r^4+\left(\frac{\epsilon}{\nu}\right)r^2+O(r^6), \tag{7}$$

if the asymmetry can be taken as constant, a condition which can be satisfied for stationary turbulence excited by a random external force if $r \ll \delta = \nu/\langle (u^2) \rangle^{1/2}$, where δ is the characteristic scale of the shock wave (i.e., in the present case — the dissipation scale).[2] We note that the asymmetry of the velocity field corresponds to the presence of a nonlinear interaction between modes, which leads to the formation of inertial and dissipative regions of the spectrum.[3] In the absence of asymmetry the spectrum will consist of delta functions and its derivatives in the source (this follows from Eq. (5) for sufficiently large scale L and the relation between S_2 and the energy spectrum; see below).

The structure function S_2 is related with the correlation function according to the formula $S_2 = 2[B(0) - B(r)]$, whence using Eq. (7) we obtain in dimensionless form

$$\hat{B}(\hat{r}) = \frac{B(r)}{B(0)} = 1 - \hat{S}_2 = 1 - \beta^2 \left(\hat{r}^2 - \frac{\hat{r}^4}{\alpha^2} \right) + O(\hat{r}^6), \tag{8}$$

where $\hat{r} = r/\delta$, $\hat{S}_2 \equiv S_2/2u_0^2$, $u_0^2/2$ is the average energy, $L = u_0^3/\epsilon$, $1/\alpha^2 = (S/24) \times (\delta^2 \epsilon^{1/2}/\nu^{3/2}) + (1/6)(\delta/L)^2$, and $\beta^2 = (1/2)(\delta^2 \epsilon^{1/3}/L^{2/3}\nu)$.

The normalized energy spectrum in turn is related with the correlation function by the formula

$$\hat{E}(\hat{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\hat{k}\hat{r}) \hat{B}(\hat{r}) d\hat{r}. \tag{9}$$

Here the function $\hat{B}$ in the integrand is represented by the sign-alternating series (8). Applying the transformation $w = \hat{r}^2/(\gamma + \hat{r}^2)$, $\hat{r}^2 = \gamma w/(1-w)$, where γ is a constant, or using a Padé approximation[6] it can be shown that the convergence radius of this series is determined by the distance to the nearest singular point, which is located on the negative real axis $\hat{r}^2$.

Since for $\hat{r} > 1$ (which corresponds to the inertial interval) we have[2] $\hat{S}_2 \sim C\hat{r}$, joining the asymptotic expansions (see, for example, Refs. 3 and 7) makes it possible to treat the expansion of $\hat{S}_2$ in Eq. (8) as an inner expansion of the function $\hat{S}_2 = \hat{r}^2/(1 + 2\hat{r}^2/\alpha^2)^{1/2}$. We note that joining of a higher-order the higher-order terms of the series (8) are required. However, since for real $\hat{r} = O(\alpha)$ the approximation obtained is a completely definite function and since we are interested only in the far dissipative region of the spectrum, it can be assumed that refining the approximation will not appreciably affect the form of the spectrum because the spectrum varies quite sharply in the direction of high wave numbers.

Therefore, using the last expression for $\hat{S}_2$, we obtain

$$\hat{B}(\hat{r}) = 1 - \beta^2 \frac{\hat{r}^2}{(1 + 2\hat{r}^2/\alpha^2)^{1/2}}.$$

Confining attention to the dissipative region of the spectrum and neglecting the influence of the source, we can write

$$\hat{E}(\hat{k}) \sim \beta^2 \frac{1}{2\pi} \frac{\partial^2}{\partial \hat{k}^2} \int_{-\infty}^{\infty} \frac{\exp(-i\hat{k}\hat{r})}{(1 + 2\hat{r}^2/\alpha^2)^{1/2}} d\hat{r} = \frac{\alpha^3 \beta^2}{2^{3/2}\pi} \left(K_0(k_1) + \frac{1}{k_1} K_1(k_1) \right), \tag{10}$$

where $k_1=\alpha\hat{k}/\sqrt{2}$ and K_ν is a modified Bessel function of order ν. For $k_1\gg 1$ we obtain the desired asymptotic spectrum

$$\hat{E}(\hat{k})\sim\frac{\alpha^3\beta^2\sqrt{\pi}}{4}\left(\frac{1}{k_1^{1/2}}+\frac{1}{k_1^{3/2}}\right)\exp(-k_1), \tag{11}$$

or

$$E(k)\sim\frac{\sqrt{\pi}\alpha^3\delta^2\epsilon}{8\nu}\left(\frac{1}{(\alpha\delta k/\sqrt{2})^{1/2}}+\frac{1}{(\alpha\delta k/\sqrt{2})^{3/2}}\right)\exp(-\alpha\delta k/\sqrt{2}), \tag{12}$$

where $\alpha=1/\sqrt{(S/24)(\delta^2\epsilon^{1/2}/\nu^{3/2})+(1/6)(\delta/L)^2}$, and δ is the dissipation scale.

One can see from Eq. (12) that for fluctuations with sufficiently small-scale the ratio of the argument in of the exponential in the spectrum (12) to the argument of the exponential in the asymptotic expression (2) is of the order of $(LV/S^2\nu)^{1/4}>1$ (ordinarily, $S\sim 0.5$ (Ref. 3)). This indicates that, in the first place, the shape of the spectrum depends fundamentally on the asymmetry and, in the second place, the spectrum drops off more rapidly with increasing k and makes it possible to interpret the results of the numerical experiment.[1]

In summary, it can be concluded that in the dissipative region the shape of the spectrum of homogeneous stationary Burgers turbulence with a random source is determined by the asymmetry parameter of the velocity field.

[1] S. S. Girimaji and Ye. Zhou, Phys. Lett. A **202**, 279 (1995).
[2] P. G. Saffman, in *Topics in Nonlinear Physics, Lectures on Homogeneous Turbulence,* edited by N. J. Zabusky, Springer, Berlin, 1968.
[3] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics*, Vol. 2, MIT Press, Cambridge, Mass., 1975 [Russian original, Gidrometeoizdat, St. Petersburg, 1996].
[4] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, New York [Russian original, Nauka, Moscow, 1986].
[5] E. A. Novikov, Zh. Éksp. Teor. Fiz. **47**, 1919 (1964) [Sov. Phys. JETP **20**, 1290 (1965)].
[6] G. A. Baker and P. Graves-Morris, *Padé Approximants*, Addison-Wesley, New York, 1981.
[7] L. Sirovich, L. Smith, and V. Yakhot, Phys. Rev. Lett. **72**, 344 (1994).
[8] J. D. Cole, *Perturbation Methods in Applied Mathematics*, Blaisdell, London, 1968.

Translated by M. E. Alferieff

Accurate method for determining tilt bias angles in thin films of nematic liquid crystals

S. V. Yablonskiĭ, A. S. Mikhaĭlov, S. P. Palto, S. G. Yudin, and S. V. Yakovlev
Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russian

G. Durand
Laboratoire de Physique des Solides, Université Paris-Sud, 91405 Orsay, France

(Submitted 20 October 1997; resubmitted 4 March 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 387–392 (25 March 1998)

We have developed a new method for measuring tilt bias angles in spatially uniform and nonuniform thin films of nematic liquid crystals. The method employs modulation ellipsometry, based on the use of an exponentially decaying light wave to probe the boundary layer. Oscillations of the director of the liquid crystal, which are induced by the flexoelectric torque, are excited with an external periodic field. A periodic variation of the ellipticity of the light wave reflected from the interface is detected at both the first and second harmonics of the exciting electric field. When these two Fourier components of the electrooptic response are known, it is possible to calculate both the tilt bias angle θ_0 of the director and the dynamic deviation $\delta\theta$ of the tilt bias angle. The angles θ_0 and $\delta\theta$ measured by this method on the surface of an electrode (ITO) and on the surface of a ferroelectric film (a copolymer of vinylidene fluoride and trifluoroethylene), oriented in a corona discharge, were equal to $\theta_0 = 5.1°$, $\delta\theta = 0.5°$ and $\theta_0 = 89°$, $\delta\theta = 0.06°$, respectively. © *1998 American Institute of Physics.*
[S0021-3640(98)00606-9]

PACS numbers: 68.15.+e, 61.30.Gd, 07.60.Fs

1. INTRODUCTION

The angle between the optic axis of a liquid crystal (LC) and a boundary surface (tilt bias angle) strongly influences the working parameters of liquid-crystal devices. For example, the production of a tilt bias angle in nematic twist cells makes it possible to avoid undesirable effects due to light scattering from inversion domain walls which form during cell operation.[1] The production of a corresponding tilt bias angle makes it possible to obtain a uniform deformation of the LC in the entire sample and thereby improve the optical properties of the liquid-crystal cell.[2,3] For this reason, to optimize the working characteristics of displays it is important to have an accurate method for measuring the the tilt bias in a wide range of angles.

Some methods for measuring tilt bias are presented in Refs. 4–10. Methods such as interferometry,[4] conoscopy,[5] total internal reflection,[6] magnetic null,[7] capacitive,[7] and

crystal rotation[7] require cells with a uniform orientation of the liquid crystal. Since these methods probe the entire volume of a liquid crystal, the resulting tilt bias angle is the result of averaging over the entire thickness of the LC layer. The substantial limitations due to the averaging over the entire volume of the LC are eliminated in methods based on analysis of a light beam reflected from a thin near-surface layer of the liquid crystal.[8–10] In this case a LC layer with a thickness of the order of the wavelength of the light wave is probed. However, for example, the method based on Brewster angle measurements[8] does not work at a metal–insulator interface and therefore it is not suitable for standard LC cells. Ellipsometric analysis of Fresnel reflection[9] requires a very complicated fitting procedure.

This letter presents a new version of modulation ellipsometry for measuring tilt bias angles by means of excitation of oscillations of the director of a LC as a result of a linear interaction of an ac electric field with the flexoelectric polarization of the medium.

2. FUNDAMENTALS OF THE THEORY OF THE METHOD

Several years ago we developed a method of modulation ellipsometry[10] based on probing of a near-surface layer with an exponentially decaying optical wave arising in a LC under conditions of total reflection of light from the interface between the LC and heavy-flint glass.

Refractive-index modulation near the surface of a homeoplanar cell changes the ellipticity of the reflected beam and the corresponding phase shift σ between the extraordinary (p) and ordinary (s) waves

$$\sigma=\sigma_p-\sigma_s, \tag{1}$$

where

$$\sigma_p=2\tan^{-1}\frac{N\sqrt{N^2\sin^2\varphi-n_{\text{eff}}^2}}{n_e n_0\cos\varphi}, \quad \sigma_s=2\tan^{-1}\frac{\sqrt{N^2\sin^2\varphi-n_0^2}}{N\cos\varphi},$$

n_0, n_e, and N are, respectively, the refractive indices of the liquid crystal and prism, $\varphi>\varphi_I$ is the angle of incidence (Fig. 1), and φ_I is the angle of total internal reflection. For a uniform orientation of the director $\mathbf{n}(\cos\theta,\ \sin\theta, 0)$ the effective refractive index is determined as $n_{\text{eff}}=\sqrt{n_0^2\cos^2\theta+n_e^2\sin^2\theta}$, $\theta=\theta_0+\delta\theta$, where $\delta\theta=\delta\theta_m\sin\omega t$ the sinusoidal inclination from the static tilt bias angle θ_0, is produced by acoustic excitation. Fourier analysis of the experimentally measured time dependence of the phase shift $\sigma(t)$ makes it possible to calculate the tilt bias angle θ_0 and the amplitude $\delta\theta_m$ of the director oscillations as follows:

$$2\delta(t)=\sigma(t)=A\Delta n_{\text{eff}}, \tag{2}$$

where $\delta(t)$ is the variable part of the azimuthal angle (Fig. 1),

$$A=\left.\frac{\partial(\sigma_p-\sigma_s)}{\partial n_{\text{eff}}}\right|_{n_{\text{eff}}=n_0,n_e}, \quad \Delta n_{\text{eff}}=\frac{n_e^2-n_0^2}{2n_0}\theta^2 \ \text{ or } \ \Delta n_{\text{eff}}=\frac{n_e^2-n_0^2}{2n_e}\theta^2$$

in accordance with whether θ_0 is close to the X axis ($\theta=\theta_0+\delta\theta_m\sin\omega t\cong 0$) or the Y axis ($\theta=\pi/2-\theta_0-\delta\theta_m\sin\omega t\cong 0$). Therefore, assuming θ_0 to be small, we have

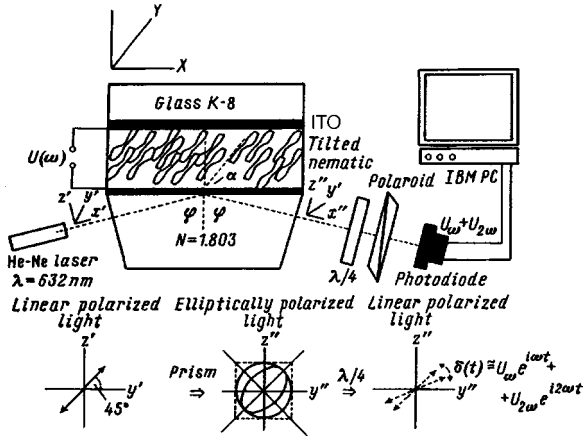

FIG. 1. Scheme for measuring the tilt bias angle θ_0 by modulation ellipsometry.

$$2\delta(t)=\sigma(t)=A\frac{n_e^2-n_0^2}{n_0}\left(\frac{2\theta_0^2+\delta_m^2}{4}+\theta_0\delta_m\sin\omega t-\frac{(\delta\theta_m)^2}{4}\cos 2\omega t\right). \tag{3}$$

For the amplitudes of the corresponding Fourier components of the electrooptic response measured at the fundamental and doubled frequencies of the modulating field we have

$$U(\omega)=A\frac{n_e^2-n_0^2}{n_0}\theta_0\delta\theta_m\left(\frac{\Delta U}{2\alpha}\right), \tag{4}$$

$$U(2\omega)=-A\frac{n_e^2-n_0^2}{4n_0}(\delta\theta_m)^2\left(\frac{\Delta U}{2\alpha}\right), \tag{5}$$

where ΔU is the change produced in the voltage at the photodetector output by a rotation of the analyzer by the angle α. Therefore we have for the tilt bias angle

$$\theta_0=\sqrt{\frac{\alpha}{A(n_e^2-n_0^2)n_0^{-1}}}\,x\,\frac{U(\omega)}{\sqrt{U(2\omega)\Delta U}}. \tag{6}$$

The last equation is the basic equation for tilt bias angle measurements and for estimating the accuracy of the method.

The key ingredient in our approach to measuring tilt bias angles is excitation of the sinusoidal oscillations of the director by means of the linear flexoelectric interaction of the nematic LC with an external electric field. To accomplish this the dielectric contribution to the torque (which is due to the quadratic effect) must be eliminated by using a compensated nematic LC with zero dielectric anisotropy. We have already found an expression for the deflection angle $\delta\theta$ for the case of zero dielectric anisotropy[11,12]

$$\delta\theta=\frac{(e_1+e_3)E_m\sin 2\theta_0}{2K[L^{-1}+(1+i)\sqrt{\omega\eta/2K}+2i\eta_s\omega]}\exp(i\omega t)\exp(iqz), \tag{7}$$

where $q=(-1+i)(\omega\eta/2K)^{1/2}$, $L=K/W$ is the extrapolation length determined by the surface interaction energy W, e_1 and e_3 are flexoelectric constants, K is a combination of

the transverse and longitudinal elastic moduli, η is the effective viscosity (approximately equal to the rotational viscosity), and η_s is the surface viscosity. We employ a sinusoidal field $E=E_m \exp(i\omega t)$, while the initial orientation of the director corresponds to the range $0 \leqslant \theta_0 \leqslant \pi/2$.

The order of magnitude of the amplitude of the flexoelectric deformation at the surface can be estimated from Eq. (7). Substituting into Eq. (7) the actual values $e=3 \times 10^{-4}$ dyn$^{1/2}$, $E_m=30$ cgs esu, $K=10^{-6}$ dyn, $L=0.1$ μm, and $z=0$, we obtain

$$\delta\theta \cong \theta_0/10. \tag{8}$$

3. EXPERIMENTAL ASPECTS OF MODULATION ELLIPSOMETRY

A diagram of the experiment is displayed in Fig. 1. The main component is the liquid-crystal cell, consisting of a high index prism ($N=1.803$) and a glass plate, which are coated with ITO electrodes. The cell is placed on an optical bench, while the induced angular oscillations of the director of the LC at the prism surface are detected by the ellipsometry method.

The principle of optical modulation ellipsometry, which we are employing, is illustrated at the bottom of Fig. 1 with the aid of a diagram of successive transformations of the polarization of the light wave.

4. RESULTS AND DISCUSSION

All measurements were performed for a compensated mixture ($\Delta\varepsilon \approx 0.02$). The tilt bias angle was produced by unidirectional rubbing.[13] A planar-oriented LC cell consisted of a prism and a glass plate with an antiparallel direction of rubbing. In this case, pure ITO electrodes were rubbed in. In a homeoplanar cell both ITO substrates were coated with thin films of a ferroelectric copolymer vinylidene fluoride and trifluoroethylene (the molar ratio of vinylidene fluoride and trifluoroethylene was equal to 70:30), after which only the film on the surface of the prism was rubbed in one direction and polarized with a corona discharge with a potential difference of -9 kV between the tungsten needle and the substrate at temperature 100 °C. The tilt bias angle was always measured at room temperature one day after the cell was filled.

To make the correct choice of the magnitude of the modulating field the amplitudes of the first and second harmonics of the electrooptic response were measured as a function of the applied ac voltage. The results are presented in Fig. 2. The linearity of the functions $U_\omega(V)$ and $\sqrt{U_{2\omega}}(V)$ agrees with Eqs. (4) and (5) in the interval from 3 to 10 V, determining the possible values of the external modulating voltage. The characteristic features in the electrooptic response that can be seen in the plots at voltages above 10 V reflect the field-induced instability in the LC.

The Fourier spectra of the electrooptic response with an 8 V modulating voltage are shown in Fig. 3. From these spectra we have for the desired Fourier components $U(498 \text{ Hz}) = 5.6 \times 10^{-4}$ V, $U(996 \text{ Hz}) = 1.27 \times 10^{-5}$ V and $U(742 \text{ Hz}) = 8.04 \times 10^{-4}$ V, $U(1484 \text{ Hz}) = 1.25 \times 10^{-5}$ V. We measured the Fourier harmonics for different frequencies of the modulating voltage only to show that the result of the measurements does not depend on the modulation frequency in the band determined by the LC and in our case equalling 1500 Hz. From Eqs. (4)–(6) for the measured amplitudes of the Fourier components, taking $n_0=1.55$, $n_e=1.77$ (Ref. 14), $N=1.803$, and $\varphi=80°$, we obtain tilt bias

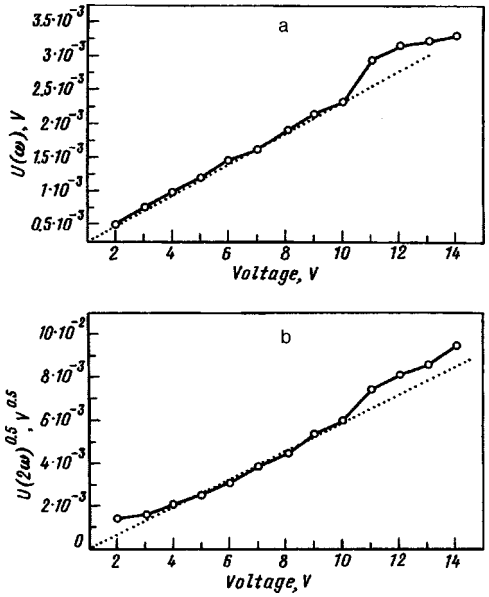

FIG. 2. First (a) and second (b) harmonics of the electrooptic response as a function of the applied ac voltage.

angles and deflections of the tilt bias angle $\theta_0 = 5.1° \pm 0.2°$, $\delta\theta = 0.5° \pm 0.05°$ and $\theta_0 = 89° \pm 0.2°$, $\delta\theta = 0.06° \pm 0.01°$ for planar and homeoplanar cells, respectively. Estimates of the same angles by the method of total internal reflection (TIR) give the values $\theta_0 = 6°$ and $\theta_0 = 89.5°$. Rough TIR measurements were also necessary in order to make the correct choice of one of the approximations in Eq. (2) (two variants are possible: $n_{\text{eff}} = n_0$ or $n_{\text{eff}} = n_e$) that is necessary to determine the derivative A. The approximation employed (Eqs. (2)–(6)) is quite accurate for angles $\theta_0 \leqslant 10°$ and $80° \leqslant \theta_0 \leqslant 90°$. It is easy to see from Eq. (6) that in the limit of small and large angles the accuracy of the method is determined by the accuracy of the measurements of the amplitude of the second harmonic, which can be estimated from the Fourier spectra and equals approximately 5%. There are not fundamental obstacles for measurements at intermediate angles; it is only necessary to make an additional analysis of the initial equation (1) and the data processing procedure is more complicated. It is also important to estimate the errors due to the dielectric interaction. From the balance of volume and surface torques we have

$$\delta\theta_B \approx \varepsilon_a E^2/4\pi\gamma\omega, \tag{9}$$

$$K\nabla\theta_s = K\delta\theta_s/L, \tag{10}$$

where the indices B and s refer, respectively, to the volume and surface contributions, ε_a is the dielectric anisotropy, and γ is the rotational viscosity. Taking into consideration the first spatial Fourier component of the director distribution, we have

$$K(\pi/d)\varepsilon_a E^2/4\pi\gamma\omega = K\delta\theta_s/L \tag{11}$$

or

$$\delta\theta_s \approx \varepsilon_a E^2 L/4\gamma\omega d. \tag{12}$$

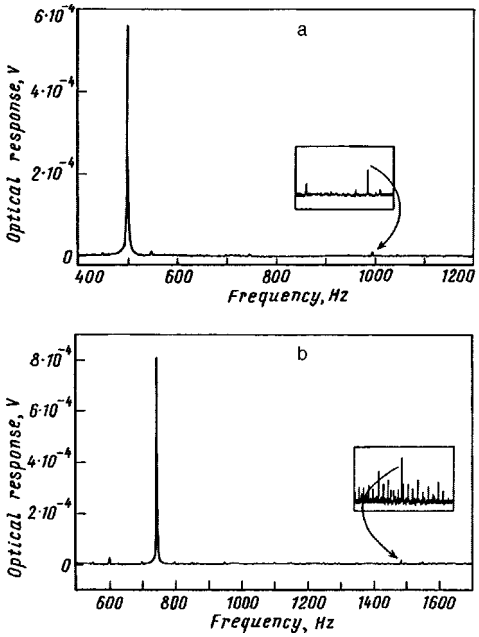

FIG. 3. Fourier spectra of the electrooptic response for: a) inclined spatially uniform orientation: $U=8$ V, $f=498$ Hz, $U(\omega)=5.6\times10^{-4}$ V, $U(2\omega)=1.27\times10^{-5}$ V, and b) hybrid orientation: $U=8$ V, $f=742$ Hz, $U(\omega)=8.04\times10^{-4}$ V, $U(2\omega)=1.25\times10^{-5}$ V .

From Eq. (7) we obtain the deflection of the angle due to the flexoelectric torque

$$\delta\theta_f \approx eEL/K. \tag{13}$$

It is obvious that the ratio of these angles should be much smaller than the ratio of the amplitudes of the second and first harmonics of the electrooptic response:

$$\frac{\delta\theta_s}{\delta\theta_f} \approx \frac{\varepsilon_a EK}{4\gamma\omega de} \ll \frac{U(2\omega)}{U(\omega)} = \frac{1}{40} \tag{14}$$

and therefore

$$\varepsilon_a \ll \gamma\omega de/10EK. \tag{15}$$

After substituting the corresponding values $\gamma=1$ P, $\omega=2\pi x1000$ Hz, $E=20$ cgs esu, $d=10^{-3}$ cm, $K=10^{-6}$ dyn, and $e=3\times10^{-4}$ dyn$^{1/2}$, we obtain

$$|\varepsilon_a| \ll 10. \tag{16}$$

For the excitation frequencies employed the penetration depth of the elastic wave is approximately ten times smaller than the penetration depth of the light. This increases the electrooptic response associated with the volume oscillations of the director and and the obvious replacement of the inequality (16) by

$$|\varepsilon_a| \ll 1, \tag{17}$$

which is in good agreement with the value of the dielectric anisotropy for our compensated mixture.

It follows from Eq. (15) that with the appropriate choice of the cell thickness the limitations imposed by the inequality (15) can be eliminated and the tilt bias angle can be measured for any NLC.

It should be noted that the method is highly sensitive to small tilt bias angles for uniform and homeotropic orientations. Since the Fourier harmonics of the electrooptic response at the fundamental frequency equal, in accordance with Eqs. (4) and (7), zero at $\theta_0 = 0°$ and $\theta_0 = 90°$, in this case we have a null method.

In conclusion, we have proposed a new method for measuring tilt bias angles in NLCs. The method has a number of advantages: 1) high accuracy, determined by the signal/noise ratio for the second harmonic of the electrooptic response; 2) constant geometry of the measurement scheme, which makes it easier to study the evolution of the tilt bias angle as a result of changes in temperature, effects due to the command surface, and others; and, 3) possibility of studying nonuniformly ordered samples.

This work was supported by COPERNICUS (IC15-CT96-0744), INTAS 95-IN-RU-128, INTAS 93-1700ext, and Russian Fund for Fundamental Research (95-02-0354) Grants.

[1] A. Stieb, G. Baur, and G. Meier, and Ber. Bunsenges, Phys. Chem. **78**, 899 (1974).
[2] G. Baur, F. Windscheid, and D. W. Berreman, Appl. Phys. **8**, 101 (1975).
[3] M. Shadt, H. Seiberle, and A. Schuster, Nature **381**, 212 (1996).
[4] R. Simon and D. M. Nicholas, J. Phys. D: Appl. Phys. **18**, 1423 (1985).
[5] E. Guyon, P. Pieranski, and M. Boix, Lett. Appl. Eng. Sci. **1**, 19 (1973).
[6] Y. Levy, D. Riviere, and C. Imbert, Opt. Commun. **26**, 225 (1978).
[7] T. J. Scheffer and J. Nehring J. Appl. Phys. **48**, 1783 (1977).
[8] R. Chiarelli, S. Faetti, and L. Fronzoni, Opt. Commun. **46**, 9 (1983).
[9] J. P. Nicholson, J. Physique **48**, 131 (1987).
[10] L. M. Blinov, D. B. Subachyus, and S. V. Yablonskii, J. Phys. II **1**, 459 (1991).
[11] L. M. Blinov, G. Durand, and S. V. Yablonskii, J. Phys. II **2**, 1287 (1992).
[12] S. V. Yablonskii, M. Rajteri, C. Oldano, and G. Durand, Proc. Soc. Photo-Opt. Intrum. Eng (SPIE) **2731**, 87 (1985).
[13] P. Chatelain, Bull. Soc. Fr. Mineral **66**, 105 (1943).

Translated by M. E. Alferieff

Interfacial electronic states in semiconductor heterostructures

A. A. Gorbatsevich and I. V. Tokatly
Moscow Institute of Electronics, 103498 Zelenograd, Moscow Region, Russia

(Submitted 3 December 1997; resubmitted 5 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 393–398 (25 March 1998)

It is shown that electronic states of a new type, with energy in the band gap can exist at a heterointerface. The interfacial states may be associated with Tamm surface states in the materials forming the heterointerface, but they can appear even if there are no surface states in the initial materials. In the plane of the heterojunction, the energy spectrum of interfacial states forms a two-dimensional band. © *1998 American Institute of Physics.* [S0021-3640(98)00706-3]

PACS numbers: 73.20.At, 73.40.Lq

Seventy years after Tamm in 1932[1] demonstrated the existence of surface states of a special type near the surface of a crystal, James[2] suggested that similar states could exist near an interface between two different materials. As far as we know, the next mention of possible interfacial states similar to Tamm levels was made in 1992.[3] An interface, like a surface, is a strong perturbation because of the discontinuity of the parameters of the material. The energy of such localized states can lie in both allowed and forbidden bands of the bulk dispersion relation. In the latter case, states localized at an interface will manifest as donor or acceptor impurities. The unusual behavior of charge carriers in InAs/AlSb heterostructures could be due to such states.[3] In Refs. 4 and 5 it was shown that localized states (''boundary states'') can indeed exist at an interface, but the energy of the boundary states always lies in the continuous spectrum (in the region of the offset of the band edges of the adjoining materials). At present only one type of very specific interfacial states with energy in the band gap has been described — states which arise in an inverted contact[6] as a result of local vanishing of the band gap at the contact point. These results were obtained in the continuous limit on the basis of the standard envelope method. A more consistent approach for describing strong perturbations localized on scales comparable to the interatomic separation is one based on the tight-binding approximation. In the present letter we obtain on the basis of the latter approach the conditions for the existence of interfacial states with energy in the band gap. It is shown that in the general case there are two types of localized interfacial states. States of the first type are genetically related with the surface states occurring in the materials forming the interface. Localized interfacial states of the second type arise even if the materials of the heteropair have no surface states, and they are entirely due to the specific structure of the interface. As a specific example we consider states arising at an interface between two cubic semiconductors with diamond or zinc blende structure.

1. Consider an atomically smooth interface between two crystals. Let the axis of growth be the z axis. To the extent that the system is uniform in a plane perpendicular to the growth axis, the problem is effectively one-dimensional. We shall enumerate the unit cells by integers n, numbers $n \leqslant 0$ referring to the left-hand crystal and numbers $n \geqslant 1$ to the right-hand crystal. We shall write out the Schrödinger equation in the tight-binding basis. For definiteness, let the dimension of the basis corresponding to one unit cell equal M. That is, each unit cell contains M orbitals and therefore the wave function Ψ_n in this basis is a vector of dimension M. The Schrödinger equation in the tight-binding approximation (in the sense that only hops between neighboring cells are taken into account) is

$$h_l \Psi^l_n + T_l \Psi^l_{n+1} + T_l^+ \Psi^l_{n-1} = E \Psi^l_n, \quad n \leqslant -1, \tag{1}$$

$$h_r \Psi^r_n + T_r \Psi^r_{n+1} + T_r^+ \Psi^r_{n-1} = E \Psi^r_n, \quad n \geqslant 2. \tag{2}$$

In Eqs. (1) and (2) h_j $(j=l,r)$ is the Hamiltonian of a unit cell, while T_j is a matrix of the hopping integrals between the states of neighboring cells. We note that for a three-dimensional system the matrices h_j and T_j depend on the transverse wave vector.

The equations for the edge (interfacial) cells with numbers $n=0$ and $n=1$ differ from Eqs. (1) and (2) to the extent that right- and leftward hops are nonequivalent:

$$h_l \Psi^l_0 + T_S \Psi^r_1 + T_l^+ \Psi^l_{-1} = E \Psi^l_0, \tag{3}$$

$$h_r \Psi^r_1 + T_r \Psi^r_2 + T_S^+ \Psi^r_0 = E \Psi^r_1, \tag{4}$$

where T_S is the hopping matrix between states of the interfacial cells. The equations (3) and (4) can be rewritten in a more compact form. We shall determine formally the function Ψ^l_1 satisfying the equation for the left-hand half space (1) at a site with $n=1$ and the function Ψ^r_0 satisfying the equation for the right-hand half space (2) at a site with $n=0$. Adding to the right- and left-hand sides of Eqs. (3) and (4) the quantities $T_l \Psi^l_1$ and $T_r^+ \Psi^r_0$, respectively, we obtain the relations

$$T_S \Psi^r_1 = T_l \Psi^l_1, \tag{5}$$

$$T_S^+ \Psi^l_0 = T_r^+ \Psi^r_0. \tag{6}$$

We note that a similar procedure was used in Ref. 7 to obtain the boundary conditions for a single-orbital one-dimensional chain. The relations (5) and (6) are actually the boundary conditions for Eqs. (1) and (2), determining the wave functions in the right- and left-hand half spaces. In the absence of interfacial hopping $(T_S=0)$ Eqs. (1) and (5) and Eqs. (2) and (6), respectively, determine the eigenfunctions of the noninteracting right- and left-hand semi-infinite crystals. It is well known that, besides delocalized itinerant states, surface Tamm states localized near the boundary can appear among the eigenfunctions of a semibounded crystal. The existence of hops between the right- and left-hand half spaces $(T_S \neq 0)$ has two effects: In the first place, a tunneling coupling appears between surface states which are genetically related with the right- and left-hand crystals and, in the second place, new localized states associated with the presence of the interface itself and not reducing to the Tamm levels of isolated crystals can appear.

2. We shall demonstrate the above-indicated possibilities for the example of an interface between two cubic crystals with diamond or zinc blende structure. The most common group-IV semiconductors and III–V compounds possess such a structure. Let

the growth axis z of the structure (for definiteness, $A_l^{III}B_l^{V}/A_r^{III}B_r^{V}$) lie in the (001) direction. We shall investigate the possibility of the appearance of localized states formed from valence-band states with zero momentum perpendicular to the axis of growth: $\mathbf{k}_\perp = 0$. These states correspond to the axis Δ of the three-dimensional Brillouin zone and project onto the Γ point of the two-dimensional Brillouin zone. We shall neglect the spin-orbit interaction. This greatly simplifies the calculations but does not change at all the qualitative results.

The valence-band spectrum on the Δ (001) axis consists of two bands. The upper twofold degenerate (neglecting spin) band is formed from atomic orbitals $|X\rangle$ and $|Y\rangle$. The second band corresponds to the mixture of the states $|Z\rangle$ and $|S\rangle$. Let us consider the localized states formed from the atomic states $|X\rangle$ and $|Y\rangle$, independent on the Δ axis of the states in the second band. In this case each unit cell of the right- and left-hand half spaces contains four orbitals: $|X_B\rangle$, $|Y_B\rangle$, $|X_A\rangle$, $|Y_A\rangle$. We shall show that this system is equivalent to contact between two Peierls chains with alternating bonds and site energies.

The intracell Hamiltonian h_j $(j=l,\ r)$ and hopping matrices T_α $(\alpha=l,r,S)$ appearing in Eqs. (1) and (2) and the boundary conditions (5) and (6) can be represented in the form[8]

$$h_j=\begin{bmatrix} E_B^j & t_{xx}^j-\sigma_x t_{xy}^j \\ t_{xx}^j-\sigma_x t_{xy}^j & E_A^j \end{bmatrix}, \quad T_\alpha=\begin{bmatrix} 0 & t_{xx}^\alpha+\sigma_x t_{xy}^\alpha \\ 0 & 0 \end{bmatrix}, \tag{7}$$

where E_B^j and E_A^j are the energies of the p states of the B and A atoms in the jth half space and t_{xx}^α and t_{xy}^α are, respectively, the hopping integrals between the orbitals of the same and different kinds on the B and A atoms. The nondiagonal 2×2 block matrices (7) can be diagonalized by a unitary transformation with the matrix

$$U=\frac{1}{\sqrt{2}}\begin{bmatrix} \sigma_x+\sigma_z & 0 \\ 0 & \sigma_x+\sigma_z \end{bmatrix}.$$

This transformation corresponds to a transition from the basis $|X\rangle$, $|Y\rangle$ to the basis functions

$$|\Delta_3\rangle=\frac{1}{\sqrt{2}}(|X\rangle+|Y\rangle), \quad |\Delta_4\rangle=\frac{1}{\sqrt{2}}(|X\rangle-|Y\rangle),$$

corresponding to the representations Δ_3 and Δ_4. The transformed matrices h_j' and T_α' are block diagonal:

$$h_j'=Uh_jU^+=\begin{bmatrix} E_B^j & t_{xx}^j-\sigma_z t_{xy}^j \\ t_{xx}^j-\sigma_z t_{xy}^j & E_A^j \end{bmatrix}, \quad T_\alpha'=UT_\alpha U^+=\begin{bmatrix} 0 & t_{xx}^\alpha+\sigma_z t_{xy}^\alpha \\ 0 & 0 \end{bmatrix}. \tag{8}$$

We shall write explicitly the system of equations for the components of the bispinor $(\Psi_n^j)^T=(u_n^j,v_n^j)$:

$$\begin{aligned} &(E_B^j-E)u_n^j+(t_{xx}^j-\sigma_z t_{xy}^j)v_n^j+(t_{xx}^j+\sigma_z t_{xy}^j)v_{n+1}^j=0,\\ &(E_A^j-E)v_n^j+(t_{xx}^j-\sigma_z t_{xy}^j)u_n^j+(t_{xx}^j+\sigma_z t_{xy}^j)u_{n-1}^j=0. \end{aligned} \tag{9}$$

The four equations (9) decompose into two uncoupled systems of two equations describing states with symmetry Δ_3 and Δ_4 and distinguished by the sign of t^j_{xy}. Each system of equations indicated above describes a Peierls chain of atoms, duplicate with respect to both bonds and site energies. The spectrum of delocalized itinerant states in the jth crystal is

$$E^j_{\pm}(k)=(E^j_{\rm A}+E^j_{\rm B})/2\pm\sqrt{(E^j_{\rm A}-E^j_{\rm B})^2/4+4(t^j_{xx})^2\,\cos^2k+4(t^j_{xy})^2\,\sin^2k/2}.$$

The lower (bonding) p band $E^j_-(k)$ corresponds to a twofold degenerate valence band. In addition, $E^j_-(0)=E^j_v$ is the valence band top located at the Γ point. The antibonding p band $E^j_+(k)$ lies above the conduction band corresponding at the Γ point to an antibonding S level.

We shall seek the localized interfacial solutions for each representation Δ_3 and Δ_4 in the form

$$\Psi^r_n=A\begin{pmatrix}u_r\\ v_re^{q_r}\end{pmatrix}e^{-2nq_r};\quad \Psi^l_n=B\begin{pmatrix}u_l\\ v_le^{-q_l}\end{pmatrix}e^{2nq_l}. \tag{10}$$

For definiteness let us assume that a plane of B_l-type atoms emerges at the surface of the left-hand crystal ($j=l$) and a plane of A_r-type atoms emerges at the surface of the right-hand crystal ($l=r$). Substituting the wave functions (10) into the boundary conditions (5) and (6) and using Eqs. (8) and (9) we obtain the following dispersion relations for the energies of the interfacial states:

$$F_{\pm}(E)=\gamma_{\pm}, \tag{11}$$

where

$$F_{\pm}(E)=\frac{(E^r_{\rm A}-E)(E^l_{\rm B}-E)}{(t^r_{\pm}e^{2q_r}-t^r_{\mp})(t^l_{\pm}e^{2q_l}-t^l_{\mp})},\quad \gamma_{\pm}=\frac{(t^S_{\pm})^2}{t^r_{\pm}t^l_{\pm}} \tag{12}$$

and $t^{\alpha}_{\pm}=t^{\alpha}_{xy}\pm t^{\alpha}_{xx}$. The upper and lower signs in Eqs. (11) and (12) correspond to the states Δ_3 and Δ_4, respectively. The decay parameters q_j are related with the energy E by the relation

$$\sinh q_j=\sqrt{(E-E^j_-)(E^j_+-E)/4t^j_+t^j_-}. \tag{13}$$

It follows from Eqs. (13) and (12) that localized solutions of the type (10) exist only if the energy lies in the common forbidden band $E^{\max}_-<E<E^{\min}_+$, where

$$E^{\min}_+=\min\{E^l_+(0),E^r_+(0)\},\quad E^{\max}_-=\max\{E^l_-(0),E^r_-(0)\}.$$

We note that the energy $E^{\max}_-$ corresponds to the real valence-band edge, while the level $E^{\min}_+$ lies above the physical conduction-band bottom E_c. Therefore solutions of Eq. (11) with energies $E_c<E<E^{\min}_+$ fall in the region of allowed states.

Let us analyze the possible solutions of Eq. (11). The qualitative form of the functions $F_{\pm}(E)$ for the case $t_{xy}>0$ is displayed in Fig. 1. The function $F_+(E)$ vanishes at the points $E^l_{\rm A}$ and $E^r_{\rm B}$. Therefore if there is no interfacial hopping ($t^S=0$), Δ_3-type surface states with energies $E^l_{\rm A}$ and $E^r_{\rm B}$ exist at the boundaries of the left- and right-hand crystals, respectively. As t^S increases, the interfacial levels move apart and vanish (not necessarily simultaneously) when the surface hopping integral reaches certain critical

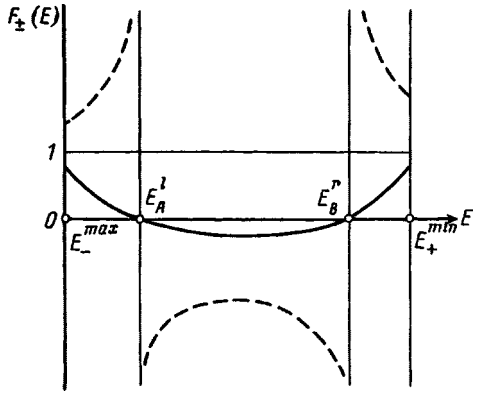

FIG. 1. Qualitative behavior of the functions $F_+(E)$ and $F_-(E)$ for the case $t_{xy}^j>0$. The solid line shows the function $F_+(E)$, while the dashed line shows the function $F_-(E)$.

values. This behavior corresponds to tunneling splitting of the Tamm levels initially present at the free surfaces. Type-Δ_4 solutions are an example of strictly interfacial states arising when there are no Tamm levels in semibounded crystals (no zeros of the function $F_-(E)$). Such states appear only if t^S exceeds a certain critical value. For example, the condition for the appearance of a level near the valence-band edge is

$$(t_-^S)^2>t_-^r t_-^l F_-(E_-^{\max}).$$

In this case the interface is a perturbing potential which localizes the charge carriers. Localized states arise with a finite value of this one-dimensional potential because of the asymmetry of the right- and left-hand crystals. The condition for the appearance of interfacial states greatly simplifies if the surface hopping integral equals the volume hopping integral in the right- or left-hand half spaces. For example, let $T_S=T_r$. Then the localized solutions of the type Δ_3 or Δ_4 arise if

$$\beta^{\pm 1}-\sqrt{\beta^{\pm 1}\frac{V(\Delta E_p^r-V)}{t_\pm^r t_\mp^r}}>1, \quad \beta=\frac{t_+^r t_-^l}{t_-^r t_+^l},$$

where V is the valence-band offset, while ΔE_p^r is the energy difference between the bonding and antibonding p levels at the Γ point. The latter formula demonstrates explicitly the influence of the asymmetry of the potential relief on the possibility of the appearance of localized states. Specifically, if the valence-band offset equals zero, then the interfacial solutions of one or another type arise with an arbitrary value of the parameter $\beta\neq 1$.

3. The results obtained above show that a semiconductor heterojunction is a strong perturbation that can result in the formation of localized states whose energies lie in the band gap. The spectrum of interfacial states in the plane of the heterojunction is of a band character and in this respect the new interfacial states are similar to Tamm surface states. However, cases when interfacial states appear even if the materials forming the heterojunction do not themselves possess Tamm surface states are possible. The new states will be manifested experimentally as characteristic features of the tunneling-resonance current-voltage characteristics and kinetics of generation-recombination processes. As

noted in Ref. 3, the existence of donor-type interfacial states can explain the unusual combination of high electron mobility and high electron density in quantum wells in undoped InAs/AlSb heterostructures. We note that the numerical solution of Eq. (11) for a InAs/AlSb system (values of the tight-binding parameters were taken from Ref. 8) shows that interfacial states indeed arise in a wide range of reasonable values of the interfacial hopping integrals.

The results of this work were obtained in their most general form in the tight-binding approximation, taking into account perturbations localized on the scale of a unit cell. It was shown that the Hamiltonian describing the states of the degenerate valence band of cubic semiconductors reduces to the Hamiltonian of a Peierls chain. It is well known that in the case of weak duplication such a system is described adequately by the method of envelopes using a two-band Dirac-type Hamiltonian. Physically, it is obvious that the results obtained in the present work should also be reproduced in such an approach. In this case the main problem is to determine the correct boundary conditions. The extension of the methods for describing interfacial states to the continuous limit and the microscopic derivation of the boundary conditions for the envelope functions will be examined in a separate work.

This work was supported by the Russian Fund for Fundamental Research and the Interdisciplinary Science and Technology Program ''Physics of Solid-State Nanostructures.''

[1] I. Tamm, Phys. Z. Sowjetunion **1**, 733 (1932).
[2] H. M. James, Phys. Rev. **76**, 1611 (1949).
[3] H. Kroemer, C. Nguen, and B. Brar, J. Vac. Sci. Technol. B **10**, 1769 (1992).
[4] R. Ą. Suris, Fiz. Tekh. Poluprovodn. **20**, 2008 (1986) [Sov. Phys. Semicond. **20**, 1259 (1986)].
[5] O. É. Raĭchev, Fiz. Tekh. Poluprovodn. **23**, 1226 (1989) [Sov. Phys. Semicond. **23**, 766 (1989)].
[6] B. A. Volkov and O. A. Pankratov, JETP Lett. **42**, 178 (1985).
[7] Q-G. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).
[8] P. Vogl, H. P. Hjalmarson, and J. D. Dow, Phys. Chem. Solids **44**, 365 (1983).

Translated by M. E. Alferieff

Model of isostructural dynamical phase transition in anharmonic crystal with possible relevance to $SrTiO_3$

V. G. Vaks
Kurchatov Institute Russian Science Center, 123182 Moscow, Russia

(Submitted 6 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 399–404 (25 March 1998)

A new type of phase transition is discussed which corresponds to a pairing of phonons of different lattice modes due to their anharmonic attraction in a crystal. It is shown that the main features of the isostructural phase transition observed in $SrTiO_3$ at $T \simeq 37$ K can be explained qualitatively by the phonon pairing phenomenon. © *1998 American Institute of Physics.* [S0021-3640(98)00806-8]

PACS numbers: 64.60.Ht, 64.70.Kb

The physical properties of the low-temperature phase of $SrTiO_3$ (STO) have recently received much attention. Phase-transition-like anomalies in the temperature dependences of a number of structural and elastic characteristics of STO have been observed at a temperature $T = T_q \simeq 37$ K (Refs. 1–6). The anomalies are rather weak, having the form of more or less smeared kinks, and no change in the crystal symmetry is seen at $T = T_q$. There are a number of speculations about the nature of this phase transition and its low-temperature phase (sometimes called ''the Müller state,'' after K. A. Müller, who discovered this phase transition and called attention to it (see reviews[4–6]). These speculations include invoking a new kind of Bose condensation,[5] some cooperative changes in the structure of dynamical polar clusters,[4] etc., but any clear physical model for this phase transition seems to be lacking as yet.

In the present work we suggest such a model. We adopt the qualitative arguments of Courtens[4] that the transition has a dynamical nature and is related to strong anharmonic interactions between the low-lying ''soft'' phonons that govern the dynamics of STO at the low temperatures under consideration. We suppose that this transition corresponds to a ''pairing'' of these phonons which in a number of formal aspects is analogous to the pairing of electrons in the standard BCS theory of superconductivity. Physically, such phonon pairing corresponds to a spontaneous breaking of the lattice symmetry for some phonon modes, which results in the relevant reconstruction of their energy spectrum and alteration of the character of their motion. We show that not only the isostructural phase transition but also some peculiar features of the low-temperature dynamics of STO (Refs. 2–4) seem to be naturally explained by this model.

To illustrate the main features of the phonon pairing we consider first a simplified model of only two interacting phonon branches, 1 and 2, with the following Hamiltonian:

$$H=\frac{1}{2}\sum_{\mathbf{k}}\sum_{\alpha=1}^{2}(\dot{x}_{\mathbf{k}}^{\alpha}\dot{x}_{-\mathbf{k}}^{\alpha}+\omega_{\mathbf{k}\alpha h}^{2}x_{\mathbf{k}}^{\alpha}x_{-\mathbf{k}}^{\alpha})+\frac{1}{4}\sum_{\mathbf{k_i}}\sum_{\alpha,\beta=1}^{2}V_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4}^{\alpha\alpha\beta\beta}x_{\mathbf{k}_1}^{\alpha}x_{\mathbf{k}_2}^{\alpha}x_{\mathbf{k}_3}^{\beta}x_{\mathbf{k}_4}^{\beta}. \tag{1}$$

Here $\mathbf{k}$ or $\mathbf{k}_i$ is the phonon wave vector, and the sum over four $\mathbf{k}_i$ in the last term obeys the quasimomentum conservation condition; $x_{\mathbf{k}}^{\alpha}$ and $\dot{x}_{\mathbf{k}}^{\alpha}$ are the phonon normal coordinate and momentum, which obey the canonical commutation relations

$$[x_{\mathbf{k}}^{\alpha},x_{\mathbf{q}}^{\beta}]=[\dot{x}_{\mathbf{k}}^{\alpha},\dot{x}_{\mathbf{q}}^{\beta}]=0,\quad [\dot{x}_{\mathbf{k}}^{\alpha},x_{\mathbf{q}}^{\beta}]=-i\hbar\,\delta_{\alpha\beta}\delta_{\mathbf{k}+\mathbf{q}}. \tag{2}$$

The quantity $\omega_{\mathbf{k}\alpha h}^{2}$ is the square of the $\mathbf{k}\alpha$ phonon frequency found in the harmonic approximation. For the soft-mode crystals under consideration this quantity in some intervals of $\mathbf{k}$ can be negative.[7] Then the system is stabilized with anharmonic interactions that in our model are described by the last term of (1). The anharmonic potentials V^{1111} and V^{2222} are supposed to be mainly positive, which ensures the lattice stability with respect to large values of x^{α}, while the intermode potential V^{1122} is supposed to be mainly negative, which corresponds to an ''attraction'' of different modes.

To investigate the phonon pairing qualitatively, we will treat the anharmonic interactions in (1) in the simplest mean-field (or Hartree–Fock, or ''self-consistent phonon'') approximation. This is a standard approach in treatments of anharmonic (in particular, of soft-mode) crystals, and in many problems it can be justified even quantitatively.[7] Within this approximation, each operator product $x_{\mathbf{k}}^{\alpha}x_{\mathbf{q}}^{\beta}$ in the last term of (1) is written as the sum of its average $\eta^{\alpha\beta}=\langle x_{\mathbf{k}}^{\alpha}x_{\mathbf{q}}^{\beta}\rangle=\delta_{\mathbf{k}+\mathbf{q}}\eta_{\mathbf{k}}^{\alpha\beta}$ and the fluctuation $\xi^{\alpha\beta}=x_{\mathbf{k}}^{\alpha}x_{\mathbf{q}}^{\beta}-\eta^{\alpha\beta}$, and the interaction of fluctuations ξ is neglected.[7] Then the terms with the ''diagonal'' averages $\eta^{\alpha\alpha}$ can be included in the values of the renormalized (and thus temperature-dependent) squared frequencies in (1), which is a conventional procedure, for example, in the perturbative treatments of soft modes.[7] However, the off-diagonal, ''anomalous'' averages $\eta_{\mathbf{k}}^{12}=\langle x_{\mathbf{k}}^{(1)}x_{-\mathbf{k}}^{(2)}\rangle$ are absent in any perturbative treatment, and their spontaneous rise at a certain temperature T_p corresponds to the phonon-pairing phase transition.

Denoting for brevity $\eta_{\mathbf{k}}^{\alpha\alpha}=\eta_{\mathbf{k}}^{\alpha}$, $\eta_{\mathbf{k}}^{12}=\eta_{\mathbf{k}}$, $V_{\mathbf{k},-\mathbf{k},\mathbf{q},-\mathbf{q}}^{\alpha\alpha\beta\beta}=U_{\mathbf{kq}}^{\alpha\beta}$, and $V_{\mathbf{k},\mathbf{q},-\mathbf{k},-\mathbf{q}}^{1122}=-V_{\mathbf{kq}}$, we can write the resulting Hamiltonian as

$$H=\frac{1}{2}\sum_{\mathbf{k}}(\dot{x}_{\mathbf{k}}^{\alpha}\dot{x}_{-\mathbf{k}}^{\alpha}+\tilde{\omega}_{\mathbf{k}\alpha}^{2}x_{\mathbf{k}}^{\alpha}x_{-\mathbf{k}}^{\alpha}-2\Delta_{\mathbf{k}}x_{\mathbf{k}}^{(1)}x_{-\mathbf{k}}^{(2)})-\frac{1}{2}\sum_{\mathbf{kq}}U_{\mathbf{kq}}^{\alpha\beta}\eta_{\mathbf{k}}^{\alpha}\eta_{\mathbf{q}}^{\beta}+\frac{1}{2}\sum_{\mathbf{kq}}V_{\mathbf{kq}}\eta_{\mathbf{k}}\eta_{\mathbf{q}}. \tag{3}$$

The repeating indices α or β here and below imply summation over α or β from 1 to 2, while the quantities $\tilde{\omega}_{\mathbf{k}\alpha}^{2}$ and $\Delta_{\mathbf{k}}$ are related to the averages $\eta_{\mathbf{k}}^{\alpha}$ and $\eta_{\mathbf{k}}$ as

$$\tilde{\omega}_{\mathbf{k}\alpha}^{2}=\omega_{\mathbf{k}\alpha h}^{2}+\sum_{\mathbf{q}}U_{\mathbf{kq}}^{\alpha\beta}\eta_{\mathbf{q}}^{\beta}, \tag{4}$$

$$\Delta_{\mathbf{k}}=\sum_{\mathbf{q}}V_{\mathbf{kq}}\eta_{\mathbf{q}}. \tag{5}$$

The Hamiltonian (3) can be diagonalized with the unitary transformation

$$x_{\mathbf{k}}^{(1)}=u_{\mathbf{k}}^{(1)}\cos\theta_{\mathbf{k}}+u_{\mathbf{k}}^{(2)}\sin\theta_{\mathbf{k}},\quad x_{\mathbf{k}}^{(2)}=-u_{\mathbf{k}}^{(1)}\sin\theta_{\mathbf{k}}+u_{\mathbf{k}}^{(2)}\cos\theta_{\mathbf{k}}, \tag{6}$$

and similarly for $\dot{x}^{\alpha}_{\mathbf{k}}$ and $\dot{u}^{\alpha}_{\mathbf{k}}$, where $u^{\alpha}_{\mathbf{k}}$ and $\dot{u}^{\alpha}_{\mathbf{k}}$ are the new canonical variables obeying the commutation relations analogous to (2). The diagonalized Hamiltonian has the form

$$H=\frac{1}{2}\sum_{\mathbf{k}}\ (\dot{u}^{\alpha}_{\mathbf{k}}\dot{u}^{\alpha}_{-\mathbf{k}}+\Omega^2_{\mathbf{k}\alpha}u^{\alpha}_{\mathbf{k}}u^{\alpha}_{-\mathbf{k}})-\frac{1}{2}\sum_{\mathbf{kq}}\ U^{\alpha\beta}_{\mathbf{kq}}\eta^{\alpha}_{\mathbf{k}}\eta^{\beta}_{\mathbf{q}}+\frac{1}{2}\sum_{\mathbf{kq}}\ V_{\mathbf{kq}}\eta_{\mathbf{k}}\eta_{\mathbf{q}}, \tag{7}$$

where the new frequencies $\Omega_{\mathbf{k}\alpha}$ are related to the original ones $\tilde{\omega}_{\mathbf{k}\alpha}$ as

$$\Omega^2_{\mathbf{k}1,2}=\frac{1}{2}\{\tilde{\omega}^2_{\mathbf{k}1}+\tilde{\omega}^2_{\mathbf{k}2}\pm[(\tilde{\omega}^2_{\mathbf{k}1}-\tilde{\omega}^2_{\mathbf{k}2})^2+4\Delta^2_{\mathbf{k}}]^{1/2}\} \tag{8}$$

and the parameter $\theta_{\mathbf{k}}$ in (6) is determined by the equation: $\tan 2\theta_{\mathbf{k}}=2\Delta_{\mathbf{k}}/(\tilde{\omega}^2_{\mathbf{k}1}-\tilde{\omega}^2_{\mathbf{k}2})$.

The quantity $\Delta_{\mathbf{k}}$ can be considered as the order parameter of the phase transition, and the self-consistency condition (5) for it yields the following equation:

$$\Delta_{\mathbf{k}}=\sum_{\mathbf{q}}\ V_{\mathbf{kq}}\varphi(\mathbf{q},\Delta^2_{\mathbf{q}})\Delta_{\mathbf{q}}, \tag{9}$$

where the function $\varphi(\mathbf{q},\Delta^2_{\mathbf{q}})$ is

$$\varphi(\mathbf{q},\Delta^2_{\mathbf{q}})=\frac{f_{\mathbf{q}2}-f_{\mathbf{q}1}}{\Omega^2_{\mathbf{q}1}-\Omega^2_{\mathbf{q}2}}\quad\text{with}\quad f_{\mathbf{q}\alpha}=\frac{\hbar}{2\Omega_{\mathbf{q}\alpha}}\coth\frac{\hbar\Omega_{\mathbf{q}\alpha}}{2T}. \tag{10}$$

One can also obtain Eq. (9) by minimizing the phonon free energy corresponding to the Hamiltonian (7) with respect to quantities $\eta_{\mathbf{k}}$ treated as variational parameters. The explicit form of Eq. (4) for $\tilde{\omega}^2_{\mathbf{k}\alpha}$ is obtained with the expressions for $\eta^{\alpha}_{\mathbf{k}}$ resulting from Eqs. (6)–(8):

$$\eta^{1,2}_{\mathbf{k}}=(f_{\mathbf{q}1}+f_{\mathbf{q}2})\pm\frac{\tilde{\omega}^2_{\mathbf{k}1}-\tilde{\omega}^2_{\mathbf{k}2}}{\Omega^2_{\mathbf{q}1}-\Omega^2_{\mathbf{q}2}}(f_{\mathbf{q}1}-f_{\mathbf{q}2}). \tag{11}$$

Near the phase transition point T_p the $\Delta_{\mathbf{q}}$ values are small. Writing in Eq. (9) the function $\varphi(\mathbf{q},\Delta^2_{\mathbf{q}})$ at small $\Delta_{\mathbf{q}}$ as $\varphi_{\mathbf{q}}-\Delta^2_{\mathbf{q}}\psi_{\mathbf{q}}$ with $\varphi_{\mathbf{q}}=\varphi(\mathbf{q},0)$, we obtain for T close to T_p

$$\sum_{\mathbf{q}}\ V_{\mathbf{kq}}\varphi_{\mathbf{q}}\Delta_{\mathbf{q}}-\Delta_{\mathbf{k}}=\sum_{\mathbf{q}}\ V_{\mathbf{kq}}\psi_{\mathbf{q}}\Delta^3_{\mathbf{q}}. \tag{12}$$

According to the definition (10), the function $\varphi_{\mathbf{q}}$ is positive. Below we show that for the soft-mode crystals under consideration the function $\psi_{\mathbf{q}}$ is normally positive, too. Positive values of $\psi_{\mathbf{q}}$ correspond to a second-order phase transition, while negative $\psi_{\mathbf{q}}$ would lead to a first-order transition, i.e., to an instability with respect to small values of $\Delta_{\mathbf{q}}$.

The solution of Eq. (12) can be conveniently written in terms of the eigenfunctions $\chi_{n\mathbf{k}}$ and eigenvalues λ_n of the linear integral equation corresponding to the left-hand side of (12):

$$\sum_{\mathbf{q}}\ V_{\mathbf{kq}}\varphi_{\mathbf{q}}\chi_{n\mathbf{q}}=\lambda_n\chi_{n\mathbf{k}}. \tag{13}$$

As the potential $V_{\mathbf{kq}}$ is symmetric in the variables $\mathbf{k}$ and $\mathbf{q}$, i.e., $V_{\mathbf{kq}}=V_{\mathbf{qk}}$, the functions $\chi_{n\mathbf{q}}$ can be orthonormalized with weight $\varphi_{\mathbf{q}}$ (see, e.g., Ref. 8). At the transition point T_p

the highest eigenvalue λ_0 in (13) reaches unity, and the pairing region corresponds to values $\lambda_0>1$. Using in Eq. (12) the usual perturbation expansion $\Delta_{\mathbf{k}}=\Sigma_n c_n\chi_{n\mathbf{k}}$ with $c_{n\neq 0}\ll c_0$ and neglecting higher powers of (λ_0-1), we obtain

$$\Delta_{\mathbf{k}}=(\lambda_0-1)^{1/2}A\chi_{0\mathbf{k}} \tag{14}$$

where the constant A is $(\Sigma_{\mathbf{q}}\psi_{\mathbf{q}}\chi_{0\mathbf{q}}^4)^{-1/2}$. Near T_p the factor (λ_0-1) is proportional to the difference T_p-T (having in mind STO, we suppose the pairing to occur at $T<T_p$). Therefore, the order parameter has the usual Landau-type temperature dependence $\Delta_{\mathbf{k}}\sim(T_p-T)^{1/2}$, which is natural for the mean-field approximation used. However, the observable phonon characteristics such as their spectra (8) include only even powers of $\Delta_{\mathbf{k}}$. Therefore, the temperature anomalies δf near T_p should usually have the form of kinks or weaker singularities: $\delta f\sim(T_p-T)\theta(T_p-T)$, $(T_p-T)^2\theta(T_p-T)$, etc., while the specific heat at $T=T_p$ should exhibit the jump that is characteristic of mean-field theory.

To illustrate the form of the functions $\varphi_{\mathbf{q}}$ and $\psi_{\mathbf{q}}$ in (12) we consider the case when the transition temperature T_p exceeds the values of $\hbar\tilde{\omega}_{\mathbf{k}\alpha}/2$ for significant phonons (which seems to be the case for STO, see below). Then $f_{\mathbf{q}\alpha}$ in (10) becomes $T/\Omega_{\mathbf{q}\alpha}^2$, and Eq. (4) is simplified:

$$\tilde{\omega}_{\mathbf{k}\alpha}^2=\omega_{\mathbf{k}\alpha h}^2+T\sum_{\mathbf{q}}\ U_{\mathbf{kq}}^{\alpha\beta}\frac{1}{\tilde{\omega}_{\mathbf{q}\beta}^2}\left(1-\frac{\Delta_{\mathbf{q}}^2}{\tilde{\omega}_{\mathbf{q}1}^2\tilde{\omega}_{\mathbf{q}2}^2}\right)^{-1}. \tag{15}$$

Let us write the small-Δ expansion of $\tilde{\omega}_{\mathbf{k}\alpha}^2$ as $\omega_{\mathbf{k}\alpha}^2+\Delta_{\mathbf{k}}^2\zeta_{\mathbf{k}\alpha}$ where $\omega_{\mathbf{k}\alpha}$ corresponds to the absence of pairing. Then the functions $\varphi_{\mathbf{q}}$ and $\psi_{\mathbf{q}}$ in (12) are

$$\varphi_{\mathbf{q}}=\frac{T}{\omega_{\mathbf{q}1}^2\omega_{\mathbf{q}2}^2},\quad \psi_{\mathbf{q}}=\frac{T}{\omega_{\mathbf{q}1}^2\omega_{\mathbf{q}2}^2}\left(\frac{\zeta_{\mathbf{q}1}}{\omega_{\mathbf{q}1}^2}+\frac{\zeta_{\mathbf{q}2}}{\omega_{\mathbf{q}2}^2}-\frac{1}{\omega_{\mathbf{q}1}^2\omega_{\mathbf{q}2}^2}\right), \tag{16}$$

while quantities $\zeta_{\mathbf{k}\alpha}$ are determined by the linear equations

$$\zeta_{\mathbf{k}\alpha}+T\sum_{\mathbf{q}}\left(\frac{\Delta_{\mathbf{q}}}{\Delta_{\mathbf{k}}}\right)^2U_{\mathbf{kq}}^{\alpha\beta}\frac{\zeta_{\mathbf{q}\beta}}{\omega_{\mathbf{q}\beta}^4}=T\sum_{\mathbf{q}}\left(\frac{\Delta_{\mathbf{q}}}{\Delta_{\mathbf{k}}}\right)^2U_{\mathbf{kq}}^{\alpha\beta}\frac{1}{\omega_{\mathbf{q}\beta}^2\omega_{\mathbf{q}1}^2\omega_{\mathbf{q}2}^2}. \tag{17}$$

For soft-mode crystals the sums over $\mathbf{q}$ in Eqs. (12) and (17) converge in the region of small wave vectors, where the harmonic values $\omega_{\mathbf{k}\alpha h}^2$ in (15) are negative. Therefore, for these k and $T\simeq T_p$ the sum over $\mathbf{q}$ in Eq. (15) exceeds $\omega_{\mathbf{k}\alpha}^2$, even though the main contribution to this sum is made by the thermally averaged phonons with ''average'' frequencies $\overline{\omega}_\alpha(T)$ rather than by small-q phonons with smaller frequencies $\omega_{\mathbf{q}}\sim\omega_0$. As the similar sums in Eq. (17) converge at small q and thus include additional large factors $\sim\overline{\omega}_\alpha/\omega_{0\alpha}$, one can see that the sum of two first terms in the brackets of Eq. (16) should, generally, exceed the third one, and thus the function $\psi_{\mathbf{q}}$ in (12) should be positive.

Let us now discuss the physical factors that can promote the phonon pairing. Let us suppose for simplicity that the $\mathbf{k}$ and $\mathbf{q}$ dependence of the potential $V_{\mathbf{kq}}$ in Eqs. (9) and (13) is insignificant, so that this potential can be approximated by its averaged value $V_0=\langle V_{\mathbf{kq}}\rangle$. Then the integral equations (9), (13) become algebraic equations, and Δ does not depend on $\mathbf{k}$. Supposing again that the temperature T exceeds the values $\hbar\omega_{\mathbf{q}\alpha}/2$ for significant phonons, we obtain the following equation for the transition temperature $T=T_p$:

$$V_0 T \sum_{\mathbf{q}} \frac{1}{\omega_{\mathbf{q}1}^2 \omega_{\mathbf{q}2}^2} = 1. \tag{18}$$

Equation (18) illustrates two necessary conditions for the emergence of phonon pairing: (i) the effective intermode interaction should correspond to a sufficiently strong attraction, which in our model corresponds to positive and sufficiently large values of V_0, and (ii) the phonon pairing on decreasing T can occur only in a soft-mode crystal in which the phonon frequencies $\omega_{\mathbf{k}\alpha}$ are both sufficiently small and sufficiently rapidly decreasing with decreasing T within significant intervals of the wave vectors $\mathbf{k}$.

In a discussion of possible phonon pairing in STO one should consider that two families of soft modes are present in this crystal at small k. They correspond to two kinds of order parameter: the ferroelectric one represented by the polarization $\mathbf{P}$, and the ''structural'' one represented by the angle $\mathbf{\Phi}$ describing the staggered rotation of oxygen octahedra in the perovskite structure.[5,9] At $T_a = 105$ K the structural phase transition from the cubic to the tetragonal phase occurs with the rise of the order parameter $\Phi = \Phi_z$. The ferroelectric transition does not occur down to $T = 0$ (though it can be induced by a small applied stress or by doping with a small amount of impurities), but the dielectric constant $\epsilon(T)$ rises at low T to very high values $\epsilon(0) \sim 10^5$ (Refs. 5 and 9). Accordingly, there are several soft modes in the small-k region: the ferroelectric transverse optical branches, in particular, those polarized along the z axis (which will be called for brevity P_z branches) and along the x or y axis (P_x or P_y branches), as well as the structural soft modes describing the rotation of the octahedra, both around the z axis (Φ_z branches) and around the x or y axis (Φ_x or Φ_y branches). The frequencies of all these modes are rather small and have an appreciable temperature dependence down to quite low T. For example, at $k = 0$ and $T = 4.2$ K the values of $\hbar\omega(P_z)$, $\hbar\omega(P_x)$, $\hbar\omega(\Phi_z)$, and $\hbar\omega(\Phi_x)$, according to the data of Ref. 9, are 27, 13, 65 and 21 K, respectively, while between $T = 58$ K and 22 K the value of $\hbar\omega(P_z)$ varies from 47 to 32 K.[2] Therefore, the above-mentioned condition (ii) can be satisfied in STO.

To get an idea of the anharmonic interactions between P_i and Φ_j modes we can use Uwe and Sakudo's estimates[9] of the nonlinear terms in the free energy of STO. As was discussed in Ref. 7, these terms correspond to the $k_i \to 0$ limit of the appropriate anharmonic interactions. Uwe and Sakudo[9] wrote these terms for the cubic phase as

$$F_{\text{int}} = -\sum_{ijkl} t^x_{ijkl} P_i P_j \Phi_k \Phi_l . \tag{19}$$

On account of the cubic symmetry there are only three different parameters t^x_{ijkl} in (19): $t^x_{xxxx} = t^x_{11}$, $t^x_{xxyy} = t^x_{12}$, and $t^x_{xyxy} = t^x_{44}$. As the degree of tetragonality of STO at $T < T_a$ is actually quite small, Eq. (19) can be used for estimates of anharmonic interactions at any T. Uwe and Sakudo[9] found: $t^x_{11} \simeq -7.4$, $t^x_{12} \simeq 9.6$, and $|t^x_{11}| \simeq 5$–9, in 10^{15} cgs. Therefore, the above-mentioned condition (i) can be satisfied for interactions of those P_i and Φ_j modes which are normal to each other, such as P_z and Φ_x or Φ_y modes for $\mathbf{k}$ in the xy plane, or P_x and Φ_y modes for $\mathbf{k}$ in the yz plane. Let us also note that on account of fluctuation effects (neglected in the above-described mean-field treatment and in Uwe and Sakudo's estimates[9]) the effective interactions t^x in (19) can actually rise appreciably with decreasing T, as can the dielectric constant $\epsilon(T)$, which is the case, for example, in

$BaTiO_3$, which is a structural analog of STO (see Ref. 10 or Ch. 8 of Ref. 7). Such a ''fluctuational'' rise of the effective interactions t^x can be one more factor promoting the phonon pairing transition in STO on decreasing T.

The presence of several soft P_i and Φ_j branches can imply that to describe the phonon pairing in STO one should employ not the above-described two-mode model, but more complex three- or four-mode pairing models. To this end we have considered the three-mode pairing model. Then the unitary transformation analogous to (6) includes three parameters, which can be taken as the Euler angles describing a three-dimensional rotation (see, e.g., Ref. 11). The resulting relations analogous to Eqs. (5)–(12) become more cumbersome, but all of the qualitative conclusions, including those about the character of the phase transition and the anomalies near T_p, remain the same as those for the two-mode model.

The phonon pairing can explain some anomalies that have been observed in the dynamics of STO at $T \lesssim T_q$ by Courtens *et al.*[2–4] In particular, the apparent ''mixing'' of the P_z- and Φ_x-type modes at small $\mathbf{k}=(k_x,k_y,0)$ stressed by those authors should be a direct consequence of the above-discussed (P_z,Φ_x) pairing at these $\mathbf{k}$. The other anomaly, the presence at such $\mathbf{k}$ of an additional, ''anomalous'' acoustic-like branch A, does not seem to be quite clear yet. However, this branch may be related to ''critical'' soft collective excitations under the phonon pairing, analogous to those discussed for a number of other phase transitions (see e.g., Ref. 12). Strong interactions between the acoustic modes (U modes) and the soft P_i and Φ_j modes can also be important for the existence of the A-branch. The relations between the frequencies of U, P_i, and Φ_j modes at these $\mathbf{k}$ can be approximately resonance-like, $\omega_{\mathbf{k}}^U \simeq |\omega_{\mathbf{k+q}}^P - \omega_{\mathbf{q}}^\Phi|$, in significant intervals of $\mathbf{q}$ (Refs. 2–4). It can promote the emergence of extra resonance-like excitations due to the anharmonic interactions. Under the phonon pairing (6) these interactions should rise appreciably as additional three-phonon couplings of the $UP\Phi$ type emerge, which at higher T are forbidden by the lattice symmetry. These effects can also persist at $T \gtrsim T_p$ due to fluctuational pretransition phenomena (neglected in the above-described mean-field treatment).

To summarize, the phonon pairing can be a new type of phase transitions in an anharmonic crystal, and the main features of the low-temperature isostructural phase transition in STO can be explained qualitatively by phonon pairing.

I am much indebted to A. I. Larkin for calling my attention to the problem of this phase transition in STO and for his valuable remarks, and to B. I. Shklovsky for his hospitality at the Theoretical Physics Institute, University of Minnesota, where this work was started.

[1] K. A. Müller, W. Berlinger, and E. Tosatti, Z. Phys. B **84**, 277 (1991).
[2] E. Courtens, G. Coddens, B. Hennion *et al.*, Phys. Scr. **T49**, 430 (1993).
[3] E. Courtens, B. Hehlen, G. Coddens, and B. Hennion, Physica B **219–220**, 577 (1996).
[4] E. Courtens, Ferroelectrics **183**, 25 (1996).
[5] K. A. Müller, Ferroelectrics **183**, 11 (1996).
[6] E. V. Balashova, V. V. Lemanov, R. Kunze *et al.*, Ferroelectrics **183**, 75 (1996).
[7] V. G. Vaks, *Introduction to the Microscopic Theory of Ferroelectrics*, Nauka, Moscow, 1973, Ch. 7.
[8] V. I. Smirnov, *Course of Advanced Mathematics*, Vol. 4, Fizmatgiz, Moscow, 1958.
[9] H. Uwe and T. Sąkudo, Phys. Rev. B **13**, 271 (1976).
[10] V. G. Vaks, Zh. Éksp. Teor. Fiz. **58**, 296 (1970) [Sov. Phys. JETP **31**, 161 (1970)].

[11] L. D. Landau and E. M. Lifshitz, *Mechanics*, 1st ed., Pergamon Press, Oxford, 1960 [Russian original, Fizmatgiz, Moscow, 1958, § 35].
[12] V. G. Vaks, V. M. Galitskiĭ, and A. I. Larkin, Zh. Éksp. Teor. Fiz. **51** 1592 (1966) [Sov. Phys. JETP **24**, 1071 (1967)].

Published in English in the original Russian journal. Edited by Steve Torstveit.

The excitonic spectrum of germanium in a high magnetic field

Yu. P. Kravchenko and M. A. Liberman
Department of Physics, Uppsala University, S–751 21, Uppsala, Sweden; P. Kapitsa Institute for Physical Problems, Russian Academy of Sciences, 117334 Moscow, Russia

(Submitted 28 October 1997; resubmitted 12 Februry 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 405–409 (25 March 1998)

In recently reported experiments with uniaxially deformed germanium in a magnetic field [V. B. Timofeev and A. V. Chernenko, JETP Lett. **61**, 617 (1995)], it was found that applying a magnetic field of sufficiently high intensity results in the appearance of a new line in the optical spectrum of the excitons. In the present paper a mechanism is proposed which can provide an explanation for this experimentally observed spectral feature. The new spectral line may be attributed to the formation of strongly bound biexcitonic molecules in the quantum state $^3\Pi_u$. © *1998 American Institute of Physics.*
[S0021-3640(98)00906-2]

PACS numbers: 71.35.Cc, 71.35.Ji, 78.40.Fy

Studies of the behavior of matter in high magnetic fields constitute one of the most interesting and rapidly developing areas in atomic and molecular physics.[1–4] Interest in this area is motivated by the radical changes in the electronic structure and properties of matter that occur when the paramagnetic energies of electrons become comparable to or greater than the typical energies of atomic and molecular bonds. In the case of ordinary atoms and molecules corresponding magnetic fields belong to the astrophysical domain. Indeed, the typical atomic binding energy may be estimated as 1 hartree $= m_e e^4/\hbar^2 = 27.2$ eV, while the paramagnetic energy of the electron is the distance between Landau levels, equal to $\hbar\omega_H = \hbar eH/m_e c$, where ω_H is the cyclotron frequency and H is the magnetic field intensity (we shall use cgs units). By equating these two expressions, we immediately find that the critical field strength for a ''high'' magnetic field is $H_0 = m_e^2 e^3 c/\hbar^3 = 2.35052\times 10^9$ G. Although far beyond reach by laboratory standards, such magnetic field strengths are nevertheless quite usual on the astrophysical scale. Magnetic fields in the vicinity of magnetic stars reach $0.1\ H_0$, while magnetic fields on the surfaces of neutron stars and pulsars can exceed H_0 by three orders of magnitude. For convenience, we shall measure the magnetic field in atomic units as $\gamma = H/H_0$.

This scale of magnetic fields profoundly changes, however, if one turns to the behavior of hydrogenlike excitons in semiconductors. First, the electron mass m_e changes to the reduced mass of an electron–hole pair $m = m_e m_h/(m_e + m_h)$. Second, the binding energy of excitons is reduced by a factor of ε^2, where ε is the dielectric constant of the semiconductor. As a consequence, the critical intensity of the magnetic field is reduced to

$H_0^* = m^2e^3c/\varepsilon^2\hbar^3$. This quantity is about several tesla for Ge and Si and may be as small as 0.2 T for InSb. Such a low value of the critical field implies that the structure of excitons in semiconductors must experience significant changes already in laboratory magnetic fields.

Such changes have indeed been observed in many laboratory experiments. One line of the experimental inquiry concentrates on the excited levels of excitons in semiconductors with relatively high values of the critical magnetic field. A typical such a material is Cu_2O, with a critical magnetic field $H_0^* \sim 800$ T, which demonstrates the ''chaotic'' behavior of excited excitonic levels in laboratory magnetic fields of the order of $10^{-2}H_0^*$ (Ref. 5). Another line of research is the study of excitons in semiconductors with critical magnetic fields comparable to or lower than the highest magnetic fields attainable in the laboratory. A representative example is uniaxially deformed germanium, which is especially interesting because this material allows one to investigate the influence of the magnetic field not only on separate excitons but on excitonic molecules as well.[6]

Recent experiments with Ge in magnetic fields up to 14 T have revealed new interesting features of the excitonic spectrum.[7] The authors of Ref. 7 studied the optical spectra of excitons in uniaxially deformed germanium placed in a magnetic field. The critical magnetic field strength for Ge was $H_0^* = 2.9$ T. In the absence of magnetic field the excitonic spectrum consisted of two lines: the line of excitons and the accompanying line of biexcitonic molecules. Application of the magnetic field resulted in a decrease in intensity of the biexcitonic line and its final disappearance at $\approx$ 1.5 T, which corresponds to an effective magnetic field $\gamma \approx 0.5$ a.u. When the applied magnetic field reached 4 T ($\gamma \approx 1.4$ a.u.), a new spectral line appeared. It was located on the ''red'' side of the line of free excitons and was labeled as the ''X'' line. This line was associated with the appearance of another bound state, whose energy is lower by one electron–hole pair than the energy of an isolated exciton.

The authors of Ref. 7 proposed two possible mechanisms explaining the observed spectrum. The first explanation, which they doubted, was based on the assumption of increased stability of the electron–hole liquid in a magnetic field.

An alternative explanation was the formation of a new biexcitonic molecular state. Although this possibility seemed promising, further progress in that direction was hindered by the lack of information about the behavior of excitonic (and hydrogen) molecules in high magnetic fields. Indeed, even such a basic question as the symmetry of the ground state of H_2 in magnetic field was the subject of a prolonged dispute. However, recent calculations of the electronic structure of the hydrogen molecule in high magnetic fields reveal more detailed information about the electronic states of H_2 and allow us to offer a possible explanation of the nature of the new spectral line described in Ref. 7.

We suggest that the explanation of the observed phenomenon lies in the appearance of metastable excitons in the quantum state $^3\Pi_u$. In what follows, we shall consider the electronic structure of the hydrogen molecule and make use of the fact that the hydrogen data can be scaled to describe the behavior of hydrogenlike excitons. The analysis is based on the Hartree–Fock calculations reported in Ref. 8. The molecular axis is directed along the magnetic field and the nuclei are assumed to be infinitely heavy. It was found in Ref. 8 that as the magnetic field increases, the ground state experiences two symmetry transitions. The first transition occurs at $\gamma \approx 0.18$ a.u., when the ground state changes

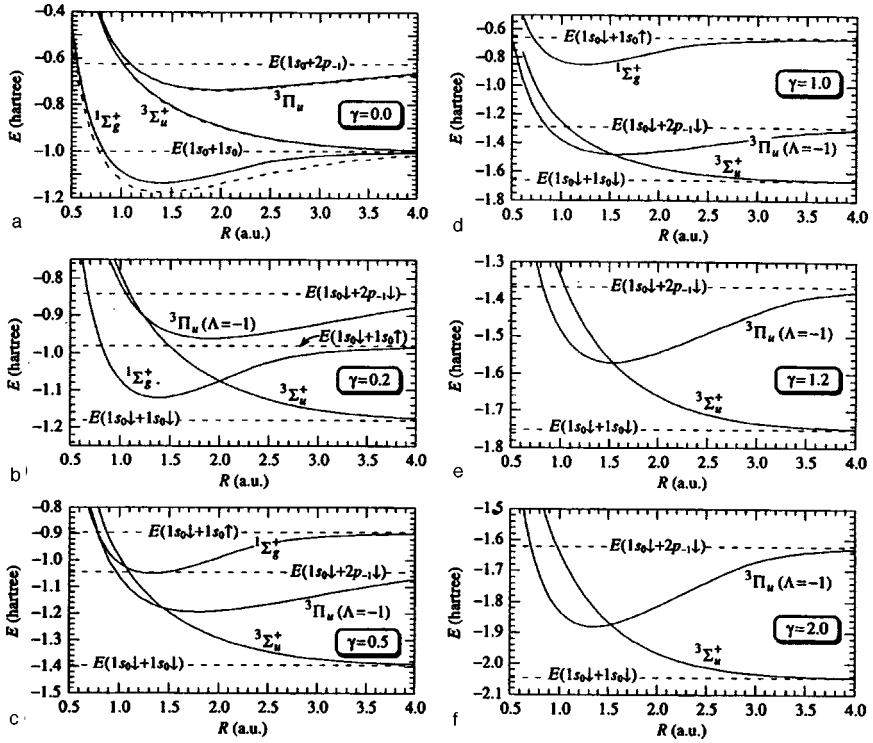

FIG. 1. Potential curves of the electronic states $^1\Sigma_g^+$, $^3\Sigma_u^+$, and $^3\Pi_u$ $(\Lambda=-1)$ of the hydrogen molecule in a high parallel magnetic field γ: (a) $\gamma=0.0$, (b) $\gamma=0.2$, (c) $\gamma=0.5$, (d) $\gamma=1.0$, (e) $\gamma=1.2$, and (f) $\gamma=2.0$ a.u. Dotted lines show energies of the corresponding states at infinite nuclear separation. For the triplet states $^3\Sigma_u^+$ and $^3\Pi_u$the total electronic spin is antiparallel to the magnetic field; the other two components of the triplet are not shown.

from the strongly bound singlet state $^1\Sigma_g^+$ to the weakly interacting triplet state $^3\Sigma_u^+$. The second transition happens at $\gamma\approx 14$ a.u. and involves a change from $^3\Sigma_u^+$ to the strongly bound triplet state $^3\Pi_u$. What is important for the present analysis, however, is the fact that, starting from $\gamma\approx 1.2$–1.4 a.u., the strongly bound state $^3\Pi_u$ may be a *metastable* state of the molecule. Therefore, starting from these intensities of the effective magnetic field, one can expect the formation of metastable biexcitons, which, we believe, provide an explanation for the ''X'' line observed by Timofeev and Chernenko.[7]

Let us consider how these ground state transitions occur. Fig. 1a shows the potential curves $^1\Sigma_g^+$, $^3\Sigma_u^+$, and $^3\Pi_u$ of the hydrogen molecule in the absence of magnetic field. The internuclear distance R is measured in the units of the Bohr radius $a_0=\hbar^2/m_e e^2 = 5.3\times 10^{-9}$ cm, and the zero of energy corresponds to the situation where all particles are at infinite separation from each other in their lowest Landau states, with their spins antiparallel to the magnetic field. The dashed curves show the total energy with allowance for the electron correlation, and the solid curves correspond to the Hartree–Fock approximation. While the correlation energy is significant for the singlet state $^1\Sigma_g^+$, its value for triplet states is much smaller due to the smaller overlap of electronic orbitals. It should be noted that the correlation energy remains approximately the same even in the

presence of magnetic fields up to $\gamma \lesssim 1$ a.u. This can be confirmed by comparing our Hartree–Fock results with detailed configuration-interaction calculations of Σ states of H_2 in magnetic field, which were recently reported in Ref. 9.

Figure 1b shows the potential curves of the same three quantum states in the presence of the parallel magnetic field $\gamma = 0.2$ a.u. We see that the symmetry of the ground state has changed. The singlet state $^1\Sigma_g^+$, which is the ground state at $\gamma = 0$, has been slightly shifted upwards, while the weakly interacting triplet state $^3\Sigma_u^+$ has been shifted lower due to the increase in the binding energy of isolated hydrogen atoms. As a result, the minimum of the potential curve of $^1\Sigma_g^+$ now lies *above* the energy of $^3\Sigma_u^+$ at $R \gg 1$, and the latter state represents the true ground state of the system. However, the hydrogen still can form tightly bound molecules in the state $^1\Sigma_g^+$, but such molecules will be metastable. As for the state $^3\Pi_u$, it lies above the potential curves of both $^1\Sigma_g^+$ and $^3\Sigma_u^+$ and is therefore unstable.

If we now increase the field to $\gamma = 0.5$ a.u., we encounter further changes. Figure 1c shows that the potential curve of the singlet state $^1\Sigma_g^+$ has shifted even higher, and $^1\Sigma_g^+$ is now an unstable state. This is manifested by the disappearance of the biexcitonic line at $\gamma \approx 0.5$ a.u., observed in Ref. 7. The potential curve of the state $^3\Pi_u$ is still above that of the state $^3\Sigma_u^+$. The hydrogen cannot form strongly bound molecules, and the ground state $^3\Sigma_u^+$ of the molecule is represented by two separated atoms. The molecule is bound very weakly, if at all, and the hydrogen now acts like a gas of weakly interacting atoms, which may exhibit such phenomena as Bose condensation and superfluidity.[2]

This situation remains essentially the same in a magnetic field $\gamma = 1$ a.u., as shown in Fig. 1d. However, the potential well of the triplet state $^3\Pi_u$ has deepened, and its minimum now lies only slightly above the curve of the state $^3\Sigma_u^+$. When we increase the magnetic field to $\gamma = 1.2$ a.u., the potential minimum of $^3\Pi_u$ crosses the energy curve of $^3\Sigma_u^+$ (Fig. 1d). This means that if the field increases even further, the hydrogen can start to form strongly bound metastable molecules in the state $^3\Pi_u$. Figure 1f shows that in a magnetic field $\gamma = 2.0$ a.u. the potential minimum of $^3\Pi_u$ lies below the potential curve of the weakly interacting state $^3\Sigma_u^+$, which is a typical picture of a metastable state.

Therefore, the quantum state $^3\Pi_u$ is an excellent candidate for the biexcitonic state, which can be responsible for the appearance of the ''X'' line observed in Ref. 7. The most convincing argument in the support of this claim is that the strength of the magnetic field at which the ''X'' line was first observed (4 T) corresponds to an effective magnetic field $\gamma \approx 1.4$ a.u. The calculations presented show that the state $^3\Pi_u$ becomes metastable starting from $\gamma \gtrsim 1.2$ a.u. It is unlikely that such a perfect agreement between the two values is accidental. However, in order to verify this claim one needs to perform similar experiments with other semiconductors having different values of the critical magnetic field.

Let us follow the further evolution of the states $^3\Sigma_u^+$ and $^3\Pi_u$. Figure 2 shows their potential curves in the magnetic field $\gamma = 10$ a.u. and $\gamma = 30$ a.u. We see that at $\gamma = 10$ a.u. the true ground state of the molecule is still the weakly interacting state $^3\Sigma_u^+$. At $\gamma = 30$ a.u. the potential minimum of the state $^3\Pi_u$ lies below the minimum of $^3\Sigma_u^+$, and the ground state of the system is the strongly bound $^3\Pi_u$. Calculations show that this second transition of the ground state symmetry occurs at $\gamma \approx 14$ a.u.[8]

Let us summarize our conclusions. We have proposed a possible theoretical expla-

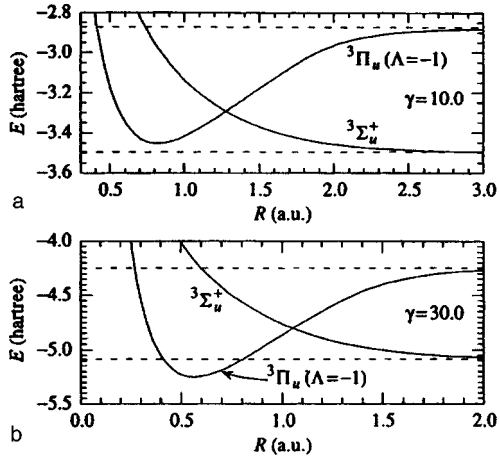

FIG. 2. Potential curves of the triplet states $^3\Sigma_u^+$ and $^3\Pi_u$ ($\Lambda=-1$) of the hydrogen molecule in the magnetic field: (a) $\gamma=10.0$ a.u., (b) $\gamma=30.0$ a.u. All designations are as in Fig. 1.

nation of the nature of the new excitonic line, observed experimentally in Ref. 7. According to our calculations, this new ''X'' line can be explained by the formation of metastable biexcitonic molecules in the quantum state $^3\Pi_u$. The calculated strength of the magnetic field at which such metastable molecules can exist, is in excellent agreement with the experimental results of Ref. 7. If the proposed mechanism is correct, it should work for other semiconductors with hydrogenlike excitons and manifest itself at the same effective magnetic field strength $\gamma\approx 1.4$ a.u. If such an effect is in fact observed in experiments with other types of semiconductors, it may open up alternative possibilities for controlling the optical spectrum of semiconductors by an applied magnetic field. Since certain kinds of semiconductors have very low values of the critical magnetic field, this mechanism may have potential technological applications.

This work was supported in part by the Swedish National Research Council (NFR), Contract No. F-AA/FU 10297-307, and in part by the Swedish Royal Academy of Sciences.

[1] M. A. Liberman and B. Johansson, Usp. Fiz. Nauk **165**, 121 (1995) [Phys. Usp. **38**, 117 (1995)].
[2] A. V. Korolev and M. A. Liberman, Phys. Rev. A **45**, 1762 (1992) (see also A. V. Korolev, Ph. D. thesis, Uppsala University, 1996, pp. 7–17); Physica A **193**, 347 (1993); Phys. Rev. B **47**, 14318 (1993); Phys. Rev. Lett. **72**, 270 (1994); Phys. Rev. B **50**, 14077 (1994); Int. J. Mod. Phys. B **10**, 729 (1996).
[3] U. Kappes and P. Schmelcher, Phys. Rev. A **53**, 3869 (1996), and references therein.
[4] Yu. P. Kravchenko, M. A. Liberman, and B. Johansson, Phys. Rev. Lett. **77**, 619 (1996); Phys. Rev. A **54**, 287 (1996); Yu. P. Kravchenko and M. A. Liberman, Int. J. Quantum Chem. **62**, 593 (1997).
[5] H. Matsumoto, K. Saito, M. Hasuo, S. Kono, and N. Nagasawa, Solid State Commun. **97**, 125 (1996); N. Nagasawa (private communication).
[6] V. D. Kulakovskii, I. V. Kukushkin, and V. B. Timofeev, Zh. Éksp. Teor. Fiz. **81**, 684 (1981) [Sov. Phys. JETP **54**, 366 (1981).
[7] V. B. Timofeev and A. V. Chernenko, JETP Lett. **61**, 617 (1995).
[8] Yu. P. Kravchenko and M. A. Liberman, Phys. Rev. A **55**, 2701 (1997).
[9] T. Detmer, P. Schmelcher, F. K. Diakonos, and L. S. Cederbaum, Phys. Rev. A **56**, 1825 (1997).

Published in English in the original Russian journal. Edited by Steve Torstveit.

Aharonov–Bohm effect for electrons on a liquid-helium surface

A. M. Dyugaev
L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

A. S. Rozhavskiĭ
Physicotechnical Institute of Low Temperatures, Ukrainian National Academy of Sciences, 310164 Kharkov, Ukraine

I. D. Vagner and P. Wyder
Grenoble High Magnetic Field Laboratory MPI-FRF and CNRS, VP 166, F-38042 Grenoble, France

(Submitted 19 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 410–414 (25 March 1998)

A method for obtaining finite electronic systems on a liquid-helium surface is proposed. If a thin film of liquid helium lies above a bottom capacitor plate made in the form of metal rings connected with one another, then electrons will accumulate in potential troughs near these rings. The purity of the helium surface, i.e., the absence of impurities and pinning centers on it, affords an excellent opportunity for investigating the Aharonov–Bohm effect in an ideal ring of a Wigner crystal and a Luttinger liquid [S0021-3640(98)01006-8]

PACS numbers: 67.70+n, 72.15.Rn

1. The Aharonov–Bohm (AB) effect[1] in ring-shaped solid-state electronic systems is suppressed by impurities and pinning centers in the case of an electronic Wigner crystal.[2] It is of interest to investigate experimentally an ideal ''pure'' electronic ring, since it is for such an object that the well-developed theory of the AB effect applies.[2,3] In this letter we propose a method for the experimental investigation of electronic rings on a liquid-helium surface experimentally. The purity of the helium surface, i.e., the absence of impurities on it, in principle permits observing AB oscillations in an ideal ring of an electronic Wigner crystal and an electronic Luttinger liquid by varying the electron density. In other words, it is possible to distinguish electron-electron interaction effects.

There exists a well-developed experimental method for obtaining two-dimensional electronic systems on a liquid-helium surface.[4] A thin helium film on a smooth substrate is placed in a capacitor, and an electron source is switched on above the helium surface. The electrons are pressed to the helium surface by electrostatic image forces and by the electric field. An electron density $n_e \approx 10^{11}$ cm^2 has been obtained on a thin helium film, and the quantum melting of an electronic Wigner crystal has been observed.[5] For low electron densities $n_e \approx 10^8$ cm^2 an anomalous increase (by four orders of magnitude) in

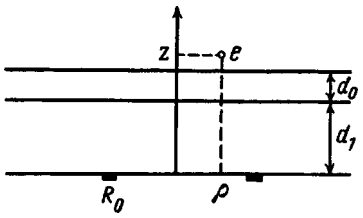

FIG. 1. Side view. The dark rectangles are transverse cross sections of the metal ring. R_0 is the radius of the ring, z and ρ are coordinates of the electron e, d_0 is the thickness of the helium film, and d_1 is the thickness of the substrate.

the mobility of a Wigner crystal on a He^3 film as the temperature is lowered to $T \approx 0.2 \times 10^{-3}$ K has been observed.[6] One-dimensional electronic systems have been investigated for helium films on diffraction gratings.[7,8]

2. Ring-shaped electronic systems can be obtained if the bottom plate of the capacitor is made in the form of metal rings connected with one another by bridges (see Fig. 1). Such a metal structure can be deposited both on the substrate directly above the helium film and on the back side of the substrate. This makes it possible to vary the distance between the metal ring and the helium surface over wide limits by varying the thickness of the substrate and the film of liquid. If the distance between the metal rings is much greater than their radius R_0, it is sufficient to take account of the effect of only one ring, which produces a potential $V(z,\rho)$ which is attractive for an electrons:

$$V(z,\rho) = -e^2 \int \frac{n(\rho_1)d\rho_1}{(z^2+(\rho-\rho_1)^2)^{1/2}}. \tag{1}$$

Here z and ρ are, respectively, the distances of the electron from the plane and center of the ring, $n(\rho)$ is the surface charge density of the metal ring, and e is the effective electron charge, which depends on the ratio of the dielectric constants of the helium film and the substrate beneath the film and the ratio of the thicknesses of the film and substrate.

The width Δ of a thin ring is much less than R_0, and we obtain from Eq. (1)

$$V(z,\rho) = -\frac{eQ}{\pi}\int_0^{\pi} \frac{d\varphi}{(z^2+(\rho-R_0)^2+4\rho R_0 \sin^2(\varphi/2))^{1/2}}, \tag{2}$$

where Q is the total effective charge of the metal ring. In terms of the dimensionless variables $x=\rho/R_0$ and $y=z/R_0$, the function $V(x,y)$ is given by the expression

$$V(x,y) = \frac{eQ}{R_0} I(x,y),$$

$$I(z,y) = -\frac{1}{\pi}\int_0^{\pi} \frac{d\varphi}{(y^2+(x-1)^2+4x \sin^2(\varphi/2))^{1/2}}. \tag{3}$$

For $y<1/\sqrt{2}$ the potential $V(x,y)$ has a minimum at $x \neq 0$. This is the case of a potential trough (ring) for electrons on the helium surface. However, if $y>1/\sqrt{2}$, then the mini-

mum of $V(x,y)$ occurs at $x=0$, i.e., for $\rho=0$ (2). This is the case of a potential well for electrons on the surface of the liquid. The expansion of $I(x,y)$ for small x and any y has the form

$$I(x,y)=-\frac{1}{\sqrt{1+y^2}}\left[1+x^2\frac{(1/2-y^2)}{(1+y^2)^2}+\frac{3}{8}x^4\frac{(y^4-3y^2+3/8)}{(1+y^2)^4}\right]. \tag{4}$$

For the case $y\ll 1$, i.e., $z\ll R_0$, replacing $\sin(\varphi/2)$ in Eq. (3) by $\varphi/2$, we have to logarithmic accuracy

$$I\cong-\frac{1}{2\pi}\ln\frac{4\pi^2}{y^2+(x-1)^2}. \tag{5}$$

Since electrons do not penetrate into the liquid helium,[4] the z coordinate of an electron is restricted by the condition $z>z_0$, where $z_0=d_0$, if the metal ring is located on top of the substrate, directly beneath the helium film of thickness d_0. However, if the ring lies beneath a substrate of thickness d_1, then $z_0=d_0+d_1$. The difference of z from z_0 is small, since the electrons are pressed to the liquid-helium surface by electrostatic image forces. The main effect for a thin helium film is attraction to the substrate, which is equivalent to switching on a strong electric field.[4] The potential $V(z,\rho)$ can be expanded in powers of two parameters: $z'=z-z_0$ and $\rho-\rho_0$, where ρ_0 is the value at which the the minimum of V with respect to ρ occurs for $z=z_0$. For example, in the case of a large effective potential ring ($z_0\ll R_0$) we obtain from Eqs. (2) and (5)

$$V=V(z_0,R_0)+\frac{eQ}{2\pi R_0}\left[\frac{(\rho-\rho_0)^2}{z_0^2}+\frac{2z'}{z_0}\right]. \tag{6}$$

In this case $\rho_0=R_0$. The term $\sim z'$ in Eq. (6) intensifies the pressing field of the electrostatic images of the helium and the substrate beneath it, while the total potential acting on an electron for small z' and $\rho-\rho_0$, minus a constant of no importance for us, has the form

$$V(\rho,z')=\frac{M\omega_0^2}{2}(\rho-\rho_0)^2+v(z'), \tag{7}$$

where M is the electron mass.

The radius ρ_0 of the effective potential trough for an electron lies in the range $0<\rho_0<R_0$ as z_0 varies over the interval $0<z_0<R_0/\sqrt{2}$, while the characteristic frequency ω_0 in Eq. (7) is proportional to the charge Q (2) of the metal ring. The characteristic frequency depends on the voltage across the capacitor plates. So, depending on the ratio of the radius R_0 of the metal ring and the distance of the ring from the helium surface z_0, it is possible to produce an effective oscillator potential (4), (7) for an electron, with a minimum as a function of ρ for arbitrarily small ρ_0.

3. The wave function Φ of an electron for the potential (7) in a magnetic field H directed along the z axis has the form

$$\Phi=\varphi(z)\Psi(\rho), \tag{8}$$

where the dependence of φ on z is of no importance for us, and $\Psi(\rho)$ is the solution of the Schrödinger equation with the Hamiltonian $\hat{H}$:[9]

$$\hat{H}=\frac{1}{2M}\left(\hat{p}-\frac{e}{c}\mathbf{A}\right)^2+\frac{M\omega_0^2}{2}(\rho-\rho_0)^2, \tag{9}$$

$\hat{p}$ is the electron momentum operator, and $\mathbf{A}$ is the vector potential. In the Fock gauge[10] $A\varphi=H\rho/2$ and $A_z=A_\rho=0$, we make the substitution

$$\Psi=e^{im\varphi}\frac{\chi(\rho)}{\sqrt{\rho}}$$

and obtain from Eq. (9) an equation for the function $\chi(\rho)$:

$$\frac{\hbar^2}{2M}\chi''+(E_m-\tilde{V}(\rho))\chi=0. \tag{10}$$

The energy E_m depends on the magnetic quantum number m, while the effective potential $\tilde{V}(\rho)$ has the form

$$\tilde{V}(\rho)=\frac{M\omega_0^2}{2}(\rho-\rho_0)^2+V_H(\rho),$$

$$V_H(\rho)=\frac{m^2-1/4}{\rho^2}\frac{\hbar^2}{2M}+\frac{M\omega_H^2}{8}\rho^2+\frac{\hbar\omega_H m}{2}, \tag{11}$$

where $\omega_H=eH/MC$ is the cyclotron frequency.

So, to determine the electron spectrum E_H it is necessary to solve Eq. (10) for χ with the potential $\tilde{V}(\rho)$ (11) and the boundary conditions $\chi(0)=\chi(\infty)=0$. We shall show that for large orbital numbers $m\gg 1$ Eq. (10) can be solved analytically. Indeed, replacing $m^2-1/4$ in Eq. (11) by m^2, we obtain an expression for V_H (11):

$$V_H=\frac{M\omega_H^2}{2}\left(\frac{\rho}{2}+\frac{\hbar m}{\rho M\omega_H}\right)^2. \tag{12}$$

The potential V_H has a minimum as a function of ρ at $\rho=\rho_m$ and its expansion around $\rho=\rho_m$ has the form

$$V_H=\frac{\hbar\omega_H}{2}(m+|m|)+\frac{M\omega_H^2}{2}(\rho-\rho_m)^2, \quad \rho_m^2=\frac{2\hbar|m|}{M\omega_H}. \tag{13}$$

It follows from Eqs. (11) and (13) that the effective potential $\tilde{V}$ is the sum of two oscillator potentials with minima at the points ρ_0 and ρ_m. Since we are interested in solutions for χ (10) that are concentrated near $\rho=\rho_0$, the boundary condition $\chi(0)=0$ is unimportant: The electron wave function at $\rho=0$ is exponentially small as it is. For this reason, by shifting the variable ρ in Eq. (10) the effective potential $\tilde{V}(\rho)$ can be put into the form

$$\tilde{V}(\rho_1)=\frac{\hbar\omega_H(m+|m|)}{2}+\frac{M}{2}\frac{\omega_H^2\omega_0^2}{\omega_H^2+\omega_0^2}(\rho_0-\rho_m)^2+\frac{M}{2}(\omega_H^2+\omega_0^2)(\rho_1-\tilde{\rho}_0)^2, \tag{14}$$

$$\tilde{\rho}_0 = \frac{\omega_0^2 \rho_0 + \omega_H^2 \rho_m}{\omega_H^2 + \omega_0^2}.$$

The equation (10) with the potential (14) has the spectrum[9,10] E_m^n:

$$E_m^n = \frac{\hbar \omega_H (m + |m|)}{2} + (n + 1/2)\hbar (\omega_H^2 + \omega_0^2)^{1/2} + \frac{M}{2} \frac{\omega_H^2 \omega_0^2}{\omega_H^2 + \omega_0^2} (\rho_0 - \rho_m)^2, \tag{15}$$

where ρ_m is determined in Eq. (13) and n is an integer. Since we have assumed that $\rho_0 \approx \rho_m$, in Eq. (15) $\rho_0 - \rho_m$ can be replaced by $\rho_0^2 - \rho_m^2/2\rho_0$ to the same degree of accuracy, and the last term in Eq. (15) can be written in the form

$$\frac{\hbar^2}{\rho_0^2 2M^*} \left(|m| - \frac{\Phi}{\Phi_0} \right)^2, \tag{16}$$

where M^* is the electron effective mass, Φ is the magnetic flux through the electronic ring, and Φ_0 is the elementary flux:

$$M^* = M \frac{\omega_H^2 + \omega_0^2}{\omega_0^2}, \quad \Phi = \pi \rho_0^2 H, \quad \Phi_0 = \frac{2\pi\hbar C}{e}. \tag{17}$$

We shall now examine limiting cases. In a weak magnetic field $\omega_0 \gg \omega_H$, in Eq. (16) M^* can be replaced by M, and we obtain from Eq. (15)

$$E_m^n = (n + 1/2)\omega_0 + \frac{\hbar^2}{2M\rho_0^2} \left(m + \frac{\Phi}{\Phi_0} \right)^2. \tag{18}$$

In this case the term proportional to $|m|$ in Eq. (15) equals zero. In a high magnetic field $\omega_H \gg \omega_0$, and we have from Eq. (15)

$$E_m^n = \hbar \omega_H \left(\frac{m + |m| + 1}{2} + n \right) + \frac{\hbar^2}{2M^* \rho_0^2} \left(|m| - \frac{\Phi}{\Phi_0} \right)^2. \tag{19}$$

The second term in Eq. (19) is important only for negative m, so that taking it into account lifts the infinite degeneracy of E_m^n with respect to m for $m < 0$. The AB effect is ordinarily studied in the weak-field limit, when expression (18) is applicable and the dependence of the spectrum E_m^n on the magnetic field H enters only via the magnetic flux Φ. For a thin helium film of thickness d on a substrate with dielectric constant ε the Coulomb interaction $V(r)$ of the electrons at large distances r is weakened[11]

$$V(r) = e^2 \left[\frac{1}{r} - \frac{\delta}{[r^2 + (2d)^2]^{1/2}} \right], \quad \delta = \frac{\varepsilon - 1}{\varepsilon + 1}. \tag{20}$$

In a high magnetic field, on account of the increase in the electron effective mass M^* (17), (19) the Coulomb interaction of the electrons becomes substantial, and a phase transition from the Luttinger liquid to a Wigner crystal can be observed in an electronic ring on a liquid-helium surface.

We thank Yu. N. Ovchinnikov for a discussion of this work.

This work was supported by the Russian Fund for Fundamental Research (Grant 96-02-18168) and the INTAS-RFBR fund (Grant 96-553).

[1] Y. Aharonov and Bohm, Phys. Rev. **115**, 485 (1959).
[2] C. L. Kane and M. P. A. Fisher, Phys. Rev. Lett. **68**, 1220 (1992).
[3] A. A. Zvyagin and I. V. Krive, Fiz. Nizk. Temp. **21**, 687 (1995) [Low Temp. Phys. **21**, 533 (1995)].
[4] V. B. Shikin and Yu. P. Monakhra, *Two-Dimensional Charged Systems in Helium* [in Russian], Nauka, Moscow, 1989.
[5] G. Mistura, T. Gunzler, B. Bitnar *et al.*, Surf. Sci. **361-362**, 831 (1996).
[6] K. Shirahama, O. I. Kirichek, and K. Kono, Phys. Rev. Lett. **79**, 4218 (1997).
[7] Yu. Z. Kovdrya and Yu. P. Monakhra, Fiz. Nizk. Temp. **12**, 1011 (1986) [Sov. J. Low Temp. Phys. **12**, 571 (1986)].
[8] V. A. Nikolaenko, H. Yayama, Yu. Z. Kovdrya, and A. Tomokiyo, Fiz. Nizk. Temp. **23**, 642 (1997) [Low Temp. Phys. **23**, 482 (1997)].
[9] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford, 1977 [Russian original, Nauka, Moscow, 1974].
[10] V. V. Fock, Z. Phys. **47**, 446 (1928).
[11] F. M. Peeters, Phys. Rev. **30**, 159 (1984).

Translated by M. E. Alferieff

Diluted generalized random energy model

D. B. Saakian

Yerevan Physics Institute, 375036 Yerevan, Armenia; LCTA, Joint Institute for Nuclear Research, 141980 Dubna, Moscow Region, Russia

(Submitted 13 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 415–419 (25 March 1998)

A layered random spin model, equivalent to the generalized random energy model (GREM), is introduced. In analogy with diluted spin systems, a diluted GREM (DGREM) is constructed. It can be applied to calculate approximately the thermodynamic properties of spin glass models in low dimensions. For the Edwards–Anderson model it gives the correct critical dimension and 5% accuracy for the ground state energy in two dimensions. © *1998 American Institute of Physics.* [S0021-3640(98)01106-2]

PACS numbers: 75.10.Nr, 05.50+q

Derrida's random energy model (REM)[1] was introduced as an archetype spin glass[2] model. In recent years it is becoming more and more popular. It has been applied in many fields of physics, biology, and even in information theory Refs. 3 and 4. The generalization of the REM (called the generalized random energy model, GREM) was introduced in Ref. 5. It has been used for approximate solution of other spin glass systems.[6,7] Unfortunately, the accuracy in describing other spin glass systems was not much better than for the REM. In this work we introduce a diluted spin model which thermodynamically resembles the GREM (in the case of large coordination number it is exactly equivalent to the GREM), then construct some new model of energy configurations — DGREM. In some cases of practical importance our spin model is thermodynamically exactly equivalent to the DGREM.

Even the simpler diluted REM (DREM)[8,9] has proven to be a good approximation for models in low dimensions (d=1,2,3). This important fact was observed in Ref. 10, where by information-theoretic arguments (mathematically leading to a DREM) a percolation threshold was found.

In the DREM one has N Ising spins interacting with each other in the z (randomly chosen from all the possible $C_N^p=N!/p!(N-p)!$) p-plets of Ising spins and quenched random couplings $\tau_{i_1,\cdots,i_p}$ having values ± 1.

The Hamiltonian reads

$$\mathcal{H}=-\sum_{(1\leqslant i_1,\cdots<i_p\leqslant N)=1}^{z}\tau_{i_1,\ldots,i_p}\sigma_{i_1}\cdots\sigma_{i_p}. \tag{1}$$

At high temperatures the system is in the paramagnetic phase and

$$\frac{F}{N} = -dT \ln\cosh\beta - T\ln 2, \tag{2}$$

where $\beta = 1/T$. Below the critical temperature $T_c = 1/\beta_c$ the system freezes in a spin-glass phase with internal energy $U/N = -\mathrm{d}\tanh\beta_c$ and vanishing entropy $S = 0$. Here $\tanh\beta_c = f(\mathrm{d})$ involves a function $f(x)$ defined by the implicit equation

$$\frac{1}{2}(1+f)\ln(1+f) + \frac{1}{2}(1-f)\ln(1-f)] = \frac{\ln 2}{x}. \tag{3}$$

For the ground state energy of the Edwards–Anderson (EA) model on a hypercubic lattice in d dimensions ($z = N\mathrm{d}$)

$$-\frac{E}{N} = f(\mathrm{d})\mathrm{d}. \tag{4}$$

In two dimensions Eq. (4) gives $E \approx -1.5599$, which is close to the result[11] of a Monte Carlo simulation for the case of random ± 1 couplings: $E/N = -1.4015 \pm 0.0008$. This estimate by formula (3) was done by Derrida in his original work,[1] long before the introduction of the DREM in Ref. 8.

Let us now construct a spin model which has properties like the GREM. It is very important to have a spin representation for the GREM (for example — in order to construct the temporal dynamics).

We consider a stacked system consisting of M planes with spin σ_i^k ordered along a ''vertical'' axis. In plane (layer) k there are N_k spins. So spins in the layer $1 < k < M$ interact with spins from the layers $k \pm 1$, the first layer interacts with the spins of the second layer, and spins from layer M interact with each other. We have the Hamiltonian

$$\mathcal{H} = -\sum_{(1 \leqslant i_1, \cdots < i_p \leqslant N_M)}^{z_M} \tau_{i_1,\cdots,i_p} \sigma_{i_1}^M \cdots \sigma_{i_p}^M$$
$$-\sum_{k=1}^{M-1} \sum_{(1 \leqslant i_1, \cdots < i_{p/2} \leqslant N_{k-1}, 1 \leqslant j_1 \cdots < j_{p/2} \leqslant N_k)}^{z_k} \tau_{i_1,\cdots,i_p} \sigma_{i_1}^{k-1} \cdots \sigma_{i_{p/2}}^{k-1} \sigma_{j_1}^k \cdots \sigma_{j_{p/2}}^k. \tag{5}$$

Let us now introduce some (equivalent) GREM like model. We consider some M level hierarchic tree. At the first level there are 2_1^N branches. At the second level every old branch fractures to 2^{N_2} new ones, and so on. At the level M there are 2^N branches, where $N = \Sigma_{i=1,M}^M = N_i$ energy configurations of our system are located on the ends of the Mth level branches. On every branch of level i there are located 2^{N_i} random variables ϵ_i^α with the distribution

$$\rho_0(\epsilon_i^\alpha, z_i) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} dk \exp[-k\epsilon_i^\alpha + z_i \ln\cosh k]. \tag{6}$$

This is a distribution for a sum of z_k random ± 1 variables. So z_k resembles the number of couplings in our diluted spin models. M branches are connected with any energy configuration. We define configuration energy as a sum (along the path on the tree,

connected with chosen energy configuration) of these M variables ϵ_i^α. We see the usual picture of the GREM, where random variables are distributed according to (6) instead of a normal partition.

We can consider the case of large M with a smooth distribution of z_k and N_k. In this case we can introduce the continuous variable $v=k/M$ between 0 and 1, labeling the levels of the planes, and define the distributions

$$z_k \equiv dz = z dv, \quad N_k \equiv dN = n'(v) dv \quad dv = \frac{1}{M}, \tag{7}$$

where $n(v)$ is a given function (the entropy in bits). The variable v $(0<v<1)$ parametrizes the level of the hierarchical tree, and z is a parameter (for our spin system z is the total number of couplings and the parameter v labels the levels of the planes).

Of course, our function $n(v)$ should be monotonic. The total number of energy configurations is $2^{n(0)}$, and $n(0)=N$. We have that 2^N energy levels E of our hierarchic model are distributed by partition $\rho(E)=\rho_0(E,z)$. If two configurations (in our GREM-like model) meet at a level of hierarchy v, they have zv common random variables. The energy difference between two configurations is related to $z(1-v)$ noncommon random variables. Therefore the distribution function of two energies E_1, E_2 reads

$$\rho_2(E_1-E_2)=\rho_0((E_1-E_2),2z(1-v))\exp(\ln 2 n(v)). \tag{8}$$

At high temperatures our system is in the paramagnetic phase. The free energy is given by Eq. (2). When we decrease the temperature, two situations are possible: first, $dz/dN \equiv z/n'(v)$ decreases monotonically with v; second, it has a local maximum.

In the first case the system has no sharp phase transition but freezes gradually. At the temperature $T=1/\beta$ all levels with $0 \leqslant v \leqslant v_f(T)$ are frozen; they are in the spin glass phase. The levels with $v_f<v\leqslant 1$ are in the paramagnetic phase; v_f is defined as the solution $v_f=v$ of the equation

$$\tanh\beta = f\left(\frac{z}{n'(v)}\right). \tag{9}$$

With this relationship between β and v we can later use functions $v(\beta)$ and $\beta(v)$. For every finite β the value of $v(\beta)$ lies between zero and unity. When $T\to\infty$, $v(\beta)\to 0$, and when $T\to 0$, $v(\beta)\to v_0>0$. So even in this limit some fraction of the spins stay in their paramagnetic phase. Let us point out that this partial freezing only is possible in the diluted GREM, and not in the original GREM. For the free energy we obtain (there is no factor of N in it):

$$-\beta F = z(1-v(\beta))\ln\cosh\beta + n(v(\beta))\ln + z\beta\int_0^{v(\beta)} dv_1 f\left(\frac{z}{n'(v_1)}\right). \tag{10}$$

The first two terms on the right-hand side describe the paramagnetic fraction of free energy ($n(v(\beta))\ln 2$ is just the entropy), while the last one describes the fraction of spins frozen in a glassy configuration (it resembles Eq. (3) with $z/f'(v_1)$ instead of d). In the second case (when the function $n'(v)$ is not monotonic) the system has a sharp first order phase transition at a finite temperature T_2. Below T_2 freezing occurs abruptly for all

levels $v<v_2$, where $v_2 \equiv v(\beta_2)$ is defined by the equation $n'(v_2)=N-n(v_2)$. We have used the fact that $n(0)=N$. The transition temperature $T_2=1/\beta_2$ follows from $\tanh(\beta_2)=f(z/n'(v_2))$. For temperatures $T<T_2$ the free energy reads

$$-\frac{\beta F}{N}=z(1-v(\beta))\ln\cosh\beta+n(v(\beta))\ln 2$$

$$+zv(\beta_2)\beta f\left(\frac{z}{n'(v_2)}\right)+z\beta\int_{v_2}^{v(\beta)}dv_1 f\left(\frac{z}{n'(v_1)}\right). \tag{11}$$

To construct the spin Hamiltonian by means of a chain of subsystems for this case is still an open problem. Let us now consider a possible approximation to the Edwards–Anderson model, following the ideas presented in Ref. 6. In the d-dimensional case our 2^N energy levels E are distributed according to the law

$$\rho(\epsilon)=\rho_0(\epsilon, N\mathrm{d}) \tag{12}$$

with ρ_0 defined in Eq. (6). Comparing with (6) one immediately notices that this is exactly equivalent to a DGREM with the choice $E=\epsilon$, $z=N\mathrm{d}$. Let us now consider the distribution of $\epsilon_1-\epsilon_2$. Following the arguments presented in Ref. 6, we find that

$$z=N\mathrm{d}, \quad n(v)=\frac{Ns(-v\,\mathrm{d}N)}{\ln 2}. \tag{13}$$

We see that the variable v corresponds to the energy per bond in the ferromagnetic model. We recall from the definition of temperature that $ds/dE=1/\tau\equiv\beta$. At given $\tilde{\beta}_1$ we can define the corresponding v_1 as the negative of the energy per bond for the ferromagnetic model at temperature $1/\tilde{\beta}_1$:

$$v_1=-\frac{E(\tilde{\beta}_1)}{N\mathrm{d}}. \tag{14}$$

We obtain for the free energy

$$-\frac{\beta F}{N\mathrm{d}}=(1-v(\beta))\ln\cosh\beta+s(v(\beta))+\beta\int_0^{v(\beta)}dv_1 f\left(\frac{\ln 2}{\tilde{\beta}}\right). \tag{15}$$

Integrating by parts in the last term, we get

$$-\frac{\beta F}{N\mathrm{d}}=(1-v(\beta))\ln\cosh\beta+s(v(\beta))-\beta\int_0^{\tilde{\beta}}d\tilde{\beta}_1\frac{2v_1(\tilde{\beta}_1)}{\ln\frac{1+y}{1-y}}+v(\beta)\beta y(\tilde{\beta}), \tag{16}$$

where y as a function of $\tilde{\beta}_1$ is defined by the equation

$$y=f\left(\frac{\ln 2}{\tilde{\beta}_1}\right), \tag{17}$$

the function $v(\beta)$ is defined by (9),(13), and $v_1(\tilde{\beta}_1)$ is the negative of the energy per bond in the ferromagnetic model at temperature $\tilde{\beta}_1$. In Eq. (16) the value of $\tilde{\beta}$ is related to the given β via the equation $\tanh(\beta)=f(\ln 2/\tilde{\beta})$. In the limit of zero temperature this reduces to

$$-\frac{\beta F}{N}=\mathrm{d}\left[1-\frac{2}{\mathrm{d}}\int_0^{\ln 2}\frac{d\tilde{\beta}_1 E(\tilde{\beta}_1)}{\ln 1+y(\tilde{\beta}_1)/1-y(\tilde{\beta}_1)}\right]. \tag{18}$$

Here $E(\tilde{\beta})=|U|$ is the negative of the energy in the ferromagnetic model, $y(\tilde{\beta})$ is defined by Eq. (17), and the function $f(x)$ is defined by Eq. (3). A calculation of the ground state energy for the two-dimensional EA model using (18) gives $E=-1.4763$. For the case of other models one can use numerical data for the ferromagnetic system.

This simple approximation to the ground state energy of disordered systems should be efficient at low dimensions.

I would like to thank B. Derrida, P. Rujan, and Th. Nieuwenhuizen for a critical discussion. This work was supported by German Ministry of Science and Technology Grant 211-5231.

[1] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).
[2] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1972 (1975).
[3] N. Sourlas, Nature **239**, 693 (1989).
[4] D. Saakian, JETP Lett. **55**, 798 (1992).
[5] B. Derrida, J. Physique Lett. **46**, 401 (1985).
[6] B. Derrida and E. Gardner, J. Phys. C **19**, 2253 (1986).
[7] B. Derrida and E. Gardner, J. Phys. C **19**, 5783 (1986).
[8] D. Dominicis and P. Mottishow, J. Phys. A **20**, L267 (1987).
[9] A. Allakhverdyan and D. Saakian, Nucl. Phys. B **498[FS]**, 604 (1997).
[10] P. Rujan, ''Coding Theory and Markov-Ansatze for Spin-Glasses,'' preprint 1997.
[11] C. De Simone, M. Diehl, M. Junger *et al.*, J. Stat. Phys. **84**, 1363 (1996).

Published in English in the original Russian journal. Edited by Steve Torstveit.

Is the Casimir effect relevant to sonoluminescence?

V. V. Nesterenko and I. G. Pirozhenko
Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, Moscow Region, Russia

(Submitted 23 February 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **67**, No. 6, 420–424 (25 March 1998)

The Casimir energy of a solid ball (or cavity in an infinite medium) is calculated by a direct frequency summation using contour integration. The dispersion is taken into account, and the divergences are removed by making use of the zeta function technique. The Casimir energy of a dielectric ball (or cavity) turns out to be positive and increasing as the radius of the ball decreases. The latter eliminates completely the possibility of explaining, via the Casimir effect, the sonoluminescence for bubbles in a liquid. Besides, the Casimir energy of the air bubbles in water proves to be immensely smaller than the amount of the energy emitted in a sonoluminescent flash. The dispersive effect is shown to be unimportant for the final result. © *1998 American Institute of Physics.* [S0021-3640(98)01206-7]

PACS numbers: 78.60.Mq, 47.55.Dz

1. Sonoluminescence, which has been observed for more than half a century,[1] has not yet found a satisfactory explanation. It is known that this phenomenon represents the emission of visual light by spherical bubbles of air or other gas injected into water and subjected to an intense acoustic wave in such a way that the radius of the bubbles changes periodically. In the last years of his life, Schwinger proposed[2] that the basis of sonoluminescence lies in the Casimir effect. When the size of the bubbles changes, so does the zero point energy of the vacuum electromagnetic field (the Casimir energy) of a cavity in a dielectric medium. According to Schwinger, it is these changes of the electromagnetic energy that are emitted as visible light in sonoluminescent flashes. In Schwinger's calculations the Casimir energy for the configuration in hand proves to be of the same order as the energy of the photons in an individual flash ($\sim$10 MeV). Other authors have obtained results both consistent with Schwinger's calculation[3] and differing from it by 10 orders or magnitude.[4,5] This disagreement is basically due to different methods used for removing the divergences in the problem under consideration.

In the present note the calculation of the Casimir energy of a dielectric ball placed in an infinite dielectric medium (or a cavity in such a medium) is carried out under following conditions. In the first place a realistic description of the dielectric properties of the media is used which takes dispersion into account.[a)] On the other hand, the simplest and most reliable method for removing the divergences, the zeta function technique, is applied. Till now these conditions have not been combined in studies of the problem in question.

2. In calculating the Casimir energy we shall use the mode-by-mode summation of the eigenfrequencies of the vacuum electromagnetic oscillations by applying contour integration in the complex frequency plane.[6,5] Consider a ball of a material which is characterized by a permittivity ε_1 and permeability μ_1. The ball is assumed to be placed in an infinite medium with permittivity ε_2 and permeability μ_2. For this configuration the frequencies of the transverse electric (TE) and transverse magnetic (TM) modes are determined by the equations[7]

$$\Delta_l^{\mathrm{TE}}(a\omega)\equiv\sqrt{\varepsilon_1\mu_2}\,\tilde{s}_l'(k_1a)\tilde{e}_l(k_2a)-\sqrt{\varepsilon_2\mu_1}\,\tilde{s}_l(k_1a)\tilde{e}_l'(k_2a)=0, \tag{1}$$

$$\Delta_l^{\mathrm{TM}}(a\omega)\equiv\sqrt{\varepsilon_2\mu_1}\,\tilde{s}_l'(k_1a)\tilde{e}_l(k_2a)-\sqrt{\varepsilon_1\mu_2}\,\tilde{s}_l(k_1a)\tilde{e}_l'(k_2a)=0, \tag{2}$$

where $\tilde{s}_l(x)=\sqrt{\pi x/2}\;J_{l+1/2}(x)$ and $\tilde{e}_l(x)=\sqrt{\pi x/2}\;H^{(1)}_{l+1/2}(x)$ are the Riccati–Bessel functions, $k_i=\sqrt{\varepsilon_i\mu_i}\,\omega, i=1,2$ are the wave numbers inside and outside the ball, respectively; a prime stands for differentiation with respect to the argument (k_1a or k_2a) of the Riccati–Bessel functions.

As usual, we define the Casimir energy by the formula

$$E=\frac{1}{2}\sum_s\;(\omega_s-\overline{\omega}_s), \tag{3}$$

where ω_s are the roots of Eqs. (1) and (2) and $\overline{\omega}_s$ are the same roots under condition $a\to\infty$. Here s is a collective index that stands for a complete set of indices specifying the roots of Eqs. (1) and (2): $s=\{l,m,n\}l=1,2,\ldots;m=-(l+1),-l,\ldots,l+1$, $n=1,2,\ldots$. The roots of Eqs. (1) and (2) do not depend on the azimuthal quantum number m. Therefore the corresponding sum gives a multiplier $(2l+1)$. Further we use the argument principle from complex analysis to represent the sum over n in terms of a contour integral. As a result, Eq. (3) can be rewritten as follows:

$$E=\sum_{l=1}^{\infty}E_l,\quad E_l=\frac{l+1/2}{2\pi i}\oint_C dz\;z\frac{d}{dz}\;\ln\frac{\Delta_l^{\mathrm{TE}}(az)\Delta_l^{\mathrm{TM}}(az)}{\Delta_l^{\mathrm{TE}}(\infty)\Delta_l^{\mathrm{TM}}(\infty)}, \tag{4}$$

where the contour C passes counterclockwise around the roots of the frequency equations (1) and (2) in the right-hand half plane. This contour can be deformed into a segment $(-i\Lambda,i\Lambda)$ of the imaginary axis and a semicircle of radius Λ with $\Lambda\to\infty$. In this limit the contribution of the semicircle to the integral (4) vanishes, with the result[5]

$$E_l=\frac{l+1/2}{\pi a}\int_0^{\infty}dy\;\ln\left\{\frac{4e^{-2(q_1-q_2)}}{(\sqrt{\varepsilon_1\mu_2}+\sqrt{\varepsilon_2\mu_1})^2}[\sqrt{\varepsilon_1\varepsilon_2\mu_1\mu_2}((s_l'(q_1)e_l(q_2))^2\right.$$

$$\left.+(s_l(q_1)e_l'(q_2))^2)-(\varepsilon_1\mu_2+\varepsilon_2\mu_1)s_l(q_1)s_l'(q_2)e_l(q_2)e_l'(q_2)]\right\}, \tag{5}$$

where $q_i=\sqrt{\varepsilon_i\mu_i}\,y, i=1,2$ and $s_l(z)$, $e_l(z)$ are the modified Riccati–Bessel functions: $s_l(z)=(\pi z/2)^{1/2}I_\nu(z)$, $e_l(z)=(2z/\pi)^{1/2}K_\nu(z)$, $\nu=l+1/2$.

Further we will content ourselves with an examination of the case when both the media are nonmagnetic $\mu_1=\mu_2=1$ and have permittivities ε_1 and ε_2 that differ only slightly. In view of this we can put $q_1=q_2$ in Eq. (5), and elsewhere keeping ε_1 and ε_2 exactly. It gives

$$E_l=\frac{l+1/2}{\pi a}\int_0^\infty dy\ \ln\{1-\xi^2[(s_l(y)e_l(y))']^2\},\quad \xi^2=\left(\frac{\sqrt{\varepsilon_1}-\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1}+\sqrt{\varepsilon_2}}\right)^2. \tag{6}$$

Now we are going to take into account the effect of dispersion, considering the parameter ξ^2 in Eq. (6) to be a function of $y=a\omega/i$. Justification of the mode-by-mode summation method in applying to dispersive and absorptive media has been considered in Ref. 8. For definiteness we put $\varepsilon_1=1+\delta$, $\varepsilon_2=1$, $\delta\ll 1$; then $\xi^2\simeq\delta^2/16$. We substitute for δ the expression

$$\delta(y)=\delta_0/[1+(y/\nu y_0)^2],\quad \nu=l+1/2, \tag{7}$$

where δ_0 is a static value of $\delta(y)$ and the parameter y_0 is determined by a ''plasma'' frequency ω_0: $y_0=a\omega_0$. The function describing dispersion in Eq. (7) is a standard one [the one-absorption-frequency Sellmeir dispersion relation] except for its dependence on l. We have introduced this dependence in order to be able to use the zeta function technique below. This complication does not contradict the main goal pursued by using this function, namely, that it should roughly simulate the behavior of $\delta(y)$ at large y. It is known[9] that general theoretical principles lead to the following properties of the function $\varepsilon(\omega)$ in the upper half of the ω plane. On the imaginary axis $\omega=iy$, $y>0$ the function $\varepsilon(iy)$ acquires real values, and with increasing y it steadily decreases from the static value $1+\delta_0>0$ (for dielectrics) to 1. Obviously formula (7) meets these requirements.

Substituting (7) into (6) and making use of the uniform asymptotic expansion for the modified Bessel functions[10] as $l\to\infty$, one obtains

$$E_l\underset{l\to\infty}{\simeq}-\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0)+\mathcal{O}(\nu^{-2}), \tag{8}$$

where

$$f(z)=\frac{z}{(1+z)^4}\left(z^3+4z^2+\frac{16}{3}z+\frac{4}{3}\right). \tag{9}$$

As z increases, the function $f(z)$ monotonically rises, approaching 1: $f(0.01)=0.013$, $f(0.1)=0.130$, $f(1)=0.729$, $f(10)=0.994$.

We carry out the summation of the partial energies (6) with the help of the zeta function technique,[11] taking into account asymptotic behavior (8)

$$\begin{aligned}E=\sum_{l=1}^\infty E_l&=\sum_{l=1}^\infty\left[E_l+\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0)-\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0)\right]\\&=\sum_{l=1}^\infty \overline{E_l}-\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0)\sum_{l=1}^\infty (l+1/2)^0\\&=\sum_{l=1}^\infty \overline{E_l}-\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0)[\zeta(0,1/2)-1].\end{aligned} \tag{10}$$

Here $\overline{E_l}=E_l+(3/64a)(\delta_0/4)^2 f(a\omega_0)$ is the renormalized partial Casimir energy, $\zeta(s,q)$ is the Hurwitz zeta function. As $\zeta(0,1/2)=0$, we get for the Casimir energy (10)

$$E=\sum_{l=1}^{\infty}\overline{E}_l+\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0). \tag{11}$$

As in the calculations without allowance for dispersion effects,[5] here we can content ourselves with the last term on the right in (11). This provides us with an accuracy of a few percent.[12] Thus the Casimir energy of a dielectric ball is

$$E\simeq\frac{3}{64a}\left(\frac{\delta_0}{4}\right)^2 f(a\omega_0), \tag{12}$$

the effect of dispersion reducing to the introduction of a positive coefficient $f(a\omega_0)<1$. Hence, a change in the sign of the energy or a considerable increase in its magnitude due to the dispersion effect[13] is out of the question.

Let us estimate the value of $f(a\omega_0)$. The parameter ω_0 can be determined by demanding that at this frequency the photons do not ''feel'' the interface between two media. This condition will be certainly met when the wavelength of the photon is less than the interatomic distance in the media, $d\sim 10^{-8}$ cm. This is actually the condition of applicability of the macroscopic description of dielectric media.[9] Sonoluminescence is observed for air bubbles in water[1] with a bubble radius $a\sim 10^{-4}$ cm. Hence it follows that $a\omega_0\simeq a/d=10^5$ and $f(10^5)=0.999\ldots$. Thus allowance for dispersion in calculating the Casimir energy of a dielectric ball (or of a spherical cavity in a dielectric slab) has practically no effect on the final result.

Certainly the real picture of dispersion in the whole frequency range $0<\omega<\infty$ for any dielectric, including water, is exceedingly complicated and cannot be described by a simple equation (7) with single a parameter ω_0. It is known that absorption of electromagnetic waves in water and, as a consequence, their dispersion take place already in the radio frequency band. Putting in this case $\lambda\sim 10^4$ cm, we obtain $a\omega_0\sim 1$ and $f(1)=0.729\ldots$ Hence one can infer that the effective value of $a\omega_0$ should be less than 10^5. In order for a more precise evaluation of this parameter to be done a more detailed consideration of the dispersion mechanism is needed. Obviously this may lead only to diminution of the absolute value of the Casimir energy. However this issue is beyond the scope of the present paper, for the main conclusion (see below) does not depend on this point.

It is worth noting two peculiarities of the final formula (12). As the radius of a bubble decreases, its Casimir energy increases. This behavior is completely opposite to what would be needed to explain sonoluminescence (it is known that the emission of light takes place at the end of the collapse of bubbles in a liquid). Besides, this energy is immensely smaller than the amount of energy emitted in a separate sonoluminescent flash ($\sim$10 MeV). Actually, taking $a=10^{-4}$ cm and $\delta_0=3/4$ (water), we arrive at a value $E\simeq 5\times 10^{-3}$ eV.

Thus the results of this paper unambiguously demonstrate that the Casimir effect is irrelevant to sonoluminescence.

This work was accomplished with financial support of Russian Fund for Fundamental Research (Grant 97-01-00745).

[a)]There is a point of view[14,13] that dispersion effects may substantially affect the final result when calculating the Casimir energy.

---

[1] H. Frenzel and H. Schultes, Z. Phys. Chem. Abt. B **27**, 421 (1934); B. P. Barber, R. A. Hiller, R. Löfstedt, Phys. Rep. **281**, 65 (1997); L. A. Crum, Phys. Today **47**(9), 22 (1994).
[2] J. Schwinger, Proc. Natl. Acad. Sci. USA **90**, 958, 2105, 4505, 7285 (1993); **91**, 6473 (1994).
[3] C. E. Carlson, C. Molina-París, J. Pérez-Mercader, and Matt Visser, Phys. Lett. B **395**, 76 (1997); Phys. Rev. D **56**, 1262 (1997); C. Molina-Paris and Matt Visser, ''Casimir effect in dielectrics: surface area contribution,'' http://xxx.lanl.gov/abs/hep-th/9707073.
[4] K. A. Milton and Y. J. Ng, Phys. Rev. E **55**, 4207 (1997); K. A. Milton and Y. J. Ng, ''Observability of the bulk Casimir effect: Can the dynamical Casimir effect be relevant to sonoluminescence?'' Preprint OKHEP-97-04, http://xxx.lanl.gov/abs/hep-th/9707122.
[5] I. Brevik, V. V. Nesterenko, and I. G. Pirozhenko, ''Direct mode summation for the Casimir energy of a solid ball,'' JINR Preprint E2-97-307, Dubna (1997), http://xxx.lanl.gov/abs/hep-th/9710101.
[6] V. V. Nesterenko and I. G. Pirozhenko, Phys. Rev. D **57**, 1284 (1998).
[7] J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941.
[8] Yu. S. Barash and V. L. Ginzburg, JETP Lett. **15**, 403 (1972); Usp. Fiz. Nauk **116**, 5 (1975) [Sov. Phys. Usp. **18**, 305 (1975)].
[9] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon, Oxford, 1984.
[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C., 1964; reprinted by Dover, New York, 1972.
[11] E. Elizalde, S. D. Odintsov, A. Romeo *et al.*, *Zeta Regularization Techniques with Applications*, World Scientific, Singapore, 1994.
[12] V. V. Nesterenko and I. G. Pirozhenko, ''Is the Casimir effect relevant to sonoluminescence?'' JINR Preprint, Dubna 1998.
[13] I. Brevik and G. Einevoll, Phys. Rev. D **37**, 2977 (1988); I. Brevik and R. Sollie, J. Math. Phys. **31**, 1445 (1990); I. Brevik, H. Skurdal, and R. Sollie, J. Math. A: Math. Gen. **27**, 6853 (1994); I. Brevik and V. N. Marachevsky, ''Casimir surface force on a dilute dielectric ball,'' Preprint Norwegian University of Science and Technology, Trondheim, 1998.
[14] P. Candelas, Ann. Phys. (N.Y.) **143**, 241 (1982).

Published in English in the original Russian journal. Edited by Steve Torstveit.