Galactic anisotropy as signature of ''top-down'' mechanisms of ultrahigh-energy cosmic rays

S. L. Dubovsky and P. G. Tinyakov
Institute for Nuclear Research, Russian Academy of Sciences, 117312 Moscow, Russia

(Submitted 10 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 99–103 (25 July 1998)

We show that ''top-down'' mechanisms of ultrahigh-energy cosmic rays which involve heavy relic particle-like objects predict a Galactic anisotropy of the highest-energy cosmic rays at the level of minimum $\sim 20\%$. This anisotropy is large enough to be either observed or ruled out in the next generation of experiments. © *1998 American Institute of Physics.* [S0021-3640(98)00114-5]

PACS numbers: 98.70.Sa, 95.30.Cq

The origin and nature of ultrahigh-energy cosmic rays (UHE CRs) with energies above 10^{19} eV[1] is one of the actively debated issues in modern astrophysics. At such a high energy the mean free path of protons, nuclei, and photons is much shorter than the size of the Universe due to their interaction with the cosmic microwave and radio background.[2,3] Protons with $E > E_{\mathrm{GZK}} \sim 5 \cdot 10^{19}$ eV can only come from distances less than $R_{\mathrm{GZK}} \sim 50$ Mpc; the corresponding region for nuclei and photons is even smaller. The energy spectrum of cosmic rays is thus expected to have a rapid falloff at energies $E \sim E_{\mathrm{GZK}}$ (the so-called Greisen–Zatsepin–Kuzmin cutoff).

The observation of few events with energies exceeding 10^{20} eV seems to indicate that a noticeable fraction of the UHE CRs comes from relatively nearby ($R < R_{\mathrm{GZK}}$) sources. The latter is not easy to reconcile with conventional astrophysical mechanisms,[4] as at these energies and distances the cosmic rays are not substantially deflected by magnetic fields and point in the direction of their source.[5] Corresponding astrophysical sources have not yet been identified. This prompts one to consider particle-physics mechanisms of UHE CR production, which are usually referred to as ''top-down'' mechanisms. Several such mechanisms involving topological defects[6] and decays of primordial heavy particles[7–10] have been proposed. In this letter we show that a large class of the top-down mechanisms predict a significant excess of the highest-energy events in the direction towards the center of our Galaxy. Thus, these mechanisms can be tested experimentally by studying the asymmetry in the angular distribution of UHE CRs. The predicted asymmetry is large enough to be either observed or excluded by future experiments.[11]

Our analysis concerns mechanisms in which the production of UHE CRs involves heavy relic particle-like objects which behave as part of the cold dark matter (CDM). We call such mechanisms CDM-related. The crucial feature of these mechanisms is that the distribution of sources of UHE CRs in the Universe between the galaxies and the inter-

galactic space is proportional to that of the CDM (cf. Ref. 12) and does not depend on their nature. In particular, the sources are mainly concentrated in galactic halos, so that their average densities in the Universe ($\overline{n}$) and in galactic halos ($\overline{n_h}$) are related by

$$\frac{\overline{n}}{\overline{n_h}} \simeq \frac{\Omega_{\rm CDM}\rho_{\rm crit}}{\overline{\rho}_{\rm halo}} \sim 10^{-5}. \tag{1}$$

On the contrary, the distribution $n(x)$ of the sources in a galactic halo depends on their interaction with each other and with other matter and does not necessarily follow that of CDM.[a)] For the decay-type mechanisms $n(x)$ is simply the density of decaying particles. For mechanisms based on collisions (see, e.g., Ref. 13) $n(x) \propto \tilde{n}^2(x)$, where $\tilde{n}(x)$ is the density of colliding particles. Note that in the latter case the distribution $n(x)$ is typically more concentrated around the galactic center.

The observed flux of UHE CR can be divided into Galactic and extragalactic parts,

$$j = j_{\rm ext} + j_h,$$

where

$$j_h = C\int_{\rm halo} \frac{d^3x}{x^2} n(x), \tag{2}$$

is the contribution of our Galaxy and

$$j_{\rm ext} = C\, 4\pi R_{\rm ext}\overline{n}$$

has extragalactic origin. Here $R_{\rm ext} = R_{\rm Universe} \sim 4$ Gpc for energies below $E_{\rm GZK}$ and $R_{\rm ext} \sim 50$ Mpc for energies above $E_{\rm GZK}$. The constant C is the same in both equations. Following Ref. 9, we note that Eq. (1) yields an estimate of the relative magnitude of these two contributions,

$$\frac{j_{\rm ext}}{j_h} = \alpha \frac{R_{\rm ext}}{R_h}\frac{\overline{n}}{\overline{n_h}} \sim \alpha \frac{R_{\rm ext}}{R_h} \times 10^{-5}, \tag{3}$$

where $R_h \sim 100$ kpc is the size of the Galactic halo and α is a constant of purely geometrical origin,

$$\alpha = \frac{3\int_{r<R_h} d^3x n(x)}{R_h^2 \int_{r<R_h} \frac{d^3x}{x^2} n(x)}. \tag{4}$$

Here $r(x,\theta) = (x^2 + r_0^2 - 2xr_0\cos\theta)^{1/2}$ is the distance between the point of integration and the Galactic center, and $r_0 = 8.5$ kpc is the distance to the Galactic center.

The numerical value of α is close to 1 for a uniform distribution $n(x) = \text{const}$ and can be small for distributions concentrated around the Galactic center. Although $n(x)$ does not have to coincide with the CDM distribution in the halo, we consider as two examples the isothermal halo model[14]

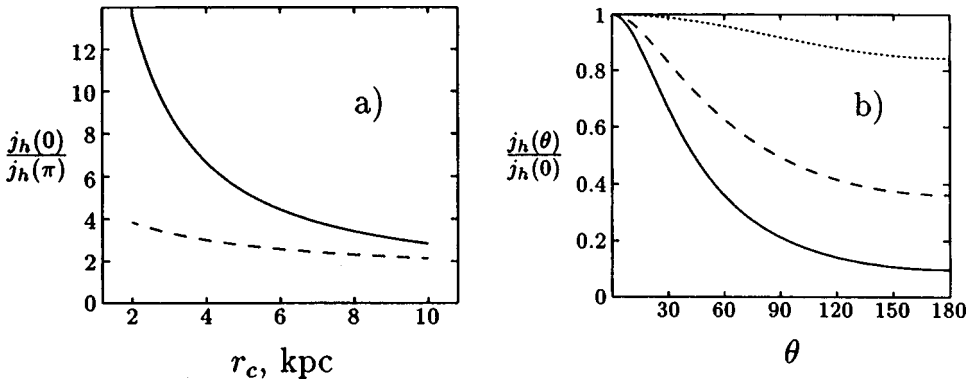

FIG. 1. a: The anisotropy $j_h(0)/j_h(\pi)$ as a function of the core size r_c for the density profiles (5) (solid line) and (6) (dashed line). b: The corresponding angular distributions at $r_c=5$ kpc. The dotted line shows the angular distribution for $n(x)=\text{const}$ (i.e., when the anisotropy is minimum).

$$n(r)\propto\frac{1}{(r_c^2+r^2)} \tag{5}$$

and the more realistic distribution of Ref. 15,

$$n(r)\propto\frac{1}{\sqrt{(r_c^2+r^2)}(R_h+r)^2}, \tag{6}$$

which we have arbitrarily regularized at $r=0$ by introducing the core size r_c. The value of α is $\alpha\simeq 0.15$ for distribution (5) and $\alpha\simeq 0.5$ for distribution (6), with no strong dependence on r_c in the range $r_c=2-10$ kpc.

From Eq. (3) we find

$$\frac{j_{\text{ext}}}{j_h}\sim\alpha \quad \text{for } E<E_{\text{GZK}}, \quad \frac{j_{\text{ext}}}{j_h}\sim 10^{-2}\times\alpha \quad \text{for } E>E_{\text{GZK}}. \tag{7}$$

Therefore, at $E<E_{\text{GZK}}$ the Galactic and extragalactic contributions can be comparable (although the Galactic one is probably somewhat larger), while at $E>E_{\text{GZK}}$ the extragalactic part is suppressed by a factor $\sim 10^{-2}$. In either case a substantial fraction of the observed UHE CRs should come from the halo of our Galaxy. In this respect our conclusions agree with that of Ref. 9.

The Galactic part of the total UHE CR flux, j_h, is anisotropic due to our position at 8.5 kpc from the center of the Galaxy. The anisotropy can be obtained from Eq. (2),

$$j_h(\theta)\propto\int dx\, n(r(x,\theta)).$$

Figure 1 shows the anisotropy $j_h(0)/j_h(\pi)$ as a function of the core radius for the trial distributions (5) and (6). For $n(x)=\text{const}$ the anisotropy is minimum and constitutes about 20%. Figure 1 shows the corresponding angular dependences of $j_h(\theta)$ at $r_c=5$ kpc. As can be seen from the picture, the anisotropy of the Galactic contribution is at

least $\sim 20\%$ and can be much larger if $n(x)$ is concentrated around the Galactic center. Also, it should be noted that the anisotropy depends exclusively on $n(x)$ and does not depend on energy, since cosmic rays with energy $E \sim E_{\mathrm{GZK}}$ are deflected by only $\sim 3^\circ$ by the Galactic magnetic field.[5]

In turn, the extragalactic contribution j_{ext} consists of an isotropic part coming from distances $R \gg 50$ Mpc, which is comparable in magnitude to j_h and is present only at $E < E_{\mathrm{GZK}}$, and the contribution from our ''neighborhood'' $R \lesssim 50$ Mpc. The latter should have peaks in the direction of nearby galaxies and clusters. The contribution of such a peak, δj_{ext}, equals

$$\frac{\delta j_{\mathrm{ext}}}{j_h} = \alpha \frac{R_h^2}{3R^2} \frac{M}{M_G},$$

where R is the distance to the astronomical object, M is its mass, and M_G is the mass of our Galaxy including the halo. For instance, contributions from the Andromeda Nebula and Virgo Cluster are comparable and close to $10^{-2} \times \alpha$, in agreement with Eq. (7) and Ref. 9.

Since at energies above the GZK cutoff the extragalactic contribution is negligible, non-observation of the anisotropy at the level of $\sim 20\%$ would rule out the CDM related mechanisms of UHE CRs. Observation of the Galactic anisotropy would enable one to reconstruct the density profile $n(x)$ and, possibly, the distribution of CDM in the Galactic halo.

At energies below the GZK cutoff, the anisotropy is smaller on account of the relative enhancement of the isotropic extragalactic part. Since anisotropy does not depend on energy and can be measured at $E > E_{\mathrm{GZK}}$, it is possible, in principle, to determine the magnitude of the extragalactic contribution. Provided the CDM related mechanisms are dominant at $E \lesssim E_{\mathrm{GZK}}$ and the coefficient α is known, the ratio j_h / j_{ext} could, in view of Eqs. (1) and (3), give important information about the distribution of matter in the Universe.

The current data are insufficient for drawing definite conclusions about the angular distribution of the highest-energy cosmic rays, on account of both the very limited statistics and the absence of data in the Southern Hemisphere, where the Galactic center is situated. However, since the anisotropy predicted by the CDM-related mechanisms is large, it will either be observed or excluded already in the next generation of experiments.[11] Among these the Pierre Auger Project has the best potential, due to the large number of expected events (600–1000 events with $E > 10^{20}$ eV in 10 years) and the ability to see both hemispheres.

The authors are grateful to D. S. Gorbunov, V. A. Kuzmin, V. A. Rubakov, M. V. Sazhin, and D. V. Semikoz for helpful discussions. This research is supported in part by Award No. RP1-187 of the U. S. Civilian Research & Development Foundation for the Independent States of the Former Soviet Union (CRDF) and by Russian Fund for Fundamental Research, Grants 96-02-17804a and 96-02-17449a.

[a)]We thank V. A. Kuz'min for drawing our attention to this point.

[1] N. Hayashida, K. Honda, M. Honda *et al.*, Phys. Rev. Lett. **73**, 3491 (1994); D. J. Bird, S. C. Corbato, H. Y. Dai *et al.*, Astrophys. J. **424**, 491 (1994); Astrophys. J. **441**, 144 (1995); T. A. Egorov *et al.*, *Proceedings of Tokyo Workshop on Techniques for the Study of Extremely High Energy Cosmic Rays*, edited by M. Nagano, ICRR, U. of Tokyo, 1993.
[2] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966); G. T. Zatsepin and V. A. Kuz'min, JETP Lett. **4**, 78 (1966).
[3] V. S. Berezinsky, Sov. J. Nucl. Phys. **1**, 222 (1970); R. J. Protheroe and P. L. Biermann, Astropart. Phys. **6**, 45 (1996).
[4] E. Waxman, Phys. Rev. Lett. **75**, 386 (1995); K. Mannheim, Astropart. Phys. **3**, 295 (1995); J. P. Rachen and P. L. Biermann, Astron. Astrophys. **272**, 161 (1993).
[5] E. Waxman and L. Miralda-Escude, Astrophys. J. **472**, L89 (1996).
[6] C. T. Hill, D. N. Schramm, and T. P. Walker, Phys. Rev. D **36**, 1007 (1987); G. Sigl, D. N. Schramm, and P. Bhattacharjee, Astropart. Phys. **2**, 401 (1994); V. Berezinsky, X. Martin, and A. Vilenkin, Phys. Rev. D **56**, 2024 (1997); for a review see G. Sigl, ''Topological defect models of ultra-high energy cosmic rays,'' http://xxx.lanl.gov/abs/astro-ph/9611190.
[7] J. Ellis, J. L. Lopez and D. V. Nanopoulos, Phys. Lett. B **247**, 257 (1990); J. Ellis, G. B. Gelmini, J. L. Lopez *et al.*, Nucl. Phys. B **373**, 399 (1992).
[8] V. A. Kuzmin and V. A. Rubakov, ''Ultra-high energy cosmic rays: a window to post-inflationary reheating epoch of the Universe?,'' http://xxx.lanl.gov/abs/astro-ph/9709187.
[9] V. Berezinsky, M. Kachelriess, and A. Vilenkin, Phys. Rev. Lett. **79**, 4302 (1997).
[10] V. Kuzmin and I. Tkachev, ''Ultrahigh-energy cosmic rays, superheavy long living particles, and matter creation after inflation,'' http://xxx.lanl.gov/abs/hep-ph/9802304.
[11] J. W. Cronin, Nucl. Phys. Proc. Suppl. **28B**, 213 (1992); *The Pierre Auger Observatory Design Report* (2nd ed.), March 1997; S. C. Corbato, H. Y. Dai, J. W. Elbert *et al.*, Nucl. Phys. Proc. Suppl. **28B**, 36 (1992).
[12] S. White, J. Navarro, A. Evard, and C. Frenk, Nature (London) **366**, 429 (1993).
[13] P. Bhattacharjee and G. Sigl, Phys. Rev. D **51**, 4079 (1995).
[14] J. Caldwell and J. Ostriker, Astrophys. J. **251**, 61 (1981); J. Bahcall *et al.* Astrophys. J. **265**, 730 (1983).
[15] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996).

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

Search for anomalous carbon atoms — evidence of violation of the Pauli principle during the period of nucleosynthesis

A. S. Barabash and V. N. Kornoukhov
''Institute of Theoretical and Experimental Physics'' State Science Center of the Russian Federation, 117259 Moscow, Russia

Yu. M. Tsipenyuk
P. L. Kapitsa Institute of Physics Problems, Russian Academy of Sciences, 117334 Moscow, Russia

B. A. Chapyzhnikov
V. I. Vernadskiĭ Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, 117975 Moscow, Russia

(Submitted 22 May 1998; resubmitted 11 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 104–108 (25 July 1998)

A search for anomalous carbon atoms $^{12}\tilde{\mathrm{C}}$ — atoms with three K-shell electrons — is made by γ-activation analysis of boron with the impurity carbon removed radiochemically. A limit on the existence of such atoms was determined: $^{12}\tilde{\mathrm{C}}/^{12}\mathrm{C} < 2.5\times 10^{-12}$. This corresponds to a limit on the lifetime with respect to violation of the Pauli principle by electrons in a carbon atom of $\tau \geqslant 2\times 10^{21}$ yr. © *1998 American Institute of Physics.* [S0021-3640(98)00214-X]

PACS numbers: 32.10.−f, 82.80.Jp

The Pauli principle is one of the most fundamental laws of nature. It was formulated by W. Pauli in 1925[1] to explain the regularities of the periodic system of elements and the characteristic features of atomic spectra. In modern quantum field theory the Pauli principle follows automatically from the principle of the identical nature of particles and the anticommutativity of the fermion creation (annihilation) operators. It consists in the fact that in a system of identical fermions two or more particles cannot occupy the same state.

The discovery in 1956 of the nonconservation of parity in β-decay[2] showed for the first time that the fundamental laws can be violated. Subsequently, as is well known, the violation of CP invariance was discovered in 1964.[3] As a result of these works, virtually all conservation laws are now being tested. For example, the possible nonconservation of leptonic quantum numbers, nonconservation of baryonic and electric charge, and violation of CPT symmetry and Lorentz invariance have been investigated intensively during the last 10–15 years.

As concerns the Pauli principle, there is no answer to the following question on the basis of modern theoretical ideas: ''To what accuracy does the Pauli principle hold?'' This is mainly due to the absence of any ''genuine'' (self-consistent and noncontradic-

tory) models admitting a weak violation of the Pauli principle, and it is quite likely that such models must fall outside the framework of the standard quantum field theory.

Okun' notes[4] that, irrespective of some model or another, it is of interest to make an experimental search for a possible violation of the Pauli principle in two directions.

1. A search for transitions of the type $2p-1s$, as a result of which a third electron appears in the $1s$ shell, in stable atoms. In 1974 Reines and Sobel[5] made the first estimate, based on a result obtained in Ref. 6 from a search for instability of the electron, of the lower limit on the time for processes involving violation of the Pauli principle in iodine atoms: $\tau > 10^{20}$ yr. Using modern limits on the stability of the electron,[7,8] even more stringent limits can be obtained for the violation of the Pauli principle in iodine and germanium atoms: $T_{1/2} > 1.2\times 10^{23}$ yr and $T_{1/2} > 2.6\times 10^{23}$ yr, respectively. It should be underscored that if the Hamiltonian possesses permutation symmetry, then spontaneous conversion of an ordinary atom into a ''non-Pauli'' atom is impossible on the basis of quantum mechanics (see, for example, Refs. 9 and 10).

2. Searches for stable ''non-Pauli'' atoms. Such atoms could be of cosmological origin, if not all 10^{80} electrons in the universe are antisymmetrized or if spontaneous transitions of ordinary atoms into ''non-Pauli'' atoms are nevertheless possible. The chemical properties of atoms with three electrons per $1s$ shell must be similar to the properties of their ''lower-order'' neighbors in the periodic table (for example, ''non-Pauli'' carbon is similar to boron).

In 1989 Novikov and Pomanskiĭ proposed a check of the Pauli principle by searching for anomalous atoms arising during the period of nucleosynthesis in the universe.[11] If the Pauli principle is violated, then every substance containing elements with the atomic number Z contains an admixture of anomalous atoms of the element with atomic number $(Z+1)$, since these anomalous atoms have the same chemical properties as the element with the atomic number Z. The concentration of anomalous atoms in a substance is highest in the case that the cosmic abundance of the parent element $(Z+1)$ is high, while that of the element Z is low. If the formation of ''non-Pauli'' atoms occurred as a result of a spontaneous transition of an outer electron into an inner shell, then the concentration of anomalous atoms in the main material will be

$$C = \frac{t}{\tau}\frac{P(Z+1)}{P(Z)}, \tag{1}$$

where t is the average time which has passed since the moment when the anomalous atom formed up to completion of nucleosynthesis ($\sim 4.5\times 10^9$ years); τ is the lifetime of an atomic electron with respect to violation of the Pauli principle; and, $P(Z+1)$ and $P(Z)$ are the cosmic abundances of elements with atomic numbers $(Z+1)$ and Z.

In Ref. 11 two pairs of atoms were proposed as most promising for investigation: boron–carbon and fluorine–neon, with the ratios $P(Z+1)/P(Z) = 2.18\times 10^6$ and 650, respectively. An experimental search for anomalous neon atoms in fluorine and argon atoms in chlorine has been performed[12] (the results obtained by accelerator mass spectrometry are $\tau(^{20}\mathrm{Ne}) > 2\times 10^{30}$ yr and $\tau(^{36}\mathrm{Ar}) > 4\times 10^{27}$ yr), whiles the boron–carbon pair has remained unstudied.

This work is devoted to a search for anomalous carbon atoms in boron. We recall that boron exists in the form of two stable isotopes, ^{10}B ($\sim$19%) and ^{11}B($\sim$81%).

Carbon likewise consists of two isotopes, ^{12}C($\sim$99%) and ^{13}C($\sim$1%). The anomalous atom $^{12}\tilde{\text{C}}$ contains three *K*-shell electrons and therefore behaves chemically like a boron atom. Such anomalous atoms should be concentrated in boron and its compounds in the process of evolution. In order to observe them a search must be conducted for anomalous ^{12}C nuclei in boron or for boron atoms with a nuclear mass of 12.

In the present work we endeavored to implement the second possibility. In discussing possible experiments in search of anomalous carbon atoms in boron, Novikov and Pomanskiĭ proposed mass spectrometry and even estimated the sensitivity of experiments of this kind. However, their estimate is much too high, because a quite large admixture of ''ordinary'' carbon (at the 0.1% level) is always present in boron samples and in the residual atmosphere of the mass spectrometer.

The idea of this experiment was to remove carbon atoms chemically from a boron sample and then to measure the content of carbon nuclei in it. The search for anomalous carbon atoms was conducted by γ-activation analysis of different boron samples. Boron is an ideal object for the γ-activation method of analysis, since irradiation with γ rays does not produce any radioactive isotopes in it and the entire induced activity is thereby wholly determined by the impurities alone.

The experiment was performed on the microtron at the Institute of Physics Problems of the Russian Academy of Sciences. The boron samples for analysis consisted of amorphous boron (powder, obtained from BCl_3 gas), α- and β-rhombohedral boron, boron whiskers on tungsten, decaborane ($B_{10}H_{14}$), and boron grown by zone melting. Experiments showed that the purest boron was obtained by zone melting, and boron from that source was used for the main measurements. In these samples the initial content of carbon atoms was equal to approximately 1.5×10^{-3} g/g. In other samples (amorphous boron, whiskers, boron anhydride) large quantities of chlorine and fluorine ($\sim10^{-4}$ g/g) as well as small quantities of different metal impurities (mainly at the 10^{-6}–10^{-8} g/g level) were found. For this reason, it is impossible to search for anomalous carbon atoms in boron by the instrumental method (nondestructive γ-spectrometry of irradiated samples), and a special method was developed for removing impurities from the boron.

A boron sample with a mass of $\sim$0.1 g was irradiated with γ-ray bremsstrahlung from 28-MeV electrons. If ^{12}C nuclei are present in the boron, then the reaction ^{12}C(γ,n)^{11}C should occur, i.e., radioactive ^{11}C nuclei with a half-life of 20.34 min are formed. In addition, in 99% of the cases ^{11}C undergoes β^+ decay and can therefore be detected by the γ–γ coincidence method (the detection of two 511 keV annihilation γ rays, which are emitted after the positrons stop and are annihilated in the material).

It should be noted that on emitting a neutron a nucleus acquires a recoil momentum, and as it moves through the material it can lose electrons, including *K*-shell electrons. Then anomalous atoms will become normal atoms as the electronic shell is subsequently filled. However, in the reaction ^{12}C(γ,n)^{11}C, which we employed, the energy of the ^{11}C recoil nucleus does not exceed 1 MeV (average energy $\sim$300 keV). According to Ref. 13, at an energy of 600 keV the probability that a carbon atom loses a *K*-shell electron does not exceed 10%. The probability of *K*-shell electrons being lost in the transfer of recoil energy to a neutral atom (on account of the ''shakeup'' of the electronic shell

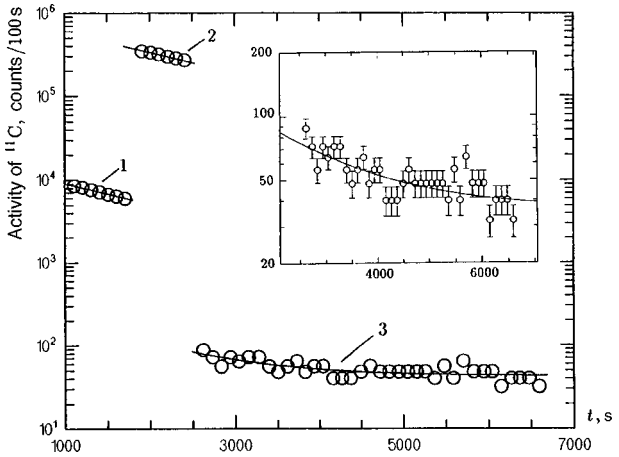

FIG. 1. Time dependence of the positron activity of carbon separated from a 100 mg boron sample after irradiation (*1*), a 6 mg carbon comparison sample irradiated together with the experimental sample (*2*), and a solution after chemical separation of carbon atoms (*3*). Inset: The detected positron activity of the solution. The solid line is a fit of the data assuming that only ^{11}C radionuclides and the background are present.

accompanying transfer of the recoil momentum to the nucleus) was also estimated. The probability of such a process does not exceed several percent.

After irradiation, the boron sample was oxidized in a KNO_3-KOH (3:1) melt for about 10 min at 600 °C. In the process of oxidative melting in an alkaline melt, the carbon-containing impurities in boron (a solid solution of substitution of carbon, i.e., boron carbide B_4C) are oxidized to CO_2, CO, $KHCO_3$, and K_2CO_3, while boron is oxidized to B_2O_3. The cooled melt was then dissolved in water containing nitric acid to convert the boron anhydride into boric acid H_3BO_3. Next, the solution was heated to the boiling point. In the process the potassium bicarbonates and carbonates decomposed, and carbon dioxide gas was released. The carbon dioxide was removed from the boric acid solution by an air flow. Experiments on measuring the activity of the released $^{11}CO_2$ showed that in 25 min such a procedure removes more than 99% of the carbon from the solution. After the CO_2 was driven off, the change in the activity of ^{11}C nuclei remaining in the boric acid solution was measured as a function of time. Graphite was irradiated simultaneously with the sample. Measurements of the activity of the graphite made it possible to calculate the mass of the remaining carbon.

The plot displayed in Fig. 1 shows the result of one of the measurements performed. The residual activity is due essentially only to the background (it remains at approximately 40 decays per 100 s) and only a slight elevation in activity at the beginning of the measurements can be attributed to radioactive carbon nuclei with a 20-min half-life. The inset in the figure displays on a large scale the result of measurements of the activity of the solution and the result of mathematical analysis of the data using the function $a+b\exp(-t/T_{1/2})$, where a is the constant background and $T_{1/2}=1224$ s is the half-life of ^{11}C.

A series of measurements of this kind established that the concentration of anomalous carbon atoms does not exceed 5×10^{-6} g/g. The point is that we cannot rule out the

possibility that the observed activity of ^{11}C nuclei is due to the residual content of ''normal'' carbon in the boron sample. For this reason, it can only be asserted that the concentration of anomalous $^{12}\tilde{\text{C}}$ atoms in boron is $\leqslant 5\times 10^{-6}$g/g. Since the cosmic abundance of carbon is 2×10^{6} times greater than that of boron, this means that the relative concentration of anomalous carbon atoms in carbon corresponds to $^{12}\tilde{\text{C}}/^{12}\text{C}\leqslant 2.5\times 10^{-12}$. Using the limit obtained, we find from Eq. (1) that since anomalous carbon atoms could have accumulated over the entire period of nucleosynthesis $t=4.5\times 10^{9}$ yr, the lifetime of electrons in a carbon atom relative to violation of the Pauli principle is

$$\tau\geqslant 2\times 10^{21}\ \text{yr}. \tag{2}$$

In closing, we note that the sensitivity of the γ-activation method, which we used to determine the ^{12}C concentration in boron, would in principle permit increasing the limit of determination of τ by an order of magnitude if one could obtain a boron sample with a low content of ''normal'' carbon impurity or if the carbon could be effectively removed from the boron sample after irradiation.

We are deeply grateful to L. B. Okun' for his interest in this work and for a discussion of the results, to V. É. Yants and I. O. Pilyugin for assistance in estimating the probability of *K*-shell electrons being lost, to A. I. Kozyrev for providing the boron samples, to K. A. Solntsev for providing the decaborane samples and for fruitful discussions of boron chemistry, and to B. S. Zakirov for assisting in the experiments.

[1] W. Pauli, Z. Phys. **31**, 765 (1925).
[2] C. S. Wu, E. Ambler, R. W. Hayward *et al.*, Phys. Rev. **105**, 1413 (1957).
[3] J. H. Christinsen, J. W. Cronin, V. L. Fitch *et al.*, Phys. Rev. Lett. **13**, 138 (1964).
[4] L. B. Okun', Usp. Fiz. Nauk **158**, 294 (1989) [Sov. Phys. Usp. **32**, 543 (1989)]; Phys. Rev. D **45**, VI.10 (1992).
[5] F. Reines and H. Sobel, Phys. Rev. Lett. **32**, 954 (1974).
[6] M. K. Moe and F. Reines, Phys. Rev. **140**, B992 (1965).
[7] H. Ejiri, H. Kinoshita, H. Sano, and H. Ohsumi, Phys. Lett. B **282**, 281 (1992).
[8] Y. Aharonov, F. T. Avignone III, R. L. Brodzinski *et al.*, Phys. Lett. B **353**, 168 (1995).
[9] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Clarendon Press, Oxford, 1958, Chap. IX.
[10] R. D. Amado and H. Primakoff, Phys. Rev. C **22**, 1338 (1980).
[11] V. M. Novikov and A. A. Pomanskiĭ, JETP Lett. **49**, 81 (1989).
[12] V. M. Novikov, A. A. Pomansky, T. Faestermann *et al.*, Phys. Lett. B **240**, 227 (1990).
[13] K. Shima, N. Kuno, M. Yamanoudri, and H. Tawara, Atomic Data Nuclear Data Tables **51**, 173 (1992).

Translated by M. E. Alferieff

String representation of $SU(3)$ gluodynamics in the Abelian projection

D. A. Komarov
Moscow Physicotechnical Institute, 141700 Dolgoprudnyĭ, Moscow Region, Russia; Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia

M. N. Chernodub
Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia

(Submitted 2 June 1998; resubmitted 18 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 109–114 (25 July 1998)

A dual Ginzburg–Landau model corresponding to $SU(3)$ gluodynamics in the Abelian projection is studied. A string theory describing QCD string dynamics is obtained in this model. The interaction of static quarks in mesons and baryons is investigated in an approximation to leading order. © *1998 American Institute of Physics.* [S0021-3640(98)00314-4]

PACS numbers: 12.38.−t, 11.25.−w

One approach to the problem of color confinement in quantum $SU(N)$ gluodynamics is the method of Abelian projections, proposed by 't Hooft.[1] This method is based on partial gauge fixing, which does not fix the Abelian gauge subgroup $[U(1)]^{N-1}$. The diagonal elements of the gluon field transform under such Abelian transformations as gauge fields, while the off-diagonal elements transform as matter vector fields. Since the $SU(N)$ group is compact, its Abelian subgroup is also compact, and Abelian monopoles exist in the system. If the monopoles are condensed, then confinement can be explained at the classical level:[2,3] A string forms between color charges (quarks) which is a (dual) analog of an Abrikosov string in a superconductor, the role of the Cooper pairs being played by the monopoles.

The confinement mechanism described above, which is often called the ''dual superconductor mechanism,'' has been confirmed by numerous computer calculations on a lattice (see, for example, the review in Ref. 4). For example, it has been shown that the contribution of Abelian monopoles to the tension of the string is almost completely identical to the total tension of a string, the monopole currents satisfy the London equation for a superconductor, and the condensate of Abelian monopoles is different from zero in the confinement phase and strictly equals zero in the deconfinement phase. Although these results were all obtained for $SU(2)$ gluodynamics, there are grounds for believing that the dual superconductor model should also work in the more realistic case of $SU(3)$ gluodynamics.[4] A substantial difference of the $SU(3)$ theory from $SU(2)$ theory is the presence of two independent string configurations — dual analogs of an Abrikosov string in a superconductor. In the present letter we study string degrees of

freedom and investigate the interaction of quarks in $SU(3)$ dynamics on the basis of the dual superconductor model (the dual Ginzburg–Landau model).

In Euclidean space, the dual Ginzburg–Landau model[5] corresponding to $SU(3)$ gluodynamics is described by the Lagrangian

$$\mathcal{L}_{\mathrm{DGL}}=\frac{1}{4}(\partial_\mu \mathbf{B}_\nu-\partial_\nu \mathbf{B}_\mu)^2+\sum_{i=1}^{3}\left[|(i\partial_\mu-g\boldsymbol{\epsilon}_i\cdot\mathbf{B}_\mu)\chi_i|^2+\lambda(|\chi_i|^2-v^2)^2\right]. \tag{1}$$

This Lagrangian contains two Abelian gauge fields, $\mathbf{B}_\mu=(B^3_\mu,B^8_\mu)$, which are the duals to the gluon fields A^3 and A^8, belonging to the Cartan subgroup of the $SU(3)$ gauge group. The model (1) also contains three Higgs fields: $\chi_k=\rho_k e^{i\theta_k}$, $k=1,2,3$, whose phases are related by the relation

$$\theta_1+\theta_2+\theta_3=0\,. \tag{2}$$

The Higgs fields χ_i correspond to the monopole fields, the monopoles being condensed since $\lambda>0$ and $v^2>0$. The Abelian charges of the Higgs fields with respect to the gauge fields B^3_μ and B^8_μ are determined by the root vectors of the group $SU(3)$: $\boldsymbol{\epsilon}_1=(1,0)$, $\boldsymbol{\epsilon}_2=(-1/2,-\sqrt{3}/2)$, and $\boldsymbol{\epsilon}_3=(-1/2,\sqrt{3}/2)$; we also used the notation $(\mathbf{a})^2=(\mathbf{a},\mathbf{a})$, where $(\mathbf{a},\mathbf{b})=a^3b^3+a^8b^8$. The Lagrangian (1) is invariant under $[U(1)]^2$ gauge transformations:

$$B^a_\mu\to B^a_\mu+\partial_\mu\alpha^a\,, \qquad a=3,8\,;$$
$$\theta_i\to\theta_i+g(\boldsymbol{\epsilon}_i,\boldsymbol{\alpha}) \bmod 2\pi\,, \quad i=1,2,3\,,$$

where $\boldsymbol{\alpha}=(\alpha^3,\alpha^8)$ are the parameters of the gauge transformation.

The model (1) contains vortex configurations which are analogous to the Abrikosov–Nielsen–Olesen (ANO) strings[6] in the Abelian Higgs model (AHM). In the AHM, the electrically charged Higgs fields Φ are condensed and the ANO strings carry a quantized magnetic flux. In a circuit around such a string the phase $\theta=\arg\Phi$ of the Higgs field acquires an increment $\theta\to\theta+2\pi n$, where n is an integer (the number of elementary fluxes inside the string). Thus, at the center of the string the phase of the Higgs field is singular and therefore the Higgs field equals zero: $\mathrm{Im}\,\Phi=\mathrm{Re}\,\Phi=0$. These last two equations define a two-dimensional manifold in a four-dimensional space, which is the world surface of the center of the string.

Three string degrees of freedom $\Sigma^{(i)}$, which correspond to three Higgs fields χ_i, $i=1,2,3$, can be introduced into the model (1). Since in this model the condensed Higgs fields possess magnetic charge, the strings $\Sigma^{(i)}$ carry an electric flux and, as we shall see below, they possess nonzero tension, which results in color confinement. By analogy to the Abelian Higgs model[7] the strings $\Sigma^{(i)}$ are determined by the equations

$$\partial_{[\mu,}\partial_{\nu]}\theta_i(x,\tilde{x})=2\pi\tilde{\Sigma}^{(i)}_{\mu\nu}(x,\tilde{x}^{(i)}), \qquad \tilde{\Sigma}^{(i)}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\Sigma^{(i)}_{\alpha\beta}\,, \tag{3}$$

$$\Sigma^{(i)}_{\alpha\beta}(x,\tilde{x}^{(i)})=\int_{\Sigma^{(i)}}d^2\sigma\epsilon^{ab}\frac{\partial\tilde{x}^{(i)}_\alpha}{\partial\sigma_a}\frac{\partial\tilde{x}^{(i)}_\beta}{\partial\sigma_b}\delta^{(4)}[x-\tilde{x}^{(i)}(\sigma)]\,,$$

where the vector $\tilde{x}=\tilde{x}^{(i)}(\sigma)$ describes the singularities of the phases θ_i parametrized by σ_1 and σ_2, and the tensor $\Sigma^{(i)}_{\mu\nu}$ determines the position of the singularities. We note that $\partial_{[\mu}\partial_{\nu]}\theta_i \neq 0$, since the phases θ_i are singular functions.

The string world surfaces $\Sigma^{(i)}$ are not independent, since the phases of the Higgs field are connected by relation (2):

$$\Sigma^{(1)}_{\mu\nu}+\Sigma^{(2)}_{\mu\nu}+\Sigma^{(3)}_{\mu\nu}=0\,. \tag{4}$$

According to numerical estimates,[8] the parameter λ in the Lagrangian of the dual Ginzburg–Landau model (1) is quite large, $\lambda\sim 65\gg 1$. For this reason, in what follows we shall consider the string degrees of freedom in the London limit $\lambda\to\infty$, which corresponds to the leading approximation to within corrections of order $O(\lambda^{-1})$.

In the London limit, the radial degrees of freedom of the Higgs fields are frozen at their vacuum values, $\chi_i=v$. Therefore the dynamical variables are the phases θ_i of the Higgs fields and the dual gauge fields B[3,8]. Then the path integral of the model assumes the form

$$\mathcal{Z}=\int_{-\infty}^{+\infty}\mathcal{D}B\int_{-\pi}^{+\pi}\mathcal{D}\theta_i\,\exp\left\{-\int d^4x\,\mathcal{L}_{\mathrm{DGL}}(B,\theta)\right\}, \tag{5}$$

where

$$\mathcal{L}_{\mathrm{DGL}}(B,\theta)=\frac{1}{4}(\partial_\mu \mathbf{B}_\nu-\partial_\nu \mathbf{B}_\mu)^2+v^2\sum_{i=1}^{3}(\partial_\mu\theta_i+g\boldsymbol{\epsilon}_i\cdot\mathbf{B}_\mu)^2\,. \tag{6}$$

Integrating over the fields $\mathbf{B}_\mu$ and $\theta^{(i)}$ in the manner of Refs. 9 and 7, we obtain

$$\mathcal{Z}=\int\mathcal{D}\boldsymbol{\Sigma}_{\mu\nu}\,\delta\left(\sum_{i=1}^{3}\Sigma^{(i)}_{\mu\nu}\right)\exp\{-S_{str}(\boldsymbol{\Sigma})\}, \tag{7}$$

$$S_{str}(\boldsymbol{\Sigma})=2\pi^2v^2\int d^4x\,d^4y\sum_{i=1}^{3}\Sigma^{(i)}_{\mu\nu}(x)\mathcal{D}_{m_B}(x-y)\Sigma^{(i)}_{\mu\nu}(y), \tag{8}$$

where $m_B^2=3g^2v^2$ is the mass of the gauge fields B^3 and B^8, $\mathcal{D}_{m_B}$ is the propagator of the massive field: $(\Delta+m_B^2)\mathcal{D}_{m_B}(x)=\delta^{(4)}(x)$, and we have introduced the notation for the string variables $\boldsymbol{\Sigma}=(\Sigma^{(1)},\Sigma^{(2)},\Sigma^{(3)})$. The integration measure $\mathcal{D}\boldsymbol{\Sigma}$ contains the Jacobian[7] of the transformation from the field variables (B^a_μ,θ_i) to the string variables $(\boldsymbol{\Sigma})$.

It is helpful to rewrite the string action (8) in terms of the independent string variables $\Sigma^{(1)}$ and $\Sigma^{(2)}$ using relation (4):

$$\begin{aligned}S_{str}(\boldsymbol{\Sigma})=4\pi^2v^2\int d^4x\,d^4y\{&\Sigma^{(1)}_{\mu\nu}(x)\mathcal{D}_{m_B}(x-y)\Sigma^{(1)}_{\mu\nu}(y)\\&+\Sigma^{(1)}_{\mu\nu}(x)\mathcal{D}_{m_B}(x-y)\Sigma^{(2)}_{\mu\nu}(y)+\Sigma^{(2)}_{\mu\nu}(x)\mathcal{D}_{m_B}(x-y)\Sigma^{(2)}_{\mu\nu}(y)\}\,.\end{aligned} \tag{9}$$

In this formulation, one can see that the model contains two types of strings, which repel one another when the electric fluxes in them are parallel and attract one another when the fluxes are antiparallel.

An interesting problem is to find the interaction potential of quarks at rest (infinitely heavy quarks). Since the Ginzburg–Landau model under study is dual to $SU(3)$ gluodynamics, the interaction potential $V_c(R)$ of quarks q_c and $q_{\bar{c}}$ is determined by the average of the 't Hooft loop:

$$V_c(R) = -\lim_{T\to+\infty} \frac{1}{T} \ln\langle H_c(\mathcal{C}_{R\times T})\rangle\,, \tag{10}$$

$$\langle H_c(\mathcal{C})\rangle = \frac{1}{\mathcal{Z}} \int_{-\infty}^{+\infty} \mathcal{D}B \int_{-\pi}^{+\pi} \mathcal{D}\theta_i \exp\left\{ -\int d^4x \left[\frac{1}{4} (\partial_\mu \mathbf{B}_\nu - \partial_\nu \mathbf{B}_\mu - \mathbf{Q}^{(c)} \Sigma^{\mathcal{C}}_{\mu\nu})^2 + v^2 \sum_{i=1}^{3} (\partial_\mu \theta_i + g \boldsymbol{\epsilon}_\alpha \cdot \mathbf{B}_\mu)^2 \right] \right\}, \tag{11}$$

where the contour $\mathcal{C}_{R\times T}$ is a rectangular loop of size $R\times T$, representing the trajectories of the quark and antiquark. The surface $\Sigma^{\mathcal{C}}_{\mu\nu}$ is a string whose boundary is the trajectory $\mathcal{C}$:

$$\partial_\mu \Sigma^{\mathcal{C}}_{\mu\nu}(x) = j^{\mathcal{C}}_\nu(x)\,, \quad j^{\mathcal{C}}_\mu(x) = \oint_{\mathcal{C}} d\tau \frac{\partial \tilde{x}_\mu(\tau)}{\partial \tau} \delta^{(4)}(x - \tilde{x}(\tau))\,, \tag{12}$$

where the vector $\tilde{x}_\mu$ parametrizes the trajectory $\mathcal{C}$. The quark q_c (antiquark $q_{\bar{c}}$) carries color charges $\mathbf{Q}^{(c)}$ ($\mathbf{Q}^{(\bar{c})} = -\mathbf{Q}^{(c)}$), which assume the values

$$\mathbf{Q}^{(c)} = (Q_3^{(c)}, Q_8^{(c)}) = \left\{ \left(\frac{e}{2}, \frac{e}{2\sqrt{3}} \right), \left(-\frac{e}{2}, \frac{e}{2\sqrt{3}} \right), \left(0, -\frac{e}{\sqrt{3}} \right) \right\} \tag{13}$$

for red ($c=R$), blue ($c=B$), and green ($c=G$) quarks, respectively. Here $e = 4\pi/g$ is the elementary Abelian electric charge.

Integrating in expression (11) over the fields $\mathbf{B}_\mu$ and θ_i, we obtain

$$\langle H_c(\mathcal{C})\rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}\boldsymbol{\Sigma}^{\mathcal{C}}_{\mu\nu} \delta\left(\sum_{i=1}^{3} \Sigma^{(i)}_{\mu\nu} \right) \exp\left\{ -\pi^2 v^2 \int d^4x\, d^4y \times \sum_{i=1}^{3} \left(\Sigma^{\mathcal{C},(i)}_{\mu\nu}(x; s_i^{(c)}) \mathcal{D}_{m_B}(x-y) \Sigma^{\mathcal{C},(i)}_{\mu\nu}(y; s_i^{(c)}) + \frac{2}{m_B^2} j^{\mathcal{C}}_\mu(x; s_i^{(c)}) \mathcal{D}_{m_B}(x-y) j^{\mathcal{C}}_\mu(y; s_i^{(c)}) \right) \right\}, \tag{14}$$

where for string surfaces of the type i, which have the contour $\mathcal{C}$ as the boundary, we have introduced the notation

$$\Sigma^{\mathcal{C},(i)}_{\mu\nu}(x; s_i^{(c)}) = s_i^{(c)}\, \Sigma^{\mathcal{C}}_{\mu\nu}(x) + \Sigma^{(i)}_{\mu\nu}(x)\,, \quad j^{\mathcal{C}}_\mu(x; s_i^{(c)}) = \partial_\nu \Sigma^{\mathcal{C},(i)}_{\mu\nu}(x; s_i^{(c)})\,. \tag{15}$$

These variables satisfy the relation

$$\sum_{i=1}^{3} \Sigma^{\mathcal{C},(i)}_{\mu\nu}(x; s_i^{(c)}) = 0\,. \tag{16}$$

TABLE I.

c	R	B	G
$s_1^{(c)}$	1	-1	0
$s_2^{(c)}$	-1	0	1
$s_3^{(c)}$	0	1	-1

The quantities $s_i^{(c)}$ possess a simple meaning: a quark of color c is the boundary for $s_i^{(c)}$ strings of type i. If $s_i^{(c)}<0$, then the corresponding string carries a negative flux. The quantities s are presented in Table I, according to which only colorless states can have a finite mass: If quarks form a colored configuration, then there exists a string which carries off the flux of the color field from this configuration to infinity. The energy of such a string is infinite, since the string possesses nonzero tension.

As a result of the condition (16), the quarks in the pairs $R-\overline{R}$, $B-\overline{B}$, and $G-\overline{G}$ are connected to one another either by two oppositely directed strings of two different types or only by any one of these strings, depending on which of the strings is integrated over to remove the δ function in Eq. (14). This will not affect the physical quantities (for example, the interaction potential of the quarks), since the tension of two oppositely directed strings of different types which are ''stuck together'' will be identical to the tension of one string. Likewise, the interaction potential of the quarks q_c and q_c^-does not depend on the color c. For this reason, we shall find the potential between the $G-\overline{G}$ pair at rest, having integrated in expression (14) over the string $i=3$:

$$\langle H_{c=G}(\mathcal{C})\rangle = \frac{1}{\mathcal{Z}}\int \mathcal{D}\Sigma_{\mu\nu}^{(1)}\mathcal{D}\Sigma_{\mu\nu}^{\mathcal{C},(2)}\exp\Bigg\{-2\pi^2 v^2\int d^4x\, d^4y \times\Bigg(\Sigma_{\mu\nu}^{(1)}(x)\mathcal{D}_{m_B}(x-y)\Sigma_{\mu\nu}^{(1)}(y)+\Sigma_{\mu\nu}^{(1)}(x)\mathcal{D}_{m_B}(x-y)\Sigma_{\mu\nu}^{\mathcal{C},(2)}(y) +\Sigma_{\mu\nu}^{\mathcal{C},(2)}(x)\mathcal{D}_{m_B}(x-y)\Sigma_{\mu\nu}^{\mathcal{C},(2)}(y)+\frac{2}{m_B^2}j_\mu^{\mathcal{C}}(x)\mathcal{D}_{m_B}(x-y)j_\mu^{\mathcal{C}}(y)\Bigg)\Bigg\}, \tag{17}$$

where the string world surface $\Sigma_{\mu\nu}^{(1)}(x)\equiv\Sigma_{\mu\nu}^{\mathcal{C},(1)}(x;0)$ is closed, while the string $\Sigma_{\mu\nu}^{\mathcal{C},(1)}(x)\equiv\Sigma_{\mu\nu}^{\mathcal{C},(1)}(x;1)$ has as its boundary the quark–antiquark trajectory.

It is well known that in the limit of large mass M_H of the scalar particle the tension σ of an ANO string becomes large (σ increases as $\ln M_H\sim\ln\lambda$). Therefore it can be expected that in the London limit the potential $V_c(R)$ is determined by the minimum of the action in Eq. (17). This minimum is reached on a configuration in which the string of the type $i=1$ is absent and the world surface of the string of type $i=2$ forms a minimal surface stretched on the contour $\mathcal{C}$:

$$\Sigma^{(1)}(x)=0\,,\quad \Sigma^{\mathcal{C},(2)}(x)=\delta(x_2)\,\delta(x_3)\,\theta(x_1)\,\theta(R-x_1)\,, \tag{18}$$

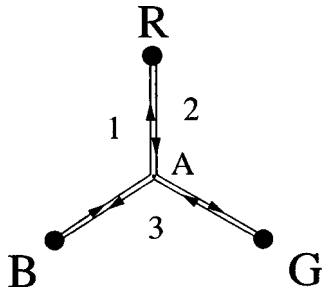

FIG. 1. Configuration of QCD strings in a baryon. The letters R, G, and B represent the quark colors; the numbers enumerate the types of strings.

the quarks being at rest at the points (0, 0, 0) and $(R,\ 0, 0)$. Substituting expressions (18) into Eq. (17) and using Eq. (10), we obtain in the momentum representation (the projection of the momentum on an axis connecting the quarks is denoted by p_1)

$$V_{cl}(R) = -\frac{2e^2}{3}\int \frac{d^3p}{(2\pi)^3}\sin^2\left(\frac{p_1 R}{2}\right)\left[\frac{1}{p^2+m_B^2}+\frac{m_B^2}{p^2+m_B^2}\frac{1}{p_1^2}\right]. \tag{19}$$

This expression is distinguished by a general numerical factor from the expression obtained in Ref. 10 for the interaction potential of static quarks in the $U(1)$ model, corresponding to the Abelian projected $SU(2)$ gluodynamics. The first term in Eq. (19) corresponds to the exchange of a massive vector boson and leads only to a Yukawa potential; the second term is of string origin. Proceeding similarly to Ref. 10, we obtain (dropping an inconsequential additive constant)

$$V_{cl}(R) = -\frac{e^2}{12\pi}\left\{\frac{e^{-m_B r}}{r}+r m_B^2 \ln\left(\frac{m_\chi}{m_B}\right)+4m_B e^{-m_B r}+4r m_B^2 \mathrm{Ei}(-m_B r)\right\}, \tag{20}$$

$$\mathrm{Ei}(x) = -\int_{-x}^{+\infty}\frac{e^{-t}}{t}dt,$$

where $\mathrm{Ei}(x)$ is the exponential integral function. In Eq. (20) we cut off the diverging integral at energies $p^2 \sim m_\chi^2 = 2\lambda v^2$, making the assumption that λ is finite (but large), which corresponds to taking the finite size of the string into account.[7,9] At small distances ($r \ll m_B^{-1}$) this potential is a Coulomb potential, while at large distances ($r \gg m_B^{-1}$) it is linearly increasing and therefore leads to quark confinement. We note that the coefficient of the term which grows linearly at large distances (the tension of the quantum-chromodynamic string) is identical to the result obtained in Ref. 11 by a different method.

The representation introduced above makes it possible to analyze the case of a bound state of three quarks. In accordance with Table I, they are connected by strings of all three types, as shown in Fig. 1.

The arrangement of the strings was chosen so that they would satisfy the condition (16). If the quarks are located at the vertices of an equilateral triangle, then the configuration where the point A lies at the center of the triangle gives a minimum of the energy. This is easy to see by integrating the δ-function in Eq. (14) and obtaining for the hadron the string action in the form (8). In this representation the quarks in a hadron are con-

nected by only two of the three strings shown in Fig. 1. Thus, if the strings attract one another, then they will ''fuse together'' on a certain segment RA. The energy of the system will be proportional to the sum of the lengths of the segments $RA+BA+GA$, whence it follows that the classical configuration corresponds to the position of the point A at the center. This result is qualitatively in agreement with the conclusions drawn in Ref. 12 on the basis of numerical calculations.

In summary, the classical limit of the string representation introduced for $SU(3)$ gluodynamics in the present work on the basis of the dual $[U(1)]^2$ Ginzburg–Landau model can serve as a good approximation for describing quark–antiquark interaction in $SU(3)$ gluodynamics.

We thank M. I. Polikarpov for helpful remarks. This work was supported in part by Grants Nos. 96-02-17230a and 96-15-96740 of the Russian Fund for Fundamental Research and by Grants INTAS-RFBR-95-0681 and INTAS-96-370.

[1] G. 't Hooft, Nucl. Phys. B **190**, 455 (1981).
[2] S. Mandelstam, Phys. Rep. C **23**, 245 (1976).
[3] G. 't Hooft, in *High Energy Physics*, *Proceedings of the EPS International Conference*, Palermo, Italy, 23–28 June 1975, edited by A. Zichichi, Compositori, Bologna, 1976.
[4] M. I. Polikarpov, Nucl. Phys. B (Proc. Suppl.) **53**, 134 (1997); M. N. Chernodub and M. I. Polikarpov, Lectures given at NATO Advanced Study Institute on Confinement, Duality and Nonperturbative Aspects of QCD, Cambridge, England, 23 June – 4 July 1997; http://xxx.lanl.gov/abs/hep-th/9710205.
[5] S. Maedan and T. Suzuki, Prog. Theor. Phys. **81**, 229 (1989).
[6] A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957) [Sov. Phys. JETP **5**, 1174 (1957)]; H. B. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973).
[7] E. T. Akhmedov *et al.*, Phys. Rev. D **53**, 2087 (1996).
[8] Y. Koma, H. Suganuma, and H. Toki, http://xxx.lanl.gov/abs/hep-ph/9804289; Talk at *INNOCOM '97, XVII RCNP the International Symposium on Innovative Computational Methods in Nuclear Many-Body Problems*, November 1997, Osaka, Japan.
[9] P. Orland, Nucl. Phys. B **428**, 221 (1994).
[10] F. V. Gubarev, M. I. Polikarpov, and V. I. Zakharov, Preprint ITEP-TH-28/98; http://xxx.lanl.gov/abs/hep-th/9805175.
[11] H. Suganuma, S. Sasaki, and H. Toki, Nucl. Phys. B **435**, 207 (1995).
[12] S. Kamizawa, Y. Matsubara, H. Shiba, and T. Suzuki, Nucl. Phys. B **389**, 563 (1993).

Translated by M. E. Alferieff

Resonance fluorescence in an atom + dielectric microsphere system excited by a single photon

V. V. Klimov
P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924, Moscow, Russia

V. S. Letokhov
Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

(Submitted 22 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 115–120 (25 July 1998)

The resonance interaction of a two-level atom with a continuum of free-space modes modified by the presence of a dielectric microsphere (modified free-space modes — MFSMs) is studied. In the case that quantized MFSMs are initially excited within the contour of one of the resonance modes of the microsphere, the spectrum of emitted photons depends strongly on the excitation method. Under optimal excitation conditions efficient excitation of the atom accompanied by the formation of a Rabi doublet in the fluorescence spectrum occur. As the excitation conditions depart from optimality, the spectrum becomes a triplet. If the departure from optimality of excitation is large, the atom remains essentially unexcited, and the fluorescence spectrum has a singlet character. © *1998 American Institute of Physics.*
[S0021-3640(98)00414-9]

PACS numbers: 32.50.+d, 32.80.−t

There are several reasons for the great interest in investigations of the interaction of atoms with a dielectric microsphere. The most important one is that a microsphere is a resonator with a high Q in the optical range.[1,2] The most interesting regime is that of strong atom–microsphere interaction. In this case either the atom or the microsphere-resonator can be excited first.[3] In the present letter we report the results of an investigation of the dynamics of the system atom + microsphere when the microsphere is initially excited by a single photon.

The procedure for quantizing the electromagnetic field in a spherical geometry is, by and large, well known.[4] In our problem, an ideally conducting sphere of finite but large radius $\Lambda \to \infty$ (Fig. 1) can be taken as the quantization volume. The transverse electric field operator can be represented in the form[4]

$$\hat{\mathbf{E}} = \sum_s \frac{a_s \mathbf{e}(s,\mathbf{r}) - a_s^+ \mathbf{e}^*(s,\mathbf{r})}{i\sqrt{2}}. \tag{1}$$

where a_s and a_s^+ are operators which annihilate and create photons in MFSMs with frequencies ω_s and which satisfy the standard commutation relations.

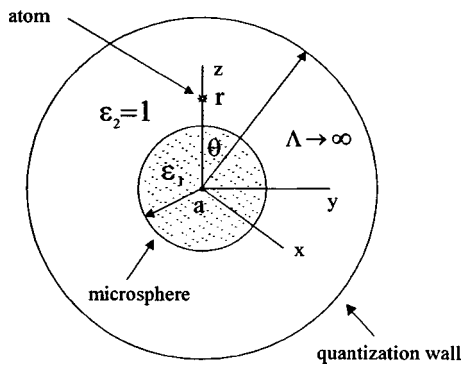

FIG. 1. Geometry of the quantum problem of the interaction of a two-level atom with a dielectric microsphere.

For definiteness, we shall consider the interaction of a radially oriented dipole moment of an atom inside a microsphere with a TM whispering-gallery mode.

It is easy to express the intensity of the electric field $\mathbf{e}(s,\mathbf{r})$ of the sth TM mode in terms of the spherical harmonics (Y_{nm}) and the spherical Bessel and Hankel functions (j and h, respectively):

$$\mathbf{e}_{TM}(n,m,\nu)=\begin{cases} -\dfrac{1}{k_0\epsilon_1}\nabla\times\{[\beta_{TM,n}J_n(k_1r)]\hat{\mathbf{L}}Y_{nm}(\vartheta,\varphi)\}, & r<a, \\ -\dfrac{1}{k_0\epsilon_2}\nabla\times\{[\alpha^{(1)}_{TM,n}h^{(1)}_n(k_2r)+\alpha^{(2)}_{TM,n}h^{(2)}_n(k_2r)]\hat{\mathbf{L}}Y_{nm}(\vartheta,\varphi)\}, & r>a. \end{cases} \tag{2}$$

In Eq. (2) n, m, and ν are, respectively, the orbital, azimuthal, and radial quantum numbers, $k_{1,2}=\sqrt{\epsilon_{1,2}}\,k=\sqrt{\epsilon_{1,2}}\,\omega(\nu)/c$ are the wave vectors inside and outside the sphere, $\mathbf{L}=-i\mathbf{r}\times\nabla$ is the angular momentum operator, and a is the radius of the microsphere. The set of quantum numbers (n,m,ν) forms the vector index $s=(n,m,\nu)$, used above.

The coefficients α_n and β_n can be found in the standard manner from the continuity of the tangential components of the field at the boundary of the sphere and the normalization of the wave functions in a sphere of radius Λ to one photon per quantized mode:

$$\alpha^{(1)}_{TM}/\alpha^{(2)}_{TM}=1-2q_n; \quad \frac{\beta_{TM}}{\alpha^{(2)}_{TM}}=\frac{2i\epsilon q_n}{ka(\epsilon(z_2j_n(z_2))'j_n(z_1)-(z_1j_n(z_1))'j_n(z_2)}, \tag{3}$$

$$q_n=\frac{\left[\epsilon\dfrac{d}{dz_2}[z_2j_n(z_2)]j_n(z_1)-\dfrac{d}{dz_1}[z_1j_n(z_1)]j_n(z_2)\right]}{\left[\epsilon\dfrac{d}{dz_2}[z_2h^{(1)}_n(z_2)]j_n(z_1)-\dfrac{d}{dz_1}[z_1j_n(z_1)]h^{(1)}_n(z_2)\right]}, \tag{4}$$

$$|\alpha^{(1)}_{TM,n}|^2=|\alpha^{(2)}_{TM,n}|^2=\frac{2\pi\hbar c}{\Lambda}\,\frac{k^3}{n(n+1)}. \tag{5}$$

Here q_n are the Mie reflection coefficients, $z_1=\sqrt{\epsilon}ka$, $z_2=ka$, and ϵ is the dielectric constant of the microsphere.

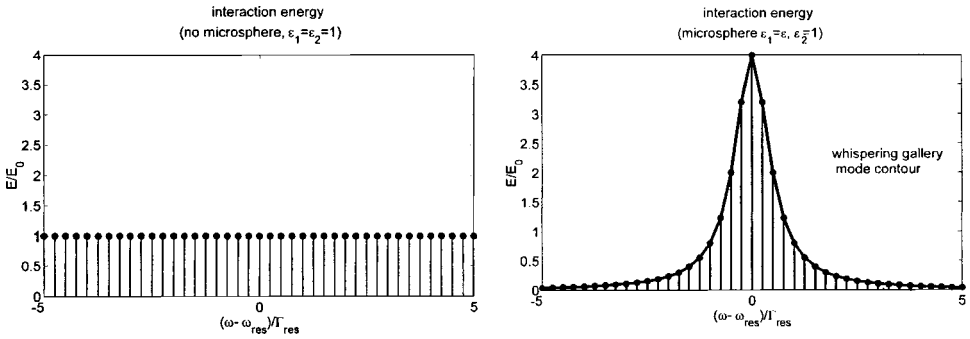

FIG. 2. Illustration of the resonance growth of the interaction energy of an atom interacting with quantized modes modified by the presence of a dielectric microsphere.

The frequencies of the quantized modes and the density of final states can be found from the requirement that the tangential components of the electric field of the TM modes vanish on the inner surface of the quantization sphere:

$$\omega_s = \left(\nu + \frac{n+1}{2}\right)\frac{\pi c}{\Lambda} + \ldots, \quad \rho_{TM}(\omega) = d\nu/d(\hbar\omega_s) = \Lambda/\pi\hbar c. \tag{6}$$

If the resonance conditions are satisfied, the energy density inside the dielectric microsphere increases considerably, i.e., whispering gallery modes arise.[5] Figure 2 shows qualitatively the dependence of the energy of the atom–MFSM interaction energy in the resonance and nonresonance cases. Our main problem is to give a correct treatment of the interaction of an atom with quantized modes falling within the contour of the whispering gallery mode.

Near resonance the Mie reflection coefficient assumes the simple form

$$q_n \approx -i\,\mathrm{Im}(\Omega_{\mathrm{res}})/(\omega - \Omega_{\mathrm{res}}), \tag{7}$$

where $\Omega_{\mathrm{res}} = \omega_{\mathrm{res}} - i\Gamma_{\mathrm{res}}^{(n)}/2$ is a complex frequency characterizing the resonance mode ($\Gamma_{\mathrm{res}}^{(n)}$ characterizes the radiation width of the resonance mode).

In the case of a resonance interaction of a two-level atom with the whispering-gallery modes, the dipole interaction Hamiltonian for a two-level atom and radiation can be represented in the standard form.[6]

In the rotating-wave approximation the ground state and the one-photon, two-photon, etc. continua do not interact and can be treated separately. In the present letter we shall confine our attention to the one-phonon continuum. The corresponding Hamiltonian has the form

$$H_1 = \begin{pmatrix} \hbar\omega_A & V_1 & V_2 & V_3 \\ V_1^* & \hbar\omega_1 & 0 & 0 \\ V_2^* & 0 & \hbar\omega_2 & 0 \\ V_3^* & 0 & 0 & \hbar\omega_3 \end{pmatrix}. \tag{8}$$

Here ω_A is the frequency of the atomic transition with allowance for the purely electrostatic frequency shift due to the interaction with the microsphere,[7,8] while the interaction matrix elements V_i have the form

$$V_i = V(\omega_i) = -f d_{\rm rad} e_{\rm rad}(n, m=0, \nu_i, \mathbf{r})/i\sqrt{2}. \tag{9}$$

In Eq. (9) $f = 3\epsilon/(2\epsilon+1)$ is a local-field factor, $d_{\rm rad}$ is the radial component of the transition dipole moment, $e_{\rm rad}$ is the radial component of the electric field of the quantized mode (2), and the coordinate system is chosen so that the atom lies on the axis of the system. The result is that the atom interacts only with the $m=0$ TM mode.

The state vector of our system has the following structure:

$$|\text{excited atom,} \quad \text{photon with energy } \omega_1, \quad \text{photon with energy } \omega_2 \ldots\rangle. \tag{10}$$

We shall assume that the microsphere is excited initially and that its excitation energy is distributed over the quantized MFSMs as follows:

$$\psi_A = 0; \quad \psi_j(t=0) = \psi_{0,j} = \psi_0(\omega_j); \quad \sum_j |\psi_{0,j}|^2 = 1. \tag{11}$$

The average energy of the superposition of states (11) equals $\hbar\omega_{\rm res}$, irrespective of the choice of the distribution $\psi_{0,j}$, while the spatial distribution of this energy (the spatial correlation function of the electric field) depends strongly on $\psi_{0,j}$. Under the initial conditions (11) the equations for the Fourier components of the probability amplitudes assume the form

$$(\omega-\omega_A)\psi_A(\omega) = \sum_j \frac{V_j}{\hbar}\psi_j(\omega), \quad (\omega-\omega_j)\psi_j(\omega) = \frac{V_j^*}{\hbar}\psi_A(\omega) - i\psi_{0,j}, \quad j = 1,2,\ldots \tag{12}$$

Let us examine first the photon emission spectrum for an arbitrary shape of the excitation line of the microsphere. From Eq. (12) it is easy to find an expression for the fluorescence spectrum in the limit $t\to\infty$:

$$\frac{dP(\omega)}{d(\hbar\omega)} = \rho_{TM}(\omega)\left|\frac{V^*(\omega)}{\hbar}\psi_A(\omega) - i\psi_0(\omega)\right|^2. \tag{13}$$

In our case the matrix elements $V(\omega)$ have a Lorentzian form (see Eq. (7))

$$V(\omega) \approx \alpha/(\omega - \Omega_{\rm res}). \tag{14}$$

Let us assume that the excitation line is also Lorentzian:

$$\psi_{0,j} = \psi_0(\omega_j) = \beta/(\omega_j - \Omega_\psi^*). \tag{15}$$

Substituting expressions (14) and (15) into Eq. (13), we obtain for the fluorescence spectrum the expression

$$\frac{dP(\omega)}{d(\hbar\omega)} = \rho_{TM}(\omega)\left|\frac{\alpha^*}{\omega-\Omega_{\rm res}^*}\frac{\psi_A(\omega)}{\hbar} - i\frac{\beta}{\omega-\Omega_\psi^*}\right|^2. \tag{16}$$

In order for a Rabi doublet to form, in Eq. (16) there can be no poles between the components of the doublet, i.e., there can be no poles at $\omega = \Omega_{\rm res}$ and $\omega = \Omega_\psi$. This is

possible only in the optimal case, if the complex resonance frequency of the external excitation source equals the complex resonance frequency of the microsphere:

$$\Omega_\psi = \Omega_{\rm res}. \tag{17}$$

If the condition (17) is not satisfied, the additional poles do not cancel and the fluorescence spectrum will have either a triplet or singlet form.

To obtain specific analytical results, let the initial distribution of the excitation over MFSMs have the form

$$\psi_0(\omega) = \frac{1}{\hbar\Omega_{\rm Rabi}}\sqrt{\frac{\Gamma_\psi}{\Gamma_{\rm res}}}\frac{\alpha^*}{\omega-\Omega_\psi^*}, \quad \Omega_\psi = \omega_\psi - i\frac{\Gamma_\psi}{2}, \quad \Gamma_\psi > 0, \tag{18}$$

where the radial Rabi frequency has the form

$$\hbar^2\Omega_{\rm Rabi}^2 = f^2\frac{d_{\rm rad}^2}{2}n(n+1)(2n+1)\frac{\pi\hbar\Gamma_{\rm res}^{(n)}}{2\epsilon^2 r^3}J_{n+1/2}^2(\sqrt{\epsilon}\,k_{\rm res}r)\left(\frac{Y_{n+1/2}(k_{\rm res}a)}{J_{n+1/2}(\sqrt{\epsilon}k_{\rm res}a)}\right)^2. \tag{19}$$

Eliminating the photon probability amplitudes from Eq. (12) and performing the frequency integration, we obtain the following expression for the Fourier components of the atomic probability amplitude:

$$\psi_A(\omega) = -\frac{\Omega_{\rm Rabi}}{D(\omega)}\frac{\sqrt{\Gamma_{\rm res}\Gamma_\psi}}{\Omega_{\rm res}-\Omega_\psi^*}, \quad D(\omega) = (\omega-\omega_A)(\omega-\Omega_{\rm res}) - \Omega_{\rm Rabi}^2. \tag{20}$$

This equation shows that the amplitude of the excited state of an atom reaches maximum values at $\Omega_\psi = \Omega_{\rm res}$. Under this condition the excitation energy is concentrated inside the sphere to the maximum degree and the atom is completely excited, as a result of which the fluorescence spectrum acquires a doublet structure.

If the condition (17) is not satisfied, then the atom will not be completely excited, and a central component will appear in the fluorescence spectrum together with the Rabi components, i.e., a triplet will form. Next, substituting expression (20) into Eq. (16), we find the fluorescence spectrum:

$$\frac{dP(\omega)}{d(\hbar\omega)} = \frac{\Gamma_\psi}{2\pi\hbar}\left|\frac{\Omega_{\rm Rabi}^2}{(\omega-\Omega_{\rm res}^*)D(\omega)}\frac{\Gamma_{\rm res}}{\Omega_{\rm res}-\Omega_\psi^*} + i\frac{1}{\omega-\Omega_\psi^*}\right|^2. \tag{21}$$

The triplet structure of the spectrum appears here explicitly. We note that the triplet structure of expression (21) is not a Mollow triplet,[9] since in the present work we are studying the case of a one-photon MFSM continuum, while the Mollow spectrum appears at high intensities.

From Eq. (21) the following expression is easily obtained for the spectrum of emitted photons in the optimal case (17):

$$\frac{dP(\omega)}{d(\hbar\omega)} = \frac{(\omega-\omega_A)^2\Gamma_{\rm res}/2\pi\hbar}{[(\omega-\omega_A)(\omega-\omega_{\rm res})-\Omega_{\rm Rabi}^2)]^2 + (\omega-\omega_A)^2\Gamma_{\rm res}^2/4}. \tag{22}$$

An important feature of Eq. (22) is that it always has a doublet character, even when the interaction is weak, i.e., for the Wigner–Weisskopf exponential decay regime.

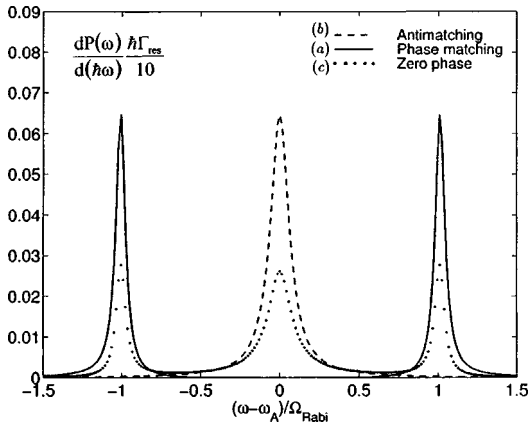

FIG. 3. The dependence of the fluorescence spectrum on the method of resonance excitation ($\omega_A=\omega_{\rm res}$) of a microsphere under the condition $|\Omega_\psi|=|\Omega_{\rm res}|$. a) $\Omega_\psi=\Omega_{\rm res}$, when the maximum possible concentration of photon energy in the microsphere is attained, i.e., under the condition (17); b) $\Omega_\psi^*=\Omega_{\rm res}$, when the minimum possible concentration of photon energy in the microsphere is attained, i.e., under the condition (23); c) intermediate case (26). For all cases ($r/a=0.8, \omega_0/\gamma_0=5\times10^7$, radial orientation of the dipole, TM1,12,12) — whispering-gallery mode, $(ka)_{\rm res}=6.924298$, $\epsilon=6$).

Other cases of excitation of MFSMs can also be investigated the present approach. Specifically, if an initial condition of the form (antimatching with respect to the phase of the MFSMs and the external excitation source)

$$\psi_0(\omega)=\sqrt{\frac{\Gamma_\psi}{\Gamma_{\rm res}}}\frac{1}{\hbar\Omega_{\rm Rabi}}\frac{\alpha^*}{\omega-\Omega_\psi^*},\quad \Omega_\psi=\omega_\psi+i\frac{\Gamma_\psi}{2},\quad \Gamma_\psi>0, \tag{23}$$

is used instead of Eq. (18), then the excitation energy inside the microsphere will be minimal (one-third as high as in the case (18)), as a result of which the Fourier component of the amplitude of the excited state of the atom will be described by the expression

$$\psi_A(\omega)=-\frac{\Omega_{\rm Rabi}}{D(\omega)}\frac{\sqrt{\Gamma_{\rm res}\Gamma_\psi}}{\omega-\Omega_\psi}, \tag{24}$$

which explicitly contains three frequencies (for sufficiently high Rabi frequency).

Correspondingly, we obtain for the fluorescence spectrum in the antimatching case (23) expression ($\omega_A=\omega_{\rm res}$)

$$\frac{dP(\omega)}{d(\hbar\omega)}=\frac{\Gamma_\psi}{2\pi\hbar}\frac{1}{|\omega-\Omega_\psi|^2}, \tag{25}$$

which is always a singlet. The case of a uniform phase distribution

$$\psi_0(\omega)=\frac{1}{\hbar\Omega_{\rm Rabi}}\sqrt{\frac{\Gamma_\psi}{\Gamma_{\rm res}}}\frac{\alpha^*}{|\omega-\Omega_\psi^*|},\quad \Omega_\psi=\omega_\psi-i\frac{\Gamma_\psi}{2},\quad \Gamma_\psi>0 \tag{26}$$

is an intermediate case between (18) and (23).

Figure 3 shows the dependence of the fluorescence spectrum on the degree of phase

matching between the field of the excitation source and the MFSMs. This figure shows clearly how a doublet structure (a), characteristic for optimal matching (17), transforms into a triplet structure (c) for nonoptimal matching (26) and then into a singlet structure (b) for the case of antimatching (23).

We thank the Russian Fund for Fundamental Research and the US Department of Defense (through the University of Arizona) for financial support of this work.

[1] V. B. Braginsky, M. L. Gorodetsky, and V. S. Ilshenko, Phys. Lett. A **137**, 393 (1989).
[2] L. Collot, V. Lefevre, M. Brune *et al.*, Phys. Lett. **23**, 327 (1993).
[3] S. Haroche, ''Cavity quantum electrodynamics,'' in *Fundamental Systems in Quantum Optics*, Elsevier Science Publishers, 1992.
[4] A. S. Davydov, *Quantum Mechanics*, Pergamon Press, New York, 2nd ed., 1976 [Russian original, Nauka, Moscow, 1973].
[5] L. A. Vaĭnshteĭn, *Electromagnetic Waves* [in Russian], Radio i Svyaz', Moscow, 1988.
[6] D. P. Craig and T. Thirunamachandran, *Molecular Quantum Electrodynamics*, Academic Press, New York, 1984.
[7] V. V. Klimov, M. Ducloy, and V. S. Letokhov, J. Mod. Opt. **43**, 549 (1996).
[8] V. V. Klimov, M. Ducloy, V. S. Letokhov, J. Mod. Opt. **43**, 2251 (1996).
[9] B. R. Mollow, Phys. Rev. **188**, 1969 (1969).

Translated by M. E. Alferieff

Self-propagating high-temperature synthesis and formation of quasicrystals in Al/Mn bilayer thin films

V. G. Myagkov and L. E. Bykova
L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036, Krasnoyarsk, Russia

G. N. Bondarenko
Institute of Chemistry and Chemical Technology, Siberian Branch of the Russian Academy of Sciences, 660036, Krasnoyarsk, Russia

(Submitted 9 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 121–124 (25 July 1998)

Self-propagating high-temperature synthesis in Al/Mn bilayer thin films was investigated. It was found that a quasicrystalline phase forms after the passage of self-propagating high-temperature synthesis through the sample. It is shown that after multiple initiation ($n>5$) of self-propagating high-temperature synthesis in a sample the quasicrystalline phase transforms into a stable Al_6Mn phase. It is conjectured that self-propagating high-temperature synthesis can play the main role in the formation of quasicrystals in other film systems as well. © *1998 American Institute of Physics.* [S0021-3640(98)00514-3]

PACS numbers: 81.20.Ka, 61.44.Br, 81.30.Bx

After the discovery of a quasicrystalline phase in the alloy Al–Mn,[1] works describing the formation of this phase in Al/Mn thin films by ion irradiation[2,3] and by the implantation of manganese in an aluminum film[4] appeared. Quasicrystals in other systems are also being investigated on ribbons or films. At the same time, the kinetics of the formation of a quasicrystalline phase is unclear in many respects, and the conditions under which a quasicrystalline phase forms are determined by experimental search.

The present work is devoted to an investigation of self-propagating high-temperature synthesis (SHS) in Al/Mn bilayer thin films and a determination of the conditions under which a quasicrystalline icosahedral (i) phase forms in them. Powder-based SHS is used to obtain divers compounds and has been well studied.[5,6] Film-based SHS exhibits features which distinguish it from powder-based SHS and is only now becoming an object of study.[7,8] The samples used for investigations of SHS in thin films consist of up to 500-nm thick layers of reactants which are deposited successively on a substrate. Initiation of SHS occurs when the temperature T_s of the sample exceeds the initiation temperature T_0. In this case, the combustion front propagates along the surface of the film sample.[7,8] An important feature of SHS in thin films is that it can be of two types. In type-I SHS, after the passage of the wave of SHS the reaction is completed by the formation of compounds. An example of this type is the system Al–Ni, where a Ni_2Al_3 phase forms as a result of the passage of a SHS front, and no traces of pure Ni and Al remain.[8] In type-II

SHS a second wave passes over the surface of the sample after the passage of the SHS wave. The passage of the second wave results in phase separation into the initial reactants. At the same time, metastable phases can form in the sample. Type-II SHS arises in Al/Ge films, where the products of the reaction contain mainly solid solutions of Al and Ge, and a negligible quantity of the metastable AlGe phase forms.[8] The temperature T_{ps} at which the phase-separation front appears equals the initiation temperature T_0 ($T_{ps}=T_0$). Since phase separation arises at substrate temperatures $T_s<T_0$, it is possible to reinitiate SHS in this sample by heating it above the temperature T_0. It follows that SHS can be initiated several times in the same sample. After the first initiation, SHS gives rise to mixing of the layers of reactants. For Al/Ge films the initiation temperature is $T_0=600-550$ K. The system Al–Ge has a simple phase diagram of the eutectic type with a eutectic temperature $T_E=697$ K. The microstructure of Al/Ge films after the passage of a wave of SHS consists of clusters which have a dense-branching morphology (DBM clusters). DBM clusters are investigated in many natural sciences and the kinetics of the formation of DBM clusters is widely studied. In Al/Ge films DBM clusters are fractals (the fractal dimension is $d_f=2$ and clusters consist of a solid solution of germanium) in an aluminum matrix.[8] Phase separation into fractal clusters, DBM clusters containing mainly germanium, and a single-crystal aluminum matrix in Al–Ge films proceeds from the amorphous phase. Different models explaining phase separation have been proposed. However, the mechanisms leading to phase separation remain unclear (see references cited in Ref. 8). It is surmised that multiple initiation of SHS in thin films is an analog of a repeated transition through the eutectic temperature in bulk samples. A very unexpected result is that SHS and phase separation in thin films occur in the solid phase, in contrast to eutectic separation, which proceeds from the liquid phase.

The Al/Mn films were obtained by successive deposition of thin layers of manganese, followed by aluminum, on glass substrates. The chosen ratio of the aluminum layer thickness d_{Al} to the manganese layer thickness d_{Mn} was $d_{\mathrm{Al}}:d_{\mathrm{Mn}}=7:1$, which corresponds to ~ 14 at. % Mn in the sample. The total film thickness was in the range 200–800 nm. The details of the production of film samples and the method of initiating SHS in them are described in Refs. 7 and 8. SHS in Al/Mn films is of the second type and proceeds similarly to SHS in Al/Ge samples. The initiation temperature T_0 is higher than in Al/Ge films and lies in the interval 750–800 K. The characteristic features of the motion of the SHS and phase-separation fronts can be easily observed visually. However, they are hardly noticeable after $n\sim 10$ cycles. The velocity V_{ps} of the phase separation front increases strongly with increasing degree of supercooling $\Delta T=T_0-T_s$, where T_s is the substrate temperature. In the experiments, two extreme values of V_{ps} permitted by the experimental conditions were used:

$$V_{ps}^{\min}=(1-2)\times 10^{-3}\ \text{m/s} \quad \text{and} \quad V_{ps}^{\max}=(0.1-0.2)\ \text{m/s}.$$

In Ref. 9 all of the reflections from the quasicrystalline phase were indexed, and the x-ray diffraction data obtained correspond well to the observed diffraction pattern. In the present work the quasicrystalline phase arising after the passage of a SHS wave in an Al(90 nm)/Mn(630 nm) sample was determined according to the two strongest reflections: $d_1=0.217$ nm and $d_2=0.2065$ nm. The phases formed in the synthesis process were identified on a DRON-4-07 diffractometer at room temperature (CuK_α radiation). Fragments of diffraction patterns obtained from the Al(90 nm)/Mn(630 nm) sample before and after SHS are presented in Fig. 1. The initial sample (Fig. 1a) contains an

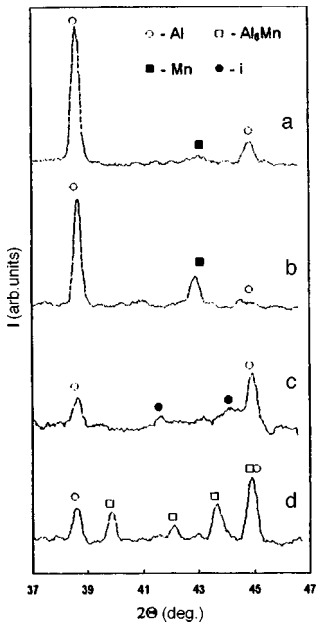

FIG. 1. Diffraction patterns of Al(80 nm)/Ge (630 nm) thin-film samples as a function of the phase separation velocity V_{ps} and the number of SHS cycles n: a) Initial sample, b) $n=1$, $V_{ps}^{\max}=(0.1-0.2)$ m/s, c) $n=1$, $V_{ps}^{\min}=(1-2)\times 10^{-3}$ m/s, d) $n=5$, $V_{ps}^{\max}$ and $n=5$, $V_{ps}^{\min}$.

amorphous layer of α-Mn (broadened peak with $d=0.210$ nm) and a polycrystalline Al film. The post-SHS phase composition depends on both the velocity V_{ps} of the phase separation front and the number n of SHS cycles. For $n=1$ and $V_{ps}^{\max}=(0.1-0.2)$ m/s (Fig. 1b) the sample contains polycrystalline aluminum and manganese. For low velocities $V_{ps}^{\max}=(1-2)\times 10^{-3}$ m/s and $n=1$ (Fig. 1c) the diffraction pattern contains reflections from polycrystalline aluminum and a quasicrystalline icosahedral (i) phase. The intensity of the reflections from aluminum and manganese decreases compared with the initial sample, which suggests formation of a substantial fraction of the quasicrystalline phase. At velocities $V_{ps}^{\max}$ and $n=1$, however, the manganese crystallizes, but there is not enough time for the quasicrystalline phase to form. The diffraction pattern changes radically with multiple initiation of SHS $n>5$ (Fig. 1d). The quasicrystalline phase transforms into an equilibrium Al_6Mn phase, irrespective of the velocity of phase separation.

The microstructure of Al/Mn films consists of clusters in an aluminum matrix and is similar in many respects to the microstructure of Al/Ge films. However, these clusters differ from the DBM clusters in Al/Ge samples, since they are round, 10–20 μm in size, and have a uniform internal structure. Apparently, the same microstructure is described in Ref. 1. In Refs. 2 and 3, layer mixing also was observed in Al/Mn films under ion irradiation. This presupposes that in this case SHS was initiated and participated in the formation of a quasicrystalline phase. The study of SHS in film samples shows that intense diffusion of reactants between layers or grains occurs exclusively at the SHS initiation temperature T_0 and, as we have said, it occurs in the solid phase. The formation of the metastable phases and amorphous state is also surmised[7,8] to occur at temperature

T_0. Since the free energy of the quasicrystalline phase falls between that of the crystal and the amorphous states,[2] the formation of a quasicrystalline phase during SHS is not unexpected. It is surmised that after the passage of the first SHS front with $T_s > T_0$ an amorphous phase forms as a result of the reaction between the aluminum and manganese layers and then decomposes at temperatures $T_s < T_0$, depending on the velocity of the phase separation front and the number of SHS cycles, into the components indicated in Fig. 1. Such a scenario, where a quasicrystalline phase forms from an amorphous phase in Al/Mn films, was studied in Refs. 2 and 3.

As we have said, the initiation temperature of SHS in Al/Ge bilayer film samples is 100–150 K lower than the eutectic temperature for the equilibrium Al–Ge system. The SHS initiation temperature in Al/Mn is also 100–150 K lower than the Al–Mn eutectic temperature of the equilibrium system $T_E = 820$ K.[10] Therefore the SHS initiation temperature T_0 for other systems must also be lower than their eutectic temperatures. Since the motion of the SHS front can be seen visually, it is not difficult to find the temperature T_0. Aluminum enters easily into SHS with many elements of the periodic system,[5] so that SHS should be expected to participate in the formation of aluminum-based quasicrystalline phases.

All this suggests that SHS can play an important role in the formation of quasicrystalline phases, and investigation of the phase composition of the products of reaction in film systems where quasicrystalline and metastable phases form will make it possible to gain a deeper understanding of the physicochemical nature of SHS and the micromechanisms of phase separation in thin films.

[1] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, Phys. Rev. Lett. **53**, 1951 (1984).
[2] D. A. Lilienfeld, M. Nastasi, H. H. Johnson *et al.*, Phys. Rev. Lett. **55**, 1587 (1985).
[3] J. A. Knapp and D. M. Follstaedt, Phys. Rev. Lett. **55**, 1591 (1985).
[4] J. D. Budai and M. J. Aziz, Phys. Rev. B **33**, 2876 (1986).
[5] A. G. Merzhanov, in *Physical Chemistry* [in Russian], collected works edited by Kolotyrkin, Khimiya, Moscow, 1983.
[6] Z. A. Munir and U. Anselmi-Tamburini, Mater. Sci. Rep. **3**, 277 (1989).
[7] V. G. Myagkov and L. E. Byova, Dokl. Akad. Nauk SSSR **354**, 777 (1997).
[8] V. G. Myagkov and L. E. Bykova, JETP Lett. **67**, 334 (1998).
[9] P. A. Bancel, P. Heiney, P. W. Stephens *et al.*, Phys. Rev. Lett. **54**, 2422 (1985).
[10] N. I. Ganina, A. M. Zakharov, V. G. Olenicheva, and L. A. Petrova, *Phase Diagrams of Metallic Systems* [in Russian], No. 41, VINITI, Moscow, 1989.

Translated by M. E. Alferieff

Formation of zero-dimensional hole states during molecular-beam epitaxy of Ge on Si (100)

A. I. Yakimov,[a] A. V. Dvurechenskiĭ, A. I. Nikiforov, and O. P. Pchelyakov
Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 10 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 125–130 (25 July 1998)

The characteristic features of electronic spectra in Ge/Si (100) heterostructures obtained by molecular-beam epitaxy are investigated by capacitance spectroscopy. It is observed that the self-organization of a Ge film into an island film when the effective germanium thickness exceeds six monolayers is accompanied by the appearance of hole bound states, which can be attributed to size quantization and the Coulomb interaction of carriers in the array of Ge quantum dots. © *1998 American Institute of Physics.* [S0021-3640(98)00614-8]

PACS numbers: 81.15.Hi, 73.20.Dx, 68.65.+g

The self-organization of a continuous semiconductor film into an island film during heteroepitaxy of systems with very different lattice constants makes it possible under certain conditions to synthesize a dense array of quantum dots (QDs), 10 nm and smaller in size.[1–4] Most of the experimental and theoretical works[5–8] on the structural and electronic properties of self-organizing QDs have concerned the heterosystem InAs/GaAs. However, the current trend in microelectronics requires the development of methods for producing quantum dots in silicon-based structures. An example of such a heterosystem exhibiting self-organization is Ge/Si.[9–13]

The formation of Ge islands on a Si surface when the Ge film reaches a critical thickness h_c is ordinarily established by HEED,[9] atomic-force microscopy (see, for example, Ref. 13), and high-resolution electron microscopy.[2] Moreover, less direct methods, making it possible to monitor the onset of new features in the carrier spectrum of a Ge film, are used—Raman scattering,[14] photoluminescence,[4,13] and admittance spectroscopy.[3] In the present work it was observed by capacitance spectroscopy that for certain thicknesses of a Ge epitaxial film on Si (100) a zero-dimensional density of states, which is due to size quantization of the hole spectrum in germanium nanocrystals, appears in the heterostructure.

The capacitance spectroscopy of QDs is based on the fact that the charge in zero-dimensional systems can be changed only by a discrete amount $\delta Q = eN$, where e is the electron charge and N is the number of quantum dots in the sample, i.e., the total number of holes in a single bound state.[b] The external voltage V_g on the Al control electrode, biasing the potential in the islands with respect to the Fermi level in the contact, which is separated from the island layer by a tunneling-transparent barrier, stimulates either the

capture of carriers from the contact to QD energy levels or the emptying of these levels, depending on the polarity of V_g (see inset in Fig. 3). When the Fermi level in the contact coincides with the energy of a bound state in a QD, the differential capacitance $C(V_g) = dQ/dV_g$ should possess a peak, attesting to the presence of a discrete energy level. The total capacitance of the structure is a sum of two contributions: The first contribution is due to the presence of a space charge region (SCR) in the carrier-depleted silicon layer, while the second contribution ($C_{\rm QD}$) is due to charging of the quantum dots. Since $C_{\rm QD}$ is proportional to the density of states in the QD, i.e., $C_{\rm QD}=e(d\mu/dV_g)(dN/d\mu)$, where μ is the chemical potential, the quantity $dN/d\mu$ can be recovered from the capacitance–voltage characteristic.

The heterostructures employed in our experiments consisted of the following sequence of layers, starting with the substrate:

— a p^+ silicon substrate with (100) orientation and hole density $p=10^{19}\,{\rm cm}^{-3}$;

— a $Si_{0.8}Ge_{0.2}$ layer of thickness $L=10$ nm, ensuring a sharp heterointerface of the subsequent Si tunneling barrier;

— a tunneling transparent Si barrier, $p=7\times10^{16}\,{\rm cm}^{-3}$, $L=7$ nm;

— a Ge film;

— a Si blocking layer, $p=7\times10^{16}\,{\rm cm}^{-3}$, $L=50$ nm; and,

— an Al control electrode, forming a Schottky barrier at the boundary with silicon; the area of the aluminum layer was equal to $\approx 8\times10^{-3}\,{\rm cm}^2$, and a cylindrical mesostructure of this size was etched out to a depth of the order of 5 μm.

The growth temperature of the Ge film was equal to 300 °C and the growth rate was equal to 0.35 Å/s. The amount of Ge deposited (effective thickness $d_{\rm eff}$) varied from 0 to 20 monatomic layers (ML, 1 ML = 1.4 Å). Figure 1 displays an STM image, obtained *ex situ* under room conditions, of a germanium layer ($d_{\rm eff}=10$ ML). The size of the islands in the growth plane was equal to 13–15 nm, and the average height of the islands was equal to 2.0–2.5 nm. The layer concentration of the islands was equal to $3\times10^{11}\,{\rm cm}^{-2}$, which is almost two orders of magnitude higher than in InAs/GaAs self-organizing structures.[6]

The capacitance–voltage (C–V) characteristics were measured at frequencies of 40–4950 Hz, using phase-sensitive detection of the complex conductivity in a scheme with a capacitive divider.[15] Figure 2 displays the C–V curves obtained at room temperature in structures with different amounts of deposited Ge. The curve for the sample without a germanium layer is featureless and has the form of standard capacitance–voltage characteristic of a *p*-Si depletion layer. Analysis of such a characteristic in the depletion-layer approximation makes it possible to reconstruct the free-carrier density profile in the epitaxial layer (Fig. 3). With allowance for the spatial resolution of the C–V method, which is about 100 Å at an impurity density of the order of $10^{17}\,{\rm cm}^{-3}$, we conclude that the boundary of the depletion layer under reverse bias is close to the heavily doped substrate. For this reason, the barrier capacitance of the structure is virtually voltage-independent.

In the case $d_{\rm eff}=2$ ML, a plateau characteristic for a two-dimensional (2D) carrier gas appears in the voltage range $V_g=0.1-0.3$ V (Fig. 2). A maximum, attesting to the

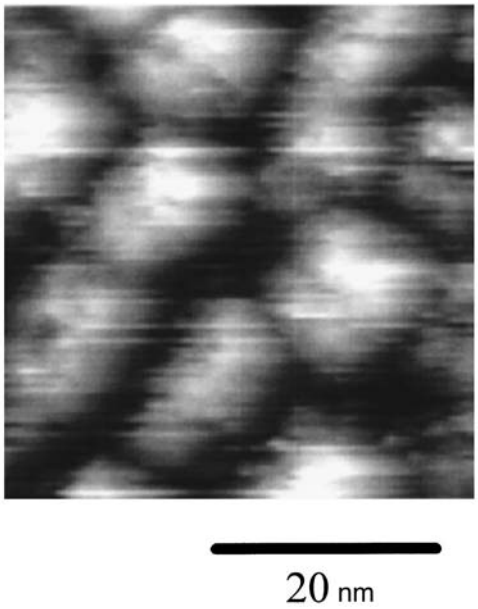

FIG. 1. STM image of the surface of a germanium film ($d_{\text{eff}}=10$ ML).

formation of a continuous Ge film and accumulation of charge in it, is observed in the density profile (Fig. 3).

According to HEED[9] and Raman scattering[14] data, the critical thickness of Ge beyond which pseudomorphic (coherent) islands form under the given growth conditions is $h_c \approx 5$ ML. The capacitance of a sample with $d_{\text{eff}}=6$ ML in the reverse-bias region ($V_g>0$) depends more weakly on V_g than in the case $d_{\text{eff}}=0$ or 2 ML. Apparently here the islands form with a large spread in sizes, and the density of states of the ensemble of

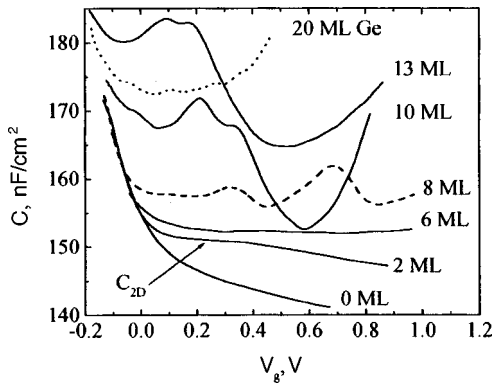

FIG. 2. Capacitance–voltage characteristics of Ge/Si (100) heterostructures with different effective thicknesses of the Ge film. The measurement temperature $T=300$ K.

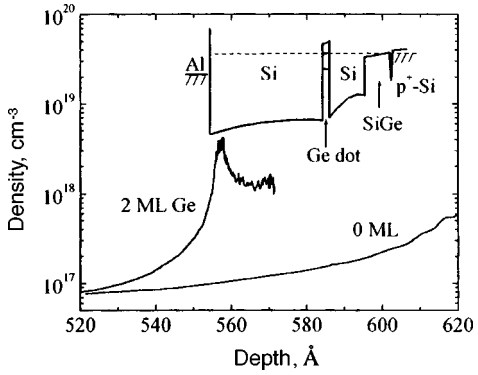

FIG. 3. Hole density profile in the epitaxial layer of a structure without a Ge film and with $d_{\text{eff}}=2$ ML. The profiles were reconstructed from the C–V characteristics in the depletion-layer approximation using the relation $p(x)=-C^3/[e\epsilon\epsilon_0 S^2(dC/dV_g)]$, where the depth x is measured from the surface of the structure and is determined as $x=\epsilon\epsilon_0 S/C(V_g)$. Here $\epsilon\epsilon_0$ is the permittivity and S is the area of the barrier. Inset: Profile of the valence band under reverse bias.

these islands does not have a pronounced δ-function form. In this case the change in capacitance of the space-charge region in Si is simply arrested, since a portion of the voltage-induced charge goes off to bound states in the islands.[d)]

In the range of effective germanium thicknesses $8 \leqslant d_{\text{eff}} \leqslant 13$ ML two peaks appear on the C–V curves at positive polarity. The splitting between these peaks and their width and position on the voltage (energy) scale depend on d_{eff}. The presence of two discrete levels becomes even clearer as the temperature decreases (Fig. 4b). Approximating the contour of the lines, after the capacitance component associated with the SCR is subtracted out, by a Gaussian function $C_{\text{QD}} \propto (\delta V_g \sqrt{\pi})^{-1} \exp[-V_g^2/2(\delta V_g)^2]$, one can find the width δV_g of the peaks and the splitting ΔV_g between them. The results are presented

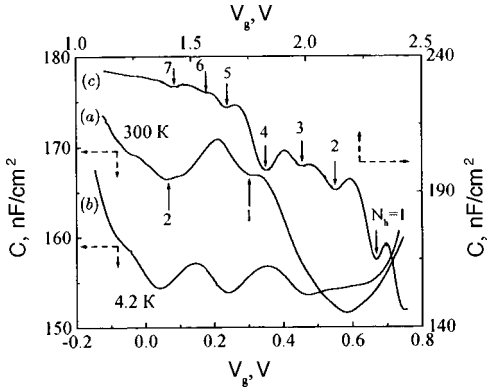

FIG. 4. Capacitance–voltage characteristics of Ge/Si ($d_{\text{eff}}=10$ ML) heterostructures with boron background impurity contents in the epitaxial Si film of $p=7\times10^{16}\,\text{cm}^{-3}$ (curves *a* — $T=300$ K, *b* — $T=4.2$ K) and $p=2\times10^{17}\,\text{cm}^{-3}$ (curve *c* — $T=300$ K). The arrows mark the voltages across the Schottky gate for which the hole levels in the islands are emptied, and the numbers indicate the number of holes in an island.

TABLE I.

d_{eff}, ML	T, K	ΔV_g, mV	δV_g, mV	ΔE, meV	A/e, $10^{11}\,\text{cm}^{-2}$
8	300	342	80	87	6.0
10	300	142	60	36	5.4
13	300	126	56	32	5.8
10	4.2	213	44	54	4.1

in Table I. For δV_g we gove the average value for the two lines. One can see that as d_{eff} increases, the peaks become narrower and the energy gap between them decreases.

We attribute the appearance of capacitance oscillations to the formation of an array of Ge nanocrystals. The nanocrystal sizes in the array are quite uniform and the density of hole states in the array is a δ-like function of energy. As the amount of Ge deposited increases, the islands increase in size. This growth in turn increases the elastic stresses near the Ge/Si interface and therefore increases the energy barrier for incorporation of adatoms diffusing along the island-free surface.[16] This effect induces small clusters to grow more rapidly than large clusters, resulting in the formation of a narrow distribution of island sizes and giving rise to a discrete carrier spectrum which is well-resolved even at room temperature. As the thickness of the Ge film increases further (20 ML), the elastic stress relax, the pseudomorphic phase vanishes, and large ($\approx$1000 Å) islands with dislocations form.[9,14] In our experiment this is observed as vanishing of the capacitance peaks on the capacitance–voltage characteristics (Fig. 2). The strong increase in the capacitance, accompanied by a sharp increase in the active part of the conductance, at voltages of 0.4–0.6 V is also apparently due to the onset of threading dislocations and breakdown of the space-charge region.

The offset of the valence bands of Si and pseudomorphic Ge equals 0.5–0.8 eV.[17] The potential well forming in the valence band of the Ge layer ''absorbs'' holes from the surrounding silicon. The number of holes N_h populating each QD at zero bias can be estimated by taking the ratio of the total number of holes in the epitaxial layer to the total number of islands in the sample: $N_h = pL_T/n_{\text{QD}}$. Here p is the hole density in silicon, L_T is the thickness of the epitaxial layer, and n_{QD} is the layer density of islands. For $p = 7 \times 10^{16}\,\text{cm}^{-3}$, $L_T = 670$ Å, and $n_{\text{QD}} = 3 \times 10^{11}\,\text{cm}^{-2}$ we obtain $N_h \approx 1.5$. This estimate is consistent with the presence of two peaks in the C–V characteristics in the reverse-bias region, since it is for this polarity that the hole-filled states are probed. Figure 4c shows the capacitance–voltage curve for a sample with a high boron impurity content in the epitaxial Si layer ($p \approx 2 \times 10^{17}\,\text{cm}^{-3}$). One can see that a proportional increase occurs in the number of filled states in QDs up to six or seven filled states per island. The stronger change in the capacitance in this case, as compared with the lightly doped sample, is due to the change in the thickness of the depletion layer in the interior of the epitaxial silicon film, while in the first case the boundary of the SCR lies in the buffer layer near the substrate (at $T = 4.2$ K it lies directly in the substrate).

One can see that if the two peaks in the C–V characteristic correspond to two levels in each island, then the area A under the peaks divided by the electron charge e should equal twice the surface density of islands ($2n_{\text{QD}} \approx 6 \times 10^{11}\,\text{cm}^{-2}$). The results of a nu-

merical integration are presented in Table I. One can see that the ratio $A/e \approx 2n_{QD}$ holds to good accuracy.

It is of interest to determine the nature of the splitting of the levels in Fig. 2: Is it due to size quantization or is it a consequence of multiparticle effects of the Coulomb interaction? Deep-level transient spectroscopy (DLTS) investigations of the energy spectrum in Ge islands similar to those shown in Fig. 1 showed that the splitting between the size-quantization levels (between the *s* and *p* shells) equals approximately 100 meV, while the splitting of the levels due to the Coulomb interaction of holes at temperatures above 100 K lies in the range 29–31 meV.[13] The energy gap ΔE between the levels can be estimated from capacitance spectroscopy data as $\Delta E = e\Delta V_g b/L_T$, where b is the distance from the island layer to the substrate. Taking $b = 170$ Å (the total width of the bottom Si layer and the solid solution $Si_{0.8}Ge_{0.2}$) and $L_T = 670$ Å we obtain $\Delta E = 36$ meV for $T = 300$ K and $\Delta E = 54$ meV for $T = 4.2$ K. The values of ΔE for other samples are also presented in Table I. These calculations make it possible to interpret the two capacitance peaks as a hole ground state split by the electron–electron interaction.[c)]

We explain the decrease of ΔE with increasing temperature by the screening of the Coulomb interaction by image charges in the nondepletion region of the SiGe buffer layer and the heavily doped substrate. Indeed, simple estimates show that as temperature increases from 4.2 K to 300 K, the width of the SCR decreases by 40 Å (this correction has virtually no effect on the ratio b/L_T). This signifies the appearance of free carriers in the film of solid solution. According to Ref. 7, the correction to the interaction energy as a result of image forces is given by the expression

$$\delta E = -\frac{e^2}{2\epsilon d} + \frac{2\pi e^2}{\epsilon} n_{QD}\left(\sqrt{4d^2 + r_{min}^2} - r_{min}\right), \tag{1}$$

where d is the distance to the metallic plane and r_{min} is the minimum distance between neighboring islands. The first term in Eq. (1) is a correction to the interaction of carriers inside a quantum dot and the second term is a correction to the interaction of carriers in neighboring QDs. Calculations based on formula (1) for $r_{min} \approx 50$ Å (see Fig. 1), $d(300\text{ K}) = 130$ Å, and $d(4.2\text{ K}) = 170$ Å show that the difference arising between δE at $T = 300$ K and 4.2 K, as a result of the displacement of the SCR by 40 Å equals 19 meV, the second term making the main contribution (18 meV). This agrees with the difference between the experimental values $\Delta E(4.2\text{ K}) - \Delta E(300\text{ K}) = 18$ meV.

The width of the levels changes very little with temperature (see Table I) and apparently is determined by the variance of the island sizes. Since $\delta V_g \approx (0.2\text{–}0.3) \cdot \Delta V_g$, while the interaction energy is inversely proportional to the size of the QDs, the relative nonuniformity of the island sizes in the array is 20–30%, which correlates with microscopy data.

In closing, we wish to call attention to the general rise of the C–V characteristics for $d_{eff} \geqslant 10$ ML. This effect is due to the formation of a quasicontinuous region of surface states with a density of the order of 10^{12} cm^{-2} at the Ge/Si interface.

We thank B. Z. Ol'shanetskiĭ, S. A. Tiĭs, and I. G. Kozhemyako for performing the STM measurements. This work was supported by the Interdisciplinary Scientific and Technical Program ''Physics of Solid-State Nanostructures'' (Grant No. 98-1100) and

the Intercollegiate Scientific Program ''Universities of Russia — Basic Research'' (Grant No. 4103).

[a)]e-mail: yakimov@isp.nsc.ru
[b)]Of course, this requires adequate QD size uniformity.
[c)]In this variant of the structure, the *p*-shell states should be observed in the forward-bias region, but this is prevented by the strong increase in the capacitance as a result of a decrease in the width of the SCR in the epitaxial layer.
[d)]Translation editor's note: As a result of a technical difficulty, this paragraph was cut short in the Russian original; here we have printed a translation of the restored version.

[1] A. I. Yakimov, V. A. Markov, A. V. Dvurechenskii, and O. P. Pchelyakov, Philos. Mag. B **65**, 701 (1992).
[2] A. I. Yakimov, V. A. Markov, A. V. Dvurechenskii, and O. P. Pchelyakov, J. Phys. C **6**, 2573 (1994).
[3] S. K. Zhang, H. J. Zhu, F. Lu *et al.*, Phys. Rev. Lett. **80**, 3340 (1998).
[4] C. S. Peng, Q. Huang, W. Q. Cheng *et al.*, Phys. Rev. B **57**, 8805 (1998).
[5] G. Medeiros-Ribeiro, D. Leonard, and P. M. Petroff, Appl. Phys. Lett. **66**, 1767 (1995).
[6] R. J. Warburton, C. S. Durr, K. Karrai *et al.*, Phys. Rev. Lett. **79**, 5282 (1997).
[7] G. Medeiros-Ribeiro, F. G. Pikus, P. M. Petroff, and A. L. Efros, Phys. Rev. B **55**, 1568 (1997).
[8] B. T. Miller, W. Hansen, S. Manus *et al.*, Phys. Rev. B **56**, 6764 (1997).
[9] V. A. Markov, A. I. Nikiforov, and O. P. Pchelyakov, J. Cryst. Growth **175/176**, 736 (1997).
[10] G. Abstreiter, P. Schittenhelm, C. Engel *et al.*, Semicond. Sci. Technol. **11**, 1521 (1996).
[11] D. J. Eaglesham and M. Cerullo, Phys. Rev. Lett. **64**, 1943 (1990).
[12] Feing Liu and M. G. Lagally, Surf. Sci. **386**, 169 (1997).
[13] V. Ya. Aleshkin, N. A. Bekin, N. G. Kalugin *et al.*, JETP Lett. **67**, 48 (1998).
[14] A. B. Talochkin, V. A. Markov, S. P. Suprun, and A. I. Nikiforov, JETP Lett. **64**, 219 (1996).
[15] A. I. Yakimov, A. V. Dvurechenskii, V. A. Markov *et al.*, Physica E, in press.
[16] I. Goldfarb, P. T. Hayden, J. H. G. Owen, and G. A. D. Briggs, Phys. Rev. B **56**, 10459 (1997).
[17] L. Colombo, R. Resta, and S. Baroni, Phys. Rev. B **44**, 5572 (1991).

Translated by M. E. Alferieff

Electrons and photons in mesoscopic structures: quantum dots in a photonic crystal

S. V. Gaponenko and A. M. Kapitonov
Institute of Molecular and Atomic Physics, National Academy of Sciences, 220072 Minsk, Belarus

V. N. Bogomolov and A. V. Prokofiev
A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

A. Eychmüller
Institute of Physical Chemistry, University of Hamburg, 20146 Hamburg, Germany

A. L. Rogach
Institute of Physical Chemistry, University of Hamburg, 20146 Hamburg, Germany; Physicochemical Research Institute, Belarus State University, 220080 Minsk, Belarus

(Submitted 15 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 131–135 (25 July 1998)

The synthesis and properties of a photonic crystal doped with quantum dots are reported. The structure exhibits a two-stage self-organization of silica nanoparticles along with quantum confinement effects in semiconductor colloids. The interplay of electron and photon confinement results in controllable spontaneous emission from the mesoscopic structure. © *1998 American Institute of Physics.*
[S0021-3640(98)00714-2]

PACS numbers: 81.05.Ys, 42.70.Qs, 81.20.Fw, 73.23.−b

Mesoscopic structures with a characteristic size either of the order of the electron de Broglie wavelength in a semiconductor (1–10 nm) or close to the optical photon wavelength (100–1000 nm) exhibit nontrivial properties due to the modified electron or photon density of states. Three-dimensional spatial confinement of electrons in nanocrystals (''quantum dots'') results in size-dependent energies and probabilities of optical transitions.[1–3] The photon density of states can be modified in structures with strong modulation of the refractive index in three dimensions (photonic crystals).[4–6] It can be performed by means of supermolecular crystallization of matter, resulting in colloidal crystals with a submicron period.[7–10] Because of the substantially different electron and photon wavelengths, the electron and photon densities of states can be engineered separately within the same mesostructure. In this letter we report the synthesis and several properties of a photonic crystal doped with semiconductor quantum dots. The structure exhibits a two-stage self-organization of silica nanoparticles along with quantum confinement effects in semiconductor colloids.

The photonic crystals used in our studies were synthesized from a sol of monodis-

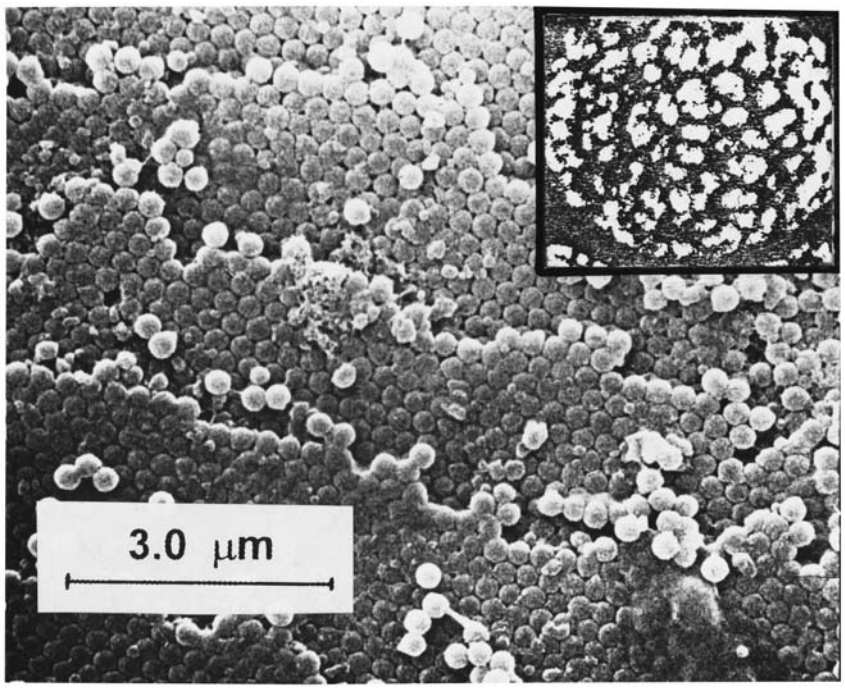

FIG. 1. Electron micrograph of the surface of a solid state silica-based colloidal crystal. Clearly seen are several successive layers consisting of close-packed spherical submicron-sized silica globules. Each globule in its turn consists of nanometer-sized silica particles (inset). The whole superstructure forms a three-dimensional face-centered cubic lattice. Because of coherence effects in the multiple scattering of optical waves, these structures inhibit the propagation of electromagnetic waves within a certain spectral range.

perse SiO_2 spherical globules by means of sedimentation and hydrothermal treatment. Close-packed silica globules are cemented together and form a solid face-centered cubic lattice, each globule consisting of nanometer-size silica particles (Fig. 1). Therefore, the resulting crystals show two stages of self-organization: aggregation of nanoparticles in a globule, and close-packing of globules in a three-dimensional lattice. The superstructures thus developed, which are called artificial opals, exhibit a photonic pseudogap in the visible range[8,11,12] which manifests itself as a pronounced stop band in transmission/reflection spectra in a spectral range determined by the globule size, the refractive index of the globules and interglobule cavities, and the crystallographic orientation of the photonic crystal (Fig. 2). The physical origin of a pseudogap in a three-dimensional colloidal crystal is multiple rescattering and interference of optical waves propagating throughout.

Since the interglobule cavities form a network, it is possible to embed therein desirable species (molecules, ions, or nanocrystals) exhibiting electron resonance in absorption or emission of light. To get overlapping size-dependent electron resonance in nanocrystals and structural optical resonance in a photonic crystal, the latter was doped with cadmium telluride nanoparticles. Semiconductor crystallites have been prepared through the addition of fresh oxygen-free NaHTe solution to N_2-saturated $Cd(ClO_4)_2 \cdot 6H_2O$ solution in the presence of thiols as stabilizing agents.[13] Well-defined CdTe nanocrystals

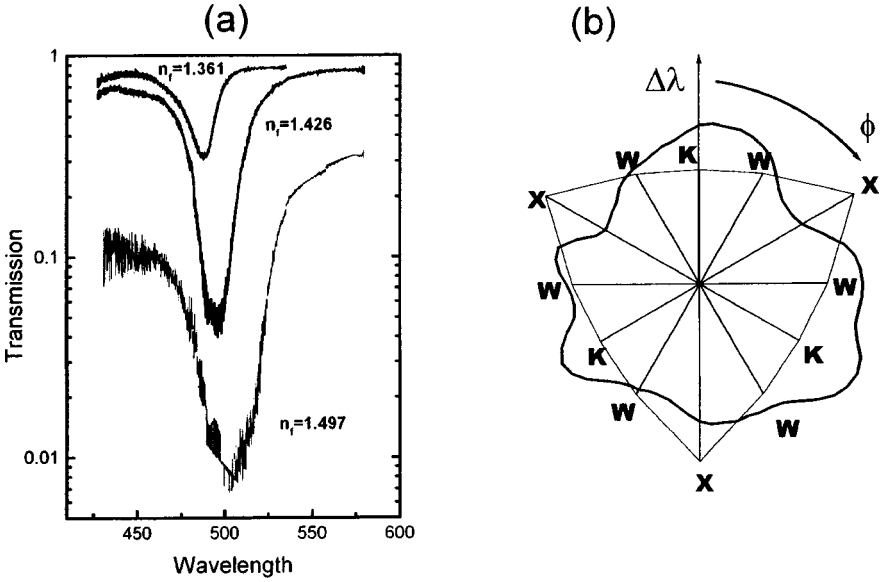

FIG. 2. Photonic stop band in an opal-based photonic crystal. Presented in (a) are transmission spectra of an opal platelet sample impregnated with fillers possessing successively growing refraction index n_f. The polar diagram (b) shows the angular dependence of the stop band. $\Delta\lambda$ is the spectral shift of the stop band as compared to the $\langle 111 \rangle$ direction, and φ is the polar angle inside the (111) plane. The deviation from $\langle 111 \rangle$ was 20°. The underlying triangular contour emphasizes the special directions of the Brillouin zone of a photonic crystal; the lengths of the radial lines are proportional to the angles between $\langle 111 \rangle$ and the directions to the special points.

were developed with predominant cubic crystalline structure possessing significant intrinsic emission in the visible spectral range (Fig. 3). Due to the size restriction, the absorption spectrum shifts by more than 1 eV (from 827 nm in the bulk CdTe crystal to 460–500 nm in the clusters). In the context of quantum size effects, this is the typical strong confinement range, when the kinetic energies of the electrons and holes are substantially larger than the energy of the electron–hole Coulomb interaction. Because of the pronounced size dependence, the emission spectrum can be tuned towards a photonic pseudogap of a photonic crystal. This was performed using 1-thioglycerol as a stabilizer. The mean size of the CdTe crystallites in this case was about 2.4 nm. The intrinsic emission band peaking at 575 nm dominates in the luminescence spectrum. The spectrum shows a pronounced inhomogeneous broadening and strongly nonexponential decay with a mean lifetime of about 10^{-8} s.

Semiconductor quantum dots embedded in a photonic crystal exhibit a noticeable change of the luminescence spectrum when the latter overlaps with the photonic pseudogap (Fig. 4). The modification of the spontaneous emission is due to a modified density of photon states in the pseudogap region. It is known that the spontaneous decay of an excited state of any quantum system is not an intrinsic property of the system but a result of its interaction with the electromagnetic vacuum. The density of photon states is redistributed in a photonic crystal as compared to free space. It vanishes within the gap and increases near the gap edges. Therefore, the spontaneous decay rate, which is directly

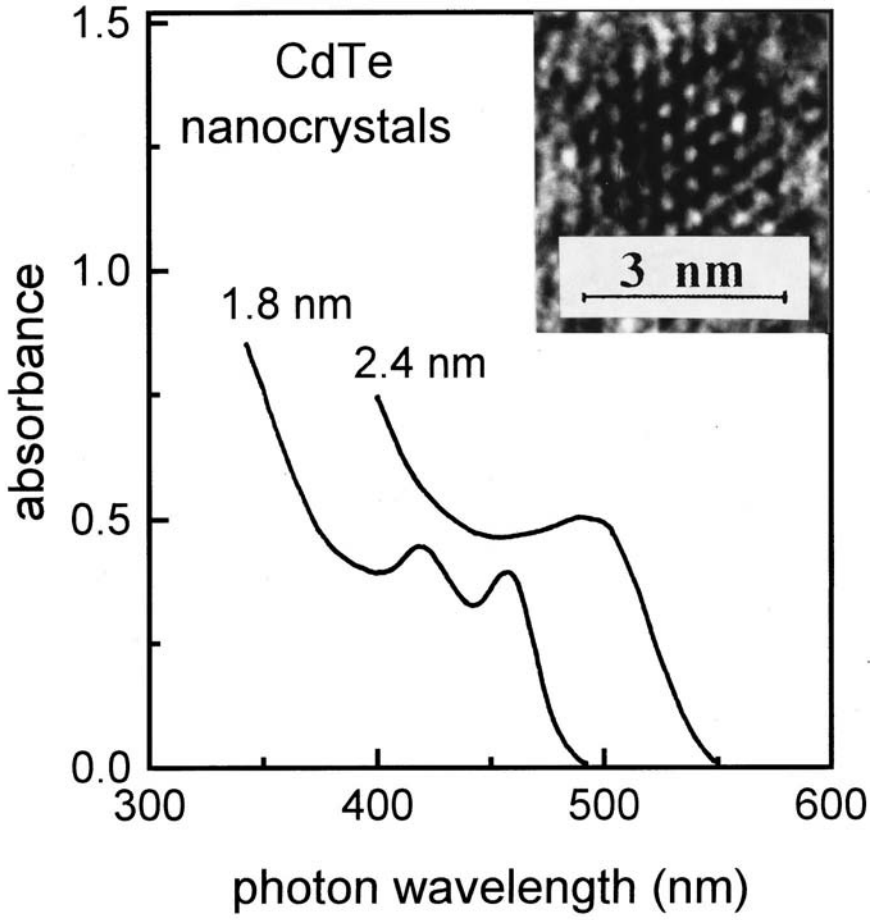

FIG. 3. Absorption spectra of two representative colloidal solutions containing cadmium telluride crystallites of mean size of 1.8 and 2.4 nm. Unlike the bulk CdTe crystal, whose fundamental absorption edge rises at 827 nm, nanocrystals exhibit multiband absorption spectra with the first absorption feature tunable up to the near ultraviolet range. Within the framework of quantum size effects, an evolution from a crystal to cluster can be interpreted in terms of the electron–hole confinement in a three-dimensional quantum box. Considered in this context, the crystallites correspond to the so-called strong confinement limit. The inset shows a single nanocrystal with pronounced crystallographic planes.

proportional to the photon density of states, is expected to be inhibited within the gap and enhanced in the close vicinity outside the gap. Our experiments clearly show a dip in the emission spectrum (Fig. 4) following the spectral position of the pseudogap (e.g., Fig. 2). The appearance of the dip is indicative of the inhibited spontaneous emission. Although a modification of the emission spectrum also takes place at the edges of the pseudogap, additional studies are necessary to reach an unambiguous conclusion about enhanced emission at the gap edges. To our knowlege,[14] however, when dye molecules are located inside a photonic crystal with a pseudogap, the rate of spontaneous emission splits into slowed and accelerated components. Unfortunatelly, the insufficiently high quantum

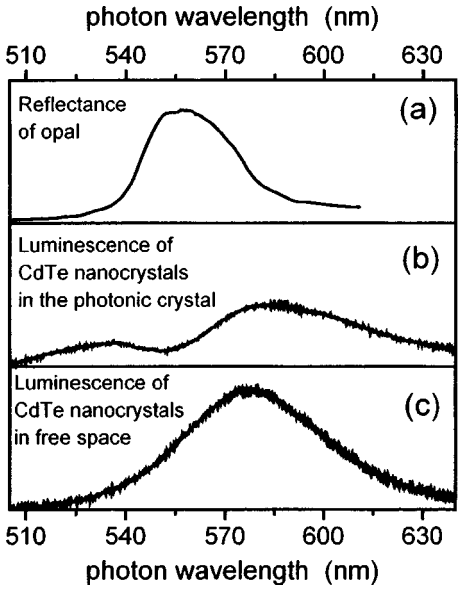

FIG. 4. Effect of the photonic band gap on the spontaneous emission of CdTe nanocrystals. Panel (a) shows the reflection spectrum of a silica colloidal crystal doped with CdTe crystallites with a size of 2.4 nm. The pronounced peak in reflection is indicative of a dip in the spectral distribution of the photon density of states due to the periodic three-dimensional lattice of silica globules. In the spectral range relevant to the photonic pseudogap the luminescence spectrum has a minimum (panel b) which is indicative of an inhibited spontaneous emission. Panel (c) shows the reference emission spectrum of the same nanocrystals outside the photonic crystal.

yield of fluorescence is an impediment to analogous experiments with semiconductor nanocrystals.

The observed modification of the spontaneous emission spectrum of quantum dots in silica-based photonic crystals is much more significant than the modification reported for organic molecules in solid-state[8] and liquid-like[15] colloidal crystals. This is due to an intrinsically narrower emission spectrum of quantum dots as compared to organic molecules, which provides a stronger overlap of the emission spectrum and the gap. The homogeneous linewidth of a given nanocrystal emission spectrum is substantially narrower than the integral emission spectrum of a nanocrystal ensemble. Inhomogeneous broadening is inherent in every ensemble of semiconductor clusters. It arises because of a distribution of transition energies and decay rates which in turn is related to the distributions in cluster size, surface structure, local environments, and impurities. Therefore, each individual component within an inhomogeneously broadened spectrum reacts individually to the modified density of photon states. Thus the total modification of the spontaneous emission is enhanced in the case of an inhomogeneously broadened spectrum as compared to a homogeneous band with the same total bandwidth. Under conditions of enhanced modulation of the refraction index, silica-based photonic crystals doped with quantum dots will afford a possibility of observing a noticeable slowing down of the spontaneous decay of a quantum system, down to the emergence of the ''frozen'' excited states predicted by the theory.[16]

To summarize, we demonstrate a novel mesostructure with separately controllable densities of photon and electron states and show that the spontaneous emission of nanocrystals, which is controlled at large by the quantum confinement effect, experiences strong modification in a photonic crystal as compared to free space on account of the modified photon density of states.

Stimulating discussions with S. Ya. Kilin are gratefully acknowledged. This research has been supported in part by the INTAS Program of the European Community and the Volkswagen Foundation.

[1] L. Banyai and S. W. Koch, *Semiconductor Quantum Dots*, World Scientific, Singapore, 1993.
[2] S. V. Gaponenko, *Optical Properties of Semiconductor Nanocrystals*, Cambridge Univ. Press, New York, 1998.
[3] M. Nirmal, B. O. Daboussi, M. G. Bawendi *et al.*, Nature (London) **383**, 802 (1996).
[4] E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987); S. John, Phys. Rev. Lett. **58**, 2486 (1987).
[5] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*, Princeton Univ. Press, Singapore, 1995.
[6] J. B. Pendry, J. Phys.: Condens. Matter **8**, 1058 (1996).
[7] W. L. Vos, R. Spirk, A. van Blaaderen *et al.*, Phys. Rev. B **53**, 16231 (1996).
[8] V. N. Bogomolov, S. V. Gaponenko, I. N. Germanenko *et al.*, Phys. Rev. E **55**, 7619 (1997).
[9] H. Mígues, C. López, F. Meseguer *et al.*, Appl. Phys. Lett. **71**, 1148 (1997).
[10] R. Biswas, M. M. Sigalas, G. Subramania, and K.-M. Ho, Phys. Rev. B **57**, 3701 (1998).
[11] V. N. Bogomolov, D. A. Kurdyukov, A. V. Prokofiev, and S. M. Samoilovich, JETP Lett. **63**, 520 (1996).
[12] Yu. A. Vlasov, V. N. Astratov, O. Z. Karimov *et al.*, Phys. Rev. B **55**, R13357 (1997).
[13] A. L. Rogach, L. Katsikas, A. Kornowski *et al.*, Ber. Bunsenges. Phys. Chem. **100**, 1772 (1996);*ibid* **101**, 1668 (1997).
[14] E. P. Petrov, V. N. Bogomolov, I. I. Kalosha, and S. V. Gaponenko, Phys. Rev. Lett. (1998), accepted for publication.
[15] J. Martorell and N. M. Lawandy, Opt. Commun. **78**, 169 (1990).
[16] S. Ya. Kilin and D. S. Mogilevtsev, Laser Phys. **2**, 153 (1992).

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

Two-dimensional excitons and hydrogenlike impurities on a curved surface

L. I. Magarill and A. V. Chaplik

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 18 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 136–140 (25 July 1998)

The shifts and splittings are found for a two-dimensional donor and a Wannier–Mott exciton on a curved surface. The appearance of fine structure in the excitonic luminescence line as a result of the motion of the system as a whole in a potential of geometric origin is estimated. The relative contributions of the Coulomb potential and kinetic energy operator to the corrections for curvature depend strongly on the geometric form of the surface. © *1998 American Institute of Physics.* [S0021-3640(98)00814-7]

PACS numbers: 71.35.Cc, 68.55.Ln

1. The properties of two-dimensional (2D) nonplanar electronic systems have been attracting the attention of investigators for a number of years now.[1–6] Until very recently, such objects were either macroscopic samples with a radius of curvature R of the order of 0.1 cm (2D electron gas on a curved substrate[7]) or carbon nanotubes with a radius six orders of magnitude smaller. This large range of curvature can be covered practically continuously by an ingenious method proposed in Ref. 8, win which initially strained GaAs/InAs layers are ''rolled up.'' Specifically, objects whose radius of curvature is comparable to (or exceeds by only an order of magnitude) the effective electronic Bohr radius a_B^* in the bulk of a semiconductor become quite practicable. In Ref. 8 it was also shown that an external load can deform a tube into an elliptic cylinder.

In this letter we calculate the energy spectrum of excitons and hydrogenlike donors placed in a curved quantum well whose thickness is less than a_B^*, i.e., the 2D limit is realized. We show that the curvature of the surface on which the particles move results in a shift and splitting of the levels of the corresponding planar system, and the relative role of the Coulomb interaction and the curvature depends strongly on the geometric shape of the surface at a fixed ratio a_B^*/R.

2. We begin by studying an exciton on the surface of an elliptic cylinder. We are actually talking about a film, of small but finite thickness, rolled up into a cylinder. As was shown in Ref. 9, the transition from a 3D to a 2D wave equation under these conditions leads to the appearance of an adiabatic potential of purely geometric origin (the motion transverse to the film is assumed to be fast, i.e., size quantization gives the greatest splitting between the energy levels). For a cylindrical film of constant width this potential has the form[9]

$$U_g(v)=\frac{\hbar^2}{2m^*a^2}\,\frac{\zeta\varepsilon^4\sin^2(2v)}{\xi(v)^3}, \tag{1}$$

where m^* is the effective mass, a and ε are the semimajor axis and eccentricity of the ellipse, v is the elliptic coordinate $(0<v<2\pi)$, $\xi(v)=1-\varepsilon^2\cos v^2$, $\zeta=(1+4\pi^2/3)/16$. We shall assume the effective Bohr radius to be small compared with both semiaxes of the ellipse. Then, it is convenient to change to the variables z (along the axis of the cylinder) and s (elliptic arc length). In these variables the electron–hole Coulomb interaction, to within the first correction for the curvature, is

$$U_{eh}=U_{eh}^{(0)}+U_{eh}^{(1)}, \tag{2}$$

$$U_{eh}^{(0)}=-\frac{\tilde{e}^{\,2}}{\rho},\quad U_{eh}^{(1)}=-\frac{\tilde{e}^{\,2}(1-\varepsilon^2)s^4}{24\rho^3(\xi(V))^3a^2}\equiv-\frac{\tilde{e}^{\,2}s^4}{24\rho^3R(V)^2}, \tag{3}$$

where

$$\rho=\sqrt{z^2+s^2},\quad z=z_e-z_h,\ s=s_e-s_h,\quad s_{e,h}=\int_0^{v_{e,h}}dt\sqrt{\xi(t)},$$

$R(V)$ is the local radius of curvature, $V=(m_ev_e+m_hv_h)/M$, $M=m_e+m_h$, and $m_{e,h}$ are, respectively, the electron and hole effective masses, and $\tilde{e}^{\,2}$ is the square of the electron charge with allowance for the background permittivity. After transforming the electron and hole kinetic energy operators to center-of-mass variables S and Z and relative-motion variables s and z, and transforming the unknown function $\psi=\chi/(\xi(v_e)\xi(v_h))^{1/4}$, we obtain for the exciton Hamiltonian

$$H=H^{(0)}+H^{(1)},$$

where $H^{(0)}$ is the Hamiltonian of the two-dimensional system (μ is the reduced mass):

$$H^{(0)}=-\frac{\hbar^2}{2M}\left[\frac{\partial^2}{\partial S^2}+\frac{\partial^2}{\partial Z^2}\right]+\frac{\hbar^2}{2\mu}\left[\frac{\partial^2}{\partial s^2}+\frac{\partial^2}{\partial z^2}\right]-\frac{\tilde{e}^{\,2}}{(s^2+z^2)^{1/2}}, \tag{4}$$

and $H^{(1)}$ takes account of the curvature:

$$H^{(1)}=\frac{\hbar^2}{2\mu a^2}\left[\left(\zeta-\frac{5}{16}\right)\frac{\varepsilon^4\sin^2(2V)}{\xi(V)^3}+\frac{\varepsilon^2\cos(2V)}{2\xi(V)^2}\right]-\frac{\tilde{e}^{\,2}s^4}{24(s^2+z^2)^{3/2}R(V)^2}. \tag{5}$$

In the perturbation $H^{(1)}$ we neglected terms of order $(a_B^*/a)^4$. In Eq. (5) the variable V must be expressed in terms of S; to leading order in a_B^*/a this relation has the form $S=\int_0^V dt\sqrt{\xi(t)}$. As one can see from the last term in Eq. (5), it is impossible to separate the relative and center-of-mass motions of an exciton on a curved surface.

In the simplest case of a circular cylinder ($\varepsilon=0$, $R(V)\equiv a$) the geometric potential vanishes. In this case the level shifts and splittings arise only as a result of a change in the Coulomb interaction. For the ground state we obtain

$$\Delta E_1=-\frac{\hbar^2}{128\mu a^2}. \tag{6}$$

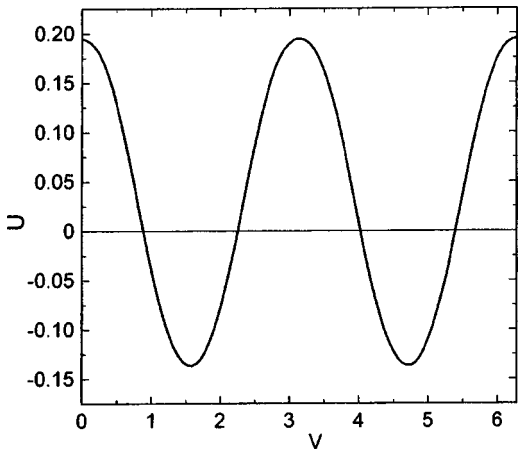

FIG. 1. Adiabatic potential of the center of mass of an exciton; $\varepsilon=1/2$, $U(V)$ is given the units $h^2/2\mu a^2$.

For the first excited (triplet) level

$$\Delta E_2^{(1)}=-\frac{7\hbar^2}{128\mu a^2},\quad \Delta E_2^{(2)}=-\frac{\hbar^2}{64\mu a^2},\quad \Delta E_2^{(3)}=-\frac{5\hbar^2}{64\mu a^2}. \tag{7}$$

We note that the first correction for the curvature does not contain the electron charge and therefore does not depend on the effective permittivity.

In the case of an elliptic cylinder an exciton as a whole moves in a potential that depends on V. If $a_B^*\ll a$, this motion is slow compared with the internal motion in the exciton. Averaging U_{eh} over the fast motion, it is possible to obtain an adiabatic potential $U(V)$ for the motion of the center of mass. A plot of this potential for the ground state of an exciton with $\varepsilon=1/2$ is presented in Fig. 1. As one can see from Eqs. (6) and (7), on account of the small numerical coefficients in the expression for ΔE the Coulomb contribution in this case is negligibly small. It becomes substantial for $\varepsilon\leqslant 0.3$. The deepest minima for any ε correspond to the tips of the semiminor axis of the ellipse, i.e., the regions of smallest curvature. Therefore exciton localization in these regions and spatially nonuniform excitonic luminescence should be expected. The number of energy levels corresponding to the motion of the center of mass of an exciton in the minimum mentioned is easily estimated to be of the order of $\sqrt{m_h/m_e}$. For this reason, for a heavy-hole exciton the lower levels can be described in the oscillator approximation. The corresponding frequency for the ground state of the exciton is given by the expression

$$\omega=\frac{\hbar}{a^2\sqrt{2m_em_h}}\left[\frac{2\pi^2}{3}\varepsilon^4-\frac{125}{32}\varepsilon^4+\frac{61}{32}\varepsilon^2\right]^{1/2}, \tag{8}$$

and for $\varepsilon=1/2$ one has $\omega\approx 0.567\hbar/a^2\sqrt{m_e/m_h}$. Therefore the excitonic luminescence line is accompanied by a fine structure of scale $\omega\sim\omega_B\sqrt{m_e/m_h}(a_B^*/a)^2$, where ω_B is the Bohr frequency of the exciton.

3. Let us now consider donor states on a curved surface. Once again, we assume the Bohr radius to be small compared with the characteristic radii of curvature. In a small

neighborhood of an arbitrary point the smooth surface can be approximated by a paraboloid $z=-(x^2/2A+y^2/2B)$. We assume the donor to be located at the vertex of the paraboloid. To obtain the effective 2D Hamiltonian, once again we consider a film with a small constant width, curved into a paraboloid. Using paraboloidal coordinates, it can be shown that the adiabatic potential U_{ad} arising with the transition to two dimensions is of the order of $(\hbar^2/m^*)(s^2/A^4, s^2/B^4)$, where s is the distance along the surface from the vertex of the paraboloid. Since the important values of s are of the order of a_B^*, the contribution of U_{ad} is small compared with the corrections arising from the kinetic energy operator and the Coulomb potential. It is convenient to calculate these corrections by changing to the variables s and ϕ, where ϕ is the azimuthal angle and s is measured along a curve which is a section of the paraboloid by the plane $\phi=$ const. After quite unwieldy calculations,[a)] we obtain

$$\Delta E_1=\frac{\hbar^2}{128m^*}\left[31\left(\frac{1}{A^2}+\frac{1}{B^2}\right)-\frac{22}{AB}\right], \tag{9}$$

for the donor ground state and

$$\Delta E_2^{(1)}=\frac{\hbar^2}{128m^*}\left[25\left(\frac{1}{A^2}+\frac{1}{B^2}\right)-\frac{26}{AB}\right];$$

$$\Delta E_2^{(2,3)}=\frac{\hbar^2}{128m^*}\left[26\left(\frac{1}{A^2}+\frac{1}{B^2}\right)-\frac{4}{AB}\pm\left(\frac{1}{A^2}-\frac{1}{B^2}\right)\right], \tag{10}$$

for the first excited level. We note that the negative shifts from the Coulomb interaction are numerically small compared with the positive contributions from the geometric potential. One would expect that the limiting values of expressions (9) and (10) with $A=B$ or in the limit $B\to\infty$ will give corrections to donor levels on a sphere or cylinder, respectively. However, this is not so. Calculations show that for a sphere of radius A

$$\Delta E_1=-\frac{9\hbar^2}{48m^*A^2}; \tag{11}$$

$$\Delta E_2^{(1)}=\frac{11\hbar^2}{3m^*A^2}; \quad \Delta E_2^{(2,3)}=-\frac{\hbar^2}{8m^*A^2}, \tag{12}$$

while for an elliptic cylinder with semimajor axis A

$$\Delta E_1=U_0(v_0)-\frac{\hbar^2}{128m^*A^2}; \tag{13}$$

$$\Delta E_2^{(1)}=U_0(v_0)-\frac{7\hbar^2}{128m^*A^2}; \quad \Delta E_2^{(2,3)}=U_0(v_0)+\frac{\hbar^2}{64m^*A^2}(-3\pm 2), \tag{14}$$

where

$$U_0(v_0)=\frac{\hbar^2}{2m^*R(v_0)^2}\,\frac{\varepsilon^2}{1-\varepsilon^2}\left[\left(\zeta-\frac{1}{16}\right)\sin^2(2v_0)+\frac{\cos(2v_0)-\varepsilon^2\cos^2 v_0}{2}\right]$$

(v_0 is the elliptic coordinate of the donor). Therefore the shape of the surface as a whole already has an effect starting with terms of order $(a_B^*/A)^2$.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-19058), the program ''Physics of Solid-State Nanostructures,'' and also by INTAS Grant (No. 95-0657).

[a)]To use the formulas of perturbation theory, where the wave functions of a planar two-dimensional Coulomb problem are used as the zeroth approximation, the perturbation operator must be put into a hermitian form. This is accomplished by a transformation of the desired function, which we do not write out here because of its complexity.

[1] V. M. Nabutovskiĭ and D. A. Romanov, Zh. Éksp. Teor. Fiz. **90**, 232 (1986) [Sov. Phys. JETP **63**, 133 (1986)].
[2] H. Aoki and H. Suezawa, Phys. Rev. A **46**, R1163 (1992).
[3] C. L. Foden, M. L. Leadbeater, J. H. Burroughes, and M. Pepper, J. Phys.: Condens. Matter **6**, L127 (1994).
[4] V. V. Rotkin and R. A. Suris, Fiz. Tverd. Tela **36**, 3569 (1994) (St. Petersburg) [Phys. Solid State **36**, 1899 (1994)].
[5] M. L. Leadbeater, C. L. Foden, J. H. Burroughes *et al.*, Phys. Rev. B **52**, 8629 (1995).
[6] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, JETP Lett. **64**, 460 (1996); Zh. Éksp. Teor. Fiz. **113**, 1411 (1998) [JETP **86**, 771 (1998)].
[7] S. Böhm, A. Lorke, J. P. Kotthaus, to be published; S. Böhm, *Diplomarbeit*, Ludwig-Maximilian-Universität, München, (1997).
[8] V. Ya. Prinz, V. A. Seleznev, V. A. Samoylov, and A. K. Gutakovsky, Microelectron. Eng. **30**, 439 (1996); V. Ya. Prinz, V. A. Seleznev, A. K. Gutakovsky, *Abstracts of the 24th International Conference on Semiconductor Physics*, Jerusalem, Israel, 1998.
[9] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, Zh. Éksp. Teor. Fiz. **110**, 669 (1996) [JETP **83**, 361 (1996)].

Translated by M. E. Alferieff

Behavior of sound velocities in the compounds $La_{1-x}Sr_xMnO_3$ near magnetic and structural phase transitions

Yu. P. Gaĭdukov[a)] and N. P. Danilova
Department of Physics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

A. A. Mukhin
Institute of General Physics, Russian Academy of Sciences, 117924 Moscow, Russia

A. M. Balbashov
Moscow Power Engineering Institute, 105835 Moscow, Russia

(Submitted 19 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 141–146 (25 July 1998)

The temperature dependences of the transverse V_t and longitudinal V_l sound velocities in single crystals of the perovskites $La_{1-x}Sr_xMnO_3$ ($x=0.1, 0.175, 0.2, 0.25$) in the temperature interval $T=70-350$ K are investigated by the resonance method. Anomalies — small minima and kinks in the temperature dependences $V_l(T)$ and $V_t(T)$ — are observed at the Curie points. A strong jumplike increase (by up to 30%) in both the longitudinal and transverse sound velocities, which attests to sizable hardening of the acoustic branches of the phonon spectrum, is observed near the temperatures of the structural transitions between the rhombohedral and orthorhombic phases (for $x=0.175, 0.2, 0.25$) and at the phase transition to the polaron-ordered state (for $x=0.1$). © *1998 American Institute of Physics.* [S0021-3640(98)00914-1]

PACS numbers: 62.65.+k, 75.30.Kz, 64.70.Kb, 43.20.Hq

Despite the enormous amount of available experimental data, there is no consensus concerning the physical nature of the unusual properties of manganites,[1] and this subject remains controversial. Specifically, the question of a strong coupling of the electronic and magnetic subsystems with the lattice is under discussion. However, there are extremely few publications devoted to the investigation of the behavior of the elastic constants, which give valuable information about this. We know of only two works — Refs. 2 and 3 — where the behavior of the sound velocity in ceramic samples of manganites was investigated and the observation of anomalies of the order of several percent at phase transition points were reported.

In the present letter we report the preliminary results of an investigation of the behavior of the sound velocities in single-crystal samples of the compounds $La_{1-x}Sr_xMnO_3$ as a function of the content $x=0.1, 0.175, 0.2$, and 0.25 and the temperature in the interval 70–350 K.

The samples were cut from single-crystal billets with a definite composition, prepared by zone melting. The procedure for preparing the samples and analyzing the composition of the billets is described in Ref. 4. A small difference in the composition of the components was observed along a billet. The experimental samples were not only disk-shaped, but half-disks, squares, or shapes close to these were also used. The samples were of the order of 1 mm thick and their dimensions in the plane ranged from 8 to 4 mm. Structurally, the samples can be called single crystals only as a matter of convention, since they contained large blocks of twin formations. However, their density corresponded to the x-ray value and for all compositions could be characterized to within 5% by the value $6.5\,\mathrm{g/cm^3}$.

The sound velocity was determined by the resonance method. For this purpose, the conventional measurement scheme for contact-free electromagnetic excitation of sound in metals with a differential passage of the signal was used.[5] In contrast to metals, where the mechanical modes of the characteristic oscillations are excited as a result of forces arising in the skin layer, in manganites, on account of their low electrical conductivity, excitation and detection are possible only through magnetostriction in the bulk. The samples were placed freely in two parallel, identical, narrow, rectangular coils (each consisting of ten loops of copper wire 0.1 mm in diameter) and connected, respectively, to a generator and to the detector of a SKCh-59 spectrum analyzer. In addition, the scan voltage of the high-frequency scanning field (the frequency range $f=100\,\mathrm{kHz}-10\,\mathrm{MHz}$) was modulated by a voltage with a frequency of 37 Hz. The coils and the sample together were placed into an electromagnet in such a way that its magnetic field H (up to 10 kOe) lay in the plane of the samples and could make an angle α with the axis of the coil from 0° to 90°. The signals from the detecting coil — either after the first detection $U(f)$ or after the second synchronous detection $A=dU/df(f)$ — were recorded on a X–Y plotter.

At frequencies above 2 MHz it was only in rare cases possible to detect in the trace $dU/df(f)$ the multimode resonance structure due to the lowest-numbered modes of the thickness resonance associated with the longitudinal sound velocity. Furthermore, the thickness resonances associated with the transverse sound velocity could not be identified because of the complicated picture of the spectral composition of $dU/df(f)$. At low frequencies, up to 1 MHz (Fig. 1), on the contrary, we observed clear, strong, single-mode resonances. We performed theoretical and experimental investigations to clarify the nature of these resonances, using nickel as a model for the manganite samples. It was found that for the above-described geometry and excitation conditions the only modes that should arise are those which are symmetric about to the central plane of the samples and which depend weakly (or not at all) on thickness over quite wide limits.[6] These modes stem from a transverse shear wave and the zeroth symmetric Lamb wave for thin plates.

In Ref. 6 the relations between the disk radius r, transverse sound velocity V_t, Poisson ratio σ, and the resonance frequencies f_n of the first ten modes were calculated and confirmed experimentally for the case of disks consisting of isotropic material. Two of them were known earlier (radial and torsional) and are widely used for exciting sound in piezoelectrics.[7]

For other shapes of flat samples, for lack of an analytical relation between the quantities indicated above, it was necessary to resort to modeling the situation using

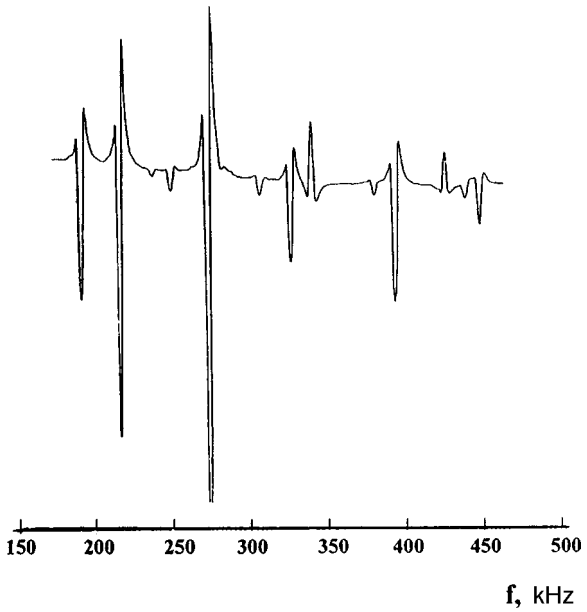

FIG. 1. Trace of $dU/df(f)$ for a sample with $x=0.2$, $\alpha=0$, and $P=800$ Oe at $T=78$ K; disk-shaped sample 8 mm in diameter and 1.5 mm thick; Curie temperature $T_C=305$ K.

nickel samples, transforming their shape from disk to square, half-disk, or other shapes. This made it possible to follow the change in the resonance frequencies of some mode or another. The most convenient for analysis is the lowest, first mode f_1 of a disk, since it is virtually independent of the Poisson ratio. For a thin disk $f_1=2.35V_t/2\pi r \times(\pm 0.15\%)$. The results presented below are associated mainly with the behavior of the frequency f_1.

We shall make one final methodological remark: The traces $U(f)$ are straight lines over a wide frequency range (from 100 kHz to ≈ 1 MHz). Weak resonance features (in the best case, of the order of several percent) can be observed against the background of these traces. The slope of these straight lines is proportional to the differential magnetic susceptibility of the sample. The observation of this quantity in a weak magnetic field ($H\sim 20$ Oe) made it possible to locate the points of the magnetic T_C and structural T_s transitions (Fig. 2). The values of the temperatures T_C and T_s agree well with data from other works.[8,9]

The resonance features of $dU/df(f)$ below the Curie point first appeared at very low fields ($H\sim 20$ Oe). Up to fields $H_{\max}\approx 500$ Oe their amplitude grows strongly, passes through a maximum, and then in the case $\alpha=0$ approaches zero in fields $H \approx 1-2$ kOe just as rapidly. In the case $\alpha=90°$, after reaching its maximum value the amplitude dropped off very little with increasing field and resonances could be observed all the way up to 10 kOe. The change (increase) in the resonance frequencies with increasing magnetic field did not exceed 1–2%. For fixed fields $H<H_{\max}$ and increasing temperature, for one of the modes the amplitude of the resonances decreases, and resonance is no longer observed as the Curie point T_C is approached. For other modes, as

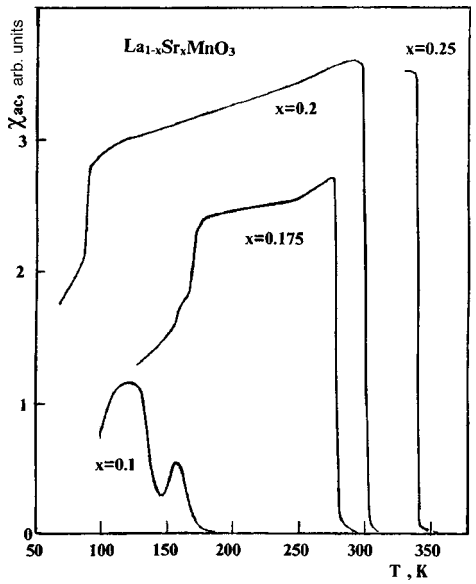

FIG. 2. Temperature dependence of the differential susceptibility, normalized to the volume, for different values of x. Measured at $f=300$ kHz.

well as thickness resonances, the amplitude increases, reaches a pronounced maximum at the Curie point, and then drops off rapidly. Sometimes, these resonances could be traced up to a temperature 10° above the Curie point (Fig. 3).

For all compositions (with a variation of 10%), at a temperature of the order of 300

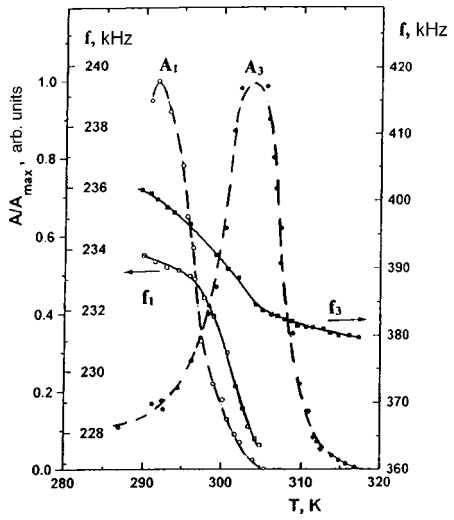

FIG. 3. Temperature dependence of the frequency and amplitude of the resonance of two vibrational modes for a sample with $x=0.2$; disk-shaped sample 8 mm in diameter and 1.5 mm thick; $T_C=305$ K.

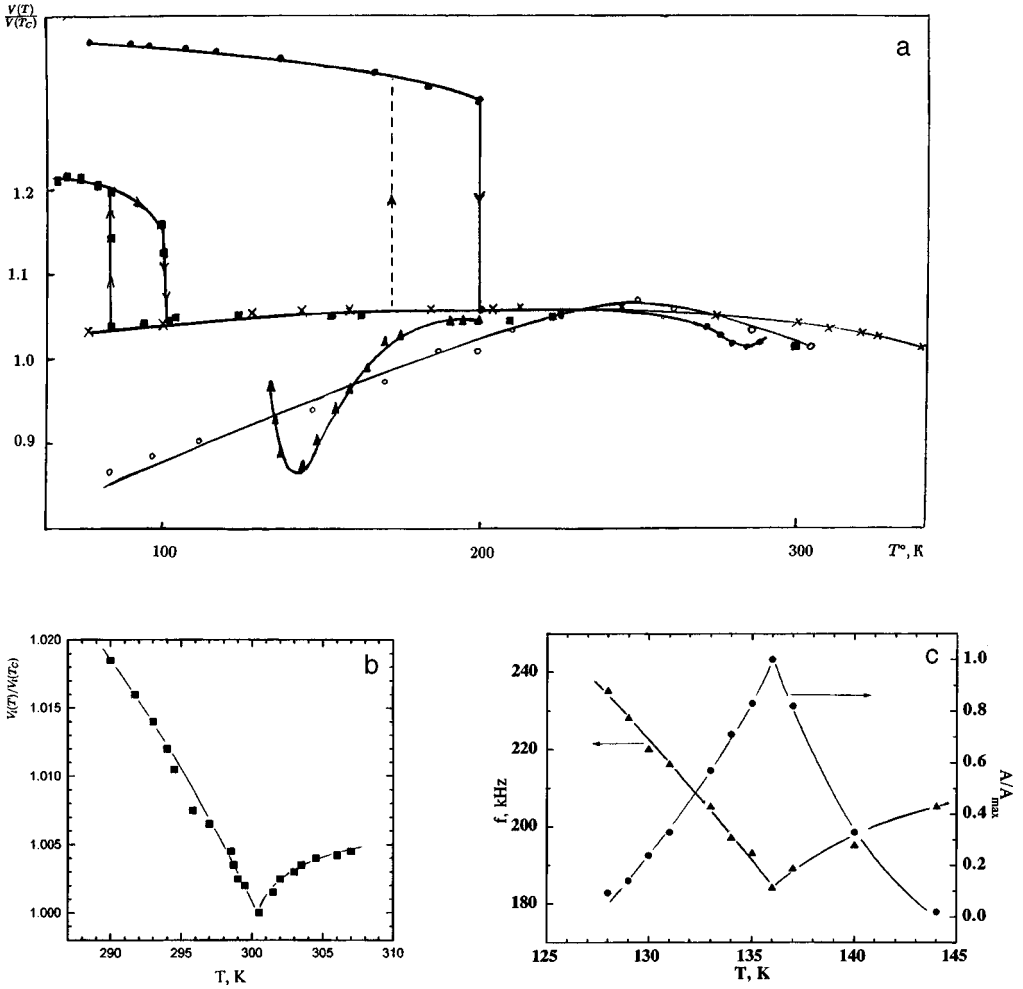

FIG. 4. Temperature dependence of the sound velocities for different x: (a) × — V_t; $x=0.25$; 5.8×5.8 mm square; thickness 0.9 mm; $T_C=340$ K; ○ — V_t; $x=0.2$; disk 8 mm in diameter and 1.5 mm thick, $T_C=305$ K; ■ — V_t; $x=0.2$; 8×7.6 mm oval; thickness 1.2 mm; $T_C=300$ K; ● — V_l; $x=0.175$; half-disk, $r=4$ mm; thickness 0.9 mm; $T_C=288$ K; ▲ — V_t; $x=0.1$; 6.5×6.1 mm oval; thickness 1.1 mm; $T_C=170$ K. (b) — Temperature dependence of the longitudinal velocity V_l for $x=0.2$; $T_C=300$ K. (c) — Temperature dependence of the frequency and amplitude of the resonance f_2 for $x=0.1$.

K the transverse velocity V_t, determined from the frequencies of the first resonance, equals ≈ 2.5 km/s. The value of the longitudinal velocity, determined, where possible, from the thickness resonances (for $x=0.2,0.175$), equals ~ 6 km/s. Such values of the velocities lead to a Poisson ratio $\sigma \approx 0.35$. Analysis of the relations between the frequencies in the low-frequency spectrum of the resonances, which was performed according to Ref. 6, also leads to the same value of σ.

Figure 4 shows the temperature dependences of the relative changes in V_t and V_l for samples with different compositions. Except for the composition with $x=0.1$, near the

Curie points the velocities have either a weak minimum or a kink (see also Fig. 3), which are characteristic for ordinary ferromagnetic materials.[10] Below T_C the velocities grow gradually, approximately by 5% down to temperatures 250–200 K, reaching a flat maximum. The subsequent behavior of the velocities depends strongly on the composition. For $x=0.25$ the velocity V_t decreases by 5% down to temperatures ~ 80 K. For one of the two samples with composition $x=0.2$ ($T_C=305$ K, $La_{0.8}Sr_{0.2}Mn_{0.9}O_{2.95}$) the velocity V_t drops by approximately 20% without exhibiting any anomalies down to $T=70$ K. At the same time, for the other sample ($T_C=300$ K, $La_{0.82}Sr_{0.18}Mn_{0.82}O_{2.66}$), after decreasing by 5%, V_t increases abruptly by 20% at $T_s \approx 90$ K, corresponding to the temperature of a structural transition from the rhombohedral to the orthorhombic phase.[11,12] For the composition with $x=0.175$, the thickness resonances could be traced over the entire temperature interval and, accordingly, the value of V_l could be determined. After reaching the maximum at the point of the structural transition, $T_s \approx 180$ K, V_l undergoes an impressive jump by 30% (a hysteresis of up to 20°) and then increases gradually by another 6%.

The $x=0.1$ sample behaves even more dramatically. Sound is first excited at $T \approx 200$ K, i.e., long before the Curie temperature ($T_C \approx 170$ K). As the temperature decreases, the velocity gradually decreases, by approximately 15%, passing through the Curie point without any appreciable anomalies. This could be due to the more broadened character of the ferromagnetic transition as compared with other compositions, as can be seen from the $\chi_{ac}(T)$ curves in Fig. 2. The velocity V_t reaches a minimum at $T_p \approx 140$ K, after which it sustains a strong, jumplike increase (with a weak, 5-degree hysteresis). The magnitude of the jump could not be determined exactly, since it was impossible to establish a correspondence between the modes before and after the jump because of the depletion of the frequency spectrum. However, it can be asserted that it is not less than 30%. The temperature T_p corresponds to the second phase transition in the composition, which, according to neutron diffraction data,[13] is due to polaron ordering. It is associated with an ordered arrangement of the heterovalent ions Mn^{3+}/Mn^{4+} in alternating (001) planes and the appearance of a corresponding superstructure. A transition to this state is accompanied by a change in the character of the conduction from metallic to semiconducting[11] and an increase (jump) in the permittivity as well as an additional anomaly (besides the Curie point) in the magnetic susceptibility $\chi_{ac}(T)$.[9] The latter is also clearly observed on our measured $\chi_{ac}(T)$ curve for the composition $x=0.1$ (Fig. 2).

In summary, the anomalies observed in the temperature dependences of the sound velocities in the compounds $La_{1-x}Sr_xMnO_3$, on the whole, agree well with the temperatures of known magnetic and structural transitions and attest to a strong change in the elastic constants of these materials at phase transformations. The magnitudes and character of the anomalies of the elastic constants at the Curie points T_C are similar to those for conventional ferromagnetic metals with unfilled d and f shells.[10] The most radical changes in the elastic constants (up to 30%) accompany structural transitions at T_s and T_p and indicate a strong restructuring of the phonon spectrum in these compounds.

In closing, we should mention that very recently there has appeared a report of measurements of the elastic constants of a $0.172\times0.139\times0.117$ cm $La_{0.83}Sr_{0.17}MnO_3$ single crystal in the temperature range 200–310 K.[14] An abrupt decrease of the elastic constants (up to 60%) was observed at the structural transition ($T_s=285$ K). This does

not agree with the increase that we observed in the velocities at nearly the samee composition, $x = 0.175$.

We thank I. F. Voloshin for assisting in the supplementary analysis of the compositions of the samples.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-17325) and the program ''Statistical Physics'' of the Ministry of Science of the Russian Federation.

[a)]e-mail: gaidukov@lt.phys.msu.su

---

[1] R. von Helmont, J. Wecker, B. Holzapfel *et al.*, Phys. Rev. Lett. **71**, 2331 (1993).
[2] A. P. Ramirez, P. Shiffer, S. W. Cheong *et al.*, Phys. Rev. Lett. **76**, 3188 (1996).
[3] H. Fujishiro, M. Ikebe, Y. Konno *et al.*, J. Phys. Soc. Jpn. **66**, 3703 (1997).
[4] A. M. Balbashov, S. G. Karabashev, Ya. M. Mukovskiy *et al.*, J. Cryst. Growth **167**, 365 (1996).
[5] V. F. Gantmakher and V. T. Dolgopolov, JETP Lett. **5**, 12 (1967).
[6] Yu. P. Gaĭdukov, N. P. Danilova, and O. A. Sapozhnikov, Akust. Zh. (1998), at press.
[7] W. P. Mason (Ed.), *Physical Acoustics*, Academic Press, New York, 1969, Vol. 5; Mir, Moscow, 1973, Vol. 5, p. 134.
[8] A. Asamitsu, Y. Moritomo, R. Kimai *et al.*, Phys. Rev. B **54**, 1716 (1996).
[9] V. Yu. Ivanov, V. D. Travkin, A. A. Mukhin *et al.*, J. Appl. Phys. **83**, 1 (1998).
[10] K. P. Belov, M. A. Belyanchikova, R. Z. Levitin, and S. A. Nikitin, *Rare-Earth Ferro- and Antiferromagnets* [in Russian], Nauka, Moscow, 1965.
[11] A. Urushibura, Y. Moritomo, T. Arima *et al.*, Phys. Rev. B **51**, 14103 (1995).
[12] Y. Moritomo, A. Asamitsu, Y. Tokura, Phys. Rev. B **56**, 12190 (1997).
[13] Y. Yamada, O. Hono, S. Nohdo *et al.*, Phys. Rev. Lett. **77**, 904 (1996).
[14] T. W. Darling, A. Migliori, E. G. Moshopoulou *et al.*, Phys. Rev. B **57**, 5093 (1998).

Translated by M. E. Alferieff

Electric-field effect in a mixed-valence ion pair $Cr^{3+}-Cr^{2+}$ in a $KZnF_3$ crystal

S. I. Nikitin,[a)] R. V. Yusupov, M. V. Eremin, N. I. Silkin, and S. Yu. Prosvirnin
Kazan State University, 420008 Kazan, Russia

(Submitted 24 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 147–149 (25 July 1998)

The influence of an external axial electric field on the absorption spectrum of a $Cr^{3+}-Cr^{2+}$ mixed-valence pair center coupled by the double-exchange mechanism in a $KZnF_3$ crystal is investigated. It is shown experimentally that the $Cr^{3+}-Cr^{2+}$ pair has an electric dipole moment. The migration of an electron is accompanied by local lattice strain. At the minima of the adiabatic potential, the intermediate fluorine atom is displaced from the lattice site. © *1998 American Institute of Physics.* [S0021-3640(98)01014-7]

PACS numbers: 71.70.Ej, 71.55.Ht, 78.40.Ha

Mixed-valence materials such as $La_{1-x}Ca_xMnO_{3+y}$ are of great research interest. Convenient model objects for understanding the main mechanisms determining the properties of these compounds are insulating crystals with mixed-valence impurity pair centers.[1] In this case, the high conductivity characteristic of concentrated compounds (like, for example, lanthanum manganites) does not interfere with the penetration of light into the crystal, and optical spectroscopy methods can be used.

The observation of quite intense and relatively narrow absorption lines of $Cr^{3+}-Cr^{2+}$ pair centers in $KZnF_3:Cr^{3+}$, Cr^{2+} crystals was reported in Ref. 1. An interpretation of all the observed absorption lines of this pair was given in Ref. 2. According to the calculations, the minima of the adiabatic potential are due to compression of the fluorine octahedron around the Cr^{3+} ion and a corresponding stretching of the octahedron along the pair axis near a Cr^{2+} ion. An important feature of the state of a pair is that the bridge fluorine ion is displaced from the lattice site. This feature is also fundamental for the interpretation of the absorption spectra of this pair center.

A direct proof of such an important conjecture about the microstructure of the pair center under study would be the observation of a linear Stark effect. This circumstance prompted us to undertake an investigation of the electric-field effect on the absorption lines of the pair center $Cr^{3+}-Cr^{2+}$.

The $KZnF_3:Cr^{3+}$, Cr^{2+} crystals were grown by the Bridgman–Stockbarger method. The concentration of each activator CrF_3 and CrF_2 in the charge was ~ 0.5 wt.%. $KZnF_3$ crystals have the perovskite structure and their symmetry space group is $Pm3m$. In Ref. 2 it was shown that the pair centers $Cr^{3+}-Cr^{2+}$ in a $KZnF_3$ crystal possess tetragonal symmetry, and the chromium ions are located in neighboring unit cells arranged along the

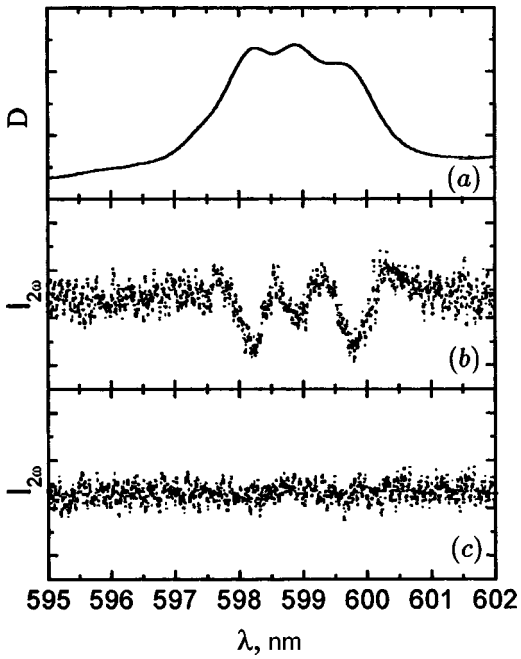

FIG. 1. Absorption spectrum (a) and the spectrum of the differential transmission signal (b, c) on an absorption line of a pair center $Cr^{3+}-Cr^{2+}$ in a $KZnF_3$ crystal near 600 nm with the application of an electric field $\epsilon \| C_4$; $\epsilon_0 = 120$ kV/cm, $T = 80$ K. Polarization of the light: b — $E \| \epsilon$, c — $E \perp \epsilon$.

C_4 axis of the crystal. For this reason, the Stark effect was measured with the electric field applied along the tetragonal axis of the crystal. The experimental samples were prepared in the form of $0.5 \times 5 \times 10$ mm slabs. The C_4 axis of the crystal was perpendicular to the plane of the slabs. Perfect cleavage surfaces were used to orient the samples. The orientation accuracy was checked on a DRON-2 x-ray diffractometer and was equal to $\pm 2°$.

The measurements were performed at sample temperature ~ 80 K. The maximum electric field intensity was 150 kV/cm. Since the absorption lines of a pair center have a comparatively large width ($\sim 100\,cm^{-1}$), a sensitive differential method, described in Ref. 3, was used to detect small splittings hidden within the linewidth.

When an ac electric field $\epsilon = \epsilon_0 \cos \omega t$ is applied to the experimental sample, a differential signal is detected at frequency 2ω. This signal is due to changes in the transmittance of the sample, which in turn are caused by shifts of the absorption lines, corresponding to orientations of the absorbing centers parallel and antiparallel to the electric field.

The experimental results for the $\lambda_{max} = 598$ nm absorption line, which according to the interpretation of Ref. 2 corresponds to a transition to the excited state $Cr^{3+}(^4A_2) - Cr^{2+}(^3E)$, are displayed in Fig. 1. One can see that when the light transmitted through the sample has the polarization $E \| \epsilon$, a differential signal corresponding to the form of the second derivative of the absorption line is observed. For the polarization $E \perp \epsilon$ the signal at the frequency 2ω is not observed.

The differential signal is not observed on the $\lambda_{max}=503$ nm line, corresponding to a transition to the excited state of the pair center $Cr^{3+}(^4T_2)-Cr^{2+}(^3T_1)$, in electric fields up to 150 kV/cm in either $E\perp\epsilon$ or $E\|\epsilon$ polarization. This can probably be explained by the fact that this line is weaker and wider. If it is assumed that the splittings of the $\lambda_{max}=503$ nm and $\lambda_{max}=598$ nm lines in an electric field are approximately the same in magnitude, then the amplitude of the differential signal on the $\lambda_{max}=503$ nm line should be an order of magnitude smaller than on the $\lambda_{max}=598$ nm line.

The observation of a signal, having the same form as the second derivative of the absorption spectrum, at the doubled modulation frequency of the electric field attests unequivocally to a linear Stark effect. In addition, the effective transition dipole moment corresponding to the $\lambda_{max}=598$ nm absorption line is oriented predominantly (this ''hedging'' is necessary because of the low signal/noise ratio) along the tetragonal axis of the crystal.

For a pair center with dynamic electron transfer, the presence of an effective dipole moment attests to a displacement of the bridge fluorine ion from the lattice site. The dynamics of ion motion associated with hopping of an electron between chromium ions is not manifested in the absorption spectra, since the electron hopping frequency is low compared with the optical transition frequency. It is logical to infer that the adiabatic potential for the excited state $Cr^{3+}(t_2^3,\ ^4A_2)-\ Cr^{2+}(t_2^3,\ ^4A_2, e;\ ^3E)$ is essentially identical to the adiabatic potential in the ground state $Cr^{3+}(t_2^3,\ ^4A_2)-\ Cr^{2+}(t_2^3,\ ^4A_2, e;\ ^5E)$. The displacements of the bridge fluorine ion at the minima of the adiabatic potentials should also be close. This is the likely explanation of the small magnitude of the effective dipole moment of the $\lambda_{max}=598$ nm line.

Thus, the present measurements have shown that the mixed-valence pair $Cr^{3+}-Cr^{2+}$ does not possess a center of inversion. Migration of excess charge is accompanied by local lattice strain. At the minima of the adiabatic potential the intermediate ion F^- is displaced from the lattice site, as a result of which the pair acquires an electric dipole moment. The results obtained reveal the fundamental features of the realization of double exchange and are apparently applicable in general to all systems with variable valence.

This work was supported by grants from the Russian Fund for Fundamental Research (Grants Nos. 97-02-185-98 and 98-02-18009).

[a)]e-mail: Sergey.Nikitin@ksu.ru

[1]M. V. Eremin, S. I. Nikitin, N. I. Silkin *et al.*, JETP Lett. **61**, 612 (1995).
[2]M. V. Eremin, S. I. Nikitin, S. Yu. Prosvirnin *et al.*, JETP (1998), in press.
[3]A. A. Kaplyanskiĭ, V. N. Medvedev, and A. P. Skvortsov, Opt. Spektrosk. **29,** 905 (1970) [Opt. Spectrosc. **29**, 481 (1970)].

Translated by M. E. Alferieff

Lunar ice: Can its origin be determined?

A. A. Berezhnoĭ
P. K. Shternberg State Astronomy Institute, 119890 Moscow, Russia

B. A. Klumov
Institute of the Geosphere Dynamics, 117979 Moscow, Russia

(Submitted 24 June 1998)
Pis'ma Zh. Éksp. Teor. Fiz. **68**, No. 2, 150–154 (25 July 1998)

The comet hypothesis of the origin of lunar ice, which was recently discovered in the polar regions of the moon by Lunar Prospector, is examined. It is shown that a comet impact produces a temporary atmosphere whose volatile component accumulates essentially completely in cold traps — the permanently shadowed regions of the Moon. The condensation of volatile compounds leads to the formation of ice with a definite chemical and isotopic composition, which contains important information about the composition of the comet. This hypothesis can be checked during the next lunar missions, and if confirmed, definite progress will have been made in understanding the nature of comets.
 [S0021-3640(98)01114-1]

PACS numbers: 96.20.Dt

The possible existence of ice in the lunar polar regions was first considered in Ref. 1. The lunar equator is tilted with respect to the ecliptic by approximately 1.5°. For this reason, permanently shadowed areas of the lunar surface exist at the bottom of numerous craters in the polar regions. The surface temperature of these areas is so low that they can be cold traps for water and for a number of other volatile compounds. On the illuminated sections, however, sunlight rapidly dissociates water molecules, and the products of dissociation escape from the Moon's gravity. For this reason, if water does exist on the lunar surface, then it exists only in cold traps and in the form of ice.[a)]

Investigation of the radar properties of the lunar surface from the space probe Clementine revealed regions with anomalous radar properties near the lunar south pole. This apparently attests indirectly to the presence of ice on the Moon.[3] Lunar Prospector was launched in January 1998. One of its main tasks was to investigate the lunar polar regions by neutron spectroscopy to confirm these results.

The first results of the Lunar Prospector mission were announced in March 1998.[4] According to preliminary data, the area S_c of the ice-occupied cold traps equals $\sim 1.5 \times 10^{14} - 7 \times 10^{14}\,\mathrm{cm}^2$, the volume fraction δ_{ice} of water ice in a half-meter layer of soil near the surface (deeper layers are inaccessible for measurements by this method) $\simeq 0.5-2\%$, while the mass of the water ice in the lunar polar regions is estimated to be $10^{13}-10^{15}$ g. What is the origin of this ice?

The content of volatiles (such as, for example, H_2O, CO_2, SO_2) in a subsurface

layer of soil in cold traps is controlled by the following processes: The sources of volatiles are comet and asteroid impacts, micrometeor bombardment, and solar wind, while losses of these compounds are due to micrometeorite bombardment and the solar wind.[5] We note that volatiles are lost only from traps; the sources, however, are distributed over the entire lunar surface and their effectiveness is determined by the probability of volatiles being captured in cold traps from the part of the surface under study.

Let us examine in greater detail the collision of a comet with the Moon. Typical impact parameters for the Moon are: Comet size $d_i \sim 2$ km, comet velocity $v_i \simeq 15 - 50$ km/s (most likely velocities $v_i \sim 20 - 25$ km/s), and comet density $\rho_i \approx 1$ g/cm^3 (comet mass $m_i \simeq 4 \times 10^{15}$ g, impact energy $E_i \sim 4 \times 10^{27} - 4 \times 10^{28}$ ergs), and lunar soil (regolith) density $\rho_t \approx 1.8$ g/cm^3. Under such an impact the following occur in succession: essentially complete vaporization of the cometary matter; vaporization, melting, and fragmentation of the regolith by the shock wave from the explosion; ejection of the mixed multiphase cometary matter and lunar soil; and, formation of an impact crater.

The mass m_{ej} of the material ejected on impact can be estimated, using the relation from Ref. 6, as

$$m_{ej}/m_i \simeq 0.2\left[\frac{v_i^2}{g}(\rho_t/m_i)^{1/3}\right]^{0.46}, \tag{1}$$

where g is the acceleration of gravity (for the Moon $g = 167$ cm/s^2) and m_i is the mass of the impactor. In our case ($v_i \sim 20 - 25$ km/s) one has $m_{ej}/m_i \sim 10^2$, the mass of the liquid phase $\sim 10\, m_i$, and the mass of the impact vapor $\sim (1-3) \cdot m_i$ (Refs. 7–9).

Some of the matter escapes from the Moon. Using the velocity distribution of the ejected matter[6]

$$m(>v_{\rm esc})/m_i \simeq 0.1(\rho_i/\rho_t)^{0.2}(v/v_{\rm esc})^{1.2}, \tag{2}$$

where $v_{\rm esc} = 2.4$ km/s is the escape velocity for the Moon and $m(>v_{\rm esc})$ is the mass of the ejected matter with velocities exceeding $v_{\rm esc}$, it is easily estimated for the impact parameters considered that $m(>v_{\rm esc})/m_i \sim 1$, i.e., the mass of the matter which escapes from the Moon is of the order of the impactor mass, i.e., mainly hot impact vapor escapes from the Moon's gravity. As numerical calculations show, the velocity distribution along the radius r in the impact vapor cloud is close to linear: $v(r) \sim v_{\rm max} r/R_{\rm max}$, where $v_{\rm max}$ is the expansion velocity of the hot cloud and in our case $v_{\rm max} \sim v_i/2 - v_i/3$. The cloud expands freely; its radius grows with time as $R_{\rm max} \simeq v_{\rm max} t$. The mass of the vapor of cometary origin that remains in the lunar gravitational field, i.e., that has a velocity less than $v_{\rm esc}$, is of the order of $m_i(v_{\rm esc}/v_{\rm max})^3 \sim (10^{-2} - 10^{-1}) m_i$. We note that as the impactor velocity increases, $v_{\rm max}$ ceases to depend on v_i: $v_{\rm max} \sim \sqrt{\eta m_i v_i^2/m_v}$, where η is the fraction of the impactor kinetic energy that is converted into the impact vapor and m_v is the mass of the impact vapor. For large v_i ($v_i^2 \gg e_v$, where e_v is the heat of vaporization of the soil) $m_v \propto v_i^2$ and $v_{\rm max}$ no longer depends on v_i. The fraction of volatiles in the impact vapor, however, decreases as the impactor velocity increases $\propto m_i/m_v \propto v_i^{-2}$. We assume that the cometary and target matter are well mixed in the impact vapor. Indeed, by estimating the growth rates of the Rayleigh–Taylor and Kelvin–Helmholtz instabilities, it can be shown that for typical collision parameters these instabilities can develop over characteristic impactor stopping times $\sim d_i/v_i$, which leads to efficient mixing of the cometary matter and the lunar soil.

In summary, the falling of a comet leads to the formation of an atmosphere on the Moon. This atmosphere consists of the cometary matter and the lunar soil, and the ratio of these materials depends on the impact velocity: For impacts with velocities $v_i \simeq 20 - 25$ km/s the ratio is of the order of 1. We note that condensation processes in the expanding impact vapor can increase the mass of cometary matter on the Moon and, correspondingly, because of the rapid vaporization of the matter by sunlight they can increase the mass of the atmosphere. What is the composition of such an atmosphere?

As the hot cloud expands, quenching occurs: The chemical composition of the impact vapor stops changing, and corresponds to the equilibrium composition at the moment of quenching. Let the parameters characterizing the quenching of the impact vapor be the same as for the comet Shoemaker–Levy 9 ($T \sim 1500$ K, $P \sim 0.01$ bar).[10] The chemical composition of the impact vapor is determined by the composition of the comet and the lunar soil. However, thermodynamic calculations show that since the lunar soil does not contain significant quantities of volatiles, the addition of lunar matter to the cometary matter essentially does not change the chemical composition of the volatiles in the impact vapor. This is valid for ratios $m_i/m_v \geqslant 0.03$, which holds for the entire range of comet–Moon collision velocities.

We shall assume that the elemental composition of a comet is identical to that of Halley's comet.[11] Then the main compounds of the impact-produced atmosphere are: H_2 ($\simeq 40\%$), CO_2 ($\simeq 30\%$), H_2O ($\simeq 20\%$), CO_2 ($\simeq 5\%$), N_2 ($\simeq 3\%$), S_2 ($\simeq 3\%$) and trace components SO ($\simeq 0.3\%$) and SO_2 ($\simeq 0.3\%$). We note that sulfur-containing compounds will also enter the composition of the impact-produced atmosphere, since at the stage of the explosion of the comet the reaction binding sulfur in the form of iron sulfide (which corresponds to fixing of sulfur into a solid phase) is kinetically inhibited.

To determine which compounds will condense in cold traps, let us estimate the surface temperature T_c of cold traps and the partial pressure of the indicated compounds of the impact atmosphere. From the balance equation $(T_c/T_{eq})^4 = a_M \cos(\theta + \alpha) d_c/D_c$, where $T_{eq} = 400$ K is the surface temperature on the day side, $a_M \approx 0.1$ is the albedo of the lunar surface, d_c and D_c are, respectively, the trap depth and diameter, $d_c/D_c \simeq 0.2$, θ is the trap latitude, and α is the declination of the Sun. This gives $T_c \simeq 50-100$ K. If the traps are not heated by the scattered solar radiation, T_c is determined by internal sources of heat and $\simeq 30$ K.

In the case of the impact of a comet with $d_i \simeq 2$ km the mass of the impact-produced atmosphere is $\sim 10^{14}$ g, the pressure at the surface is $p \simeq 10^{-8}$ bar, and the density of neutrals is $N_a \sim 10^{11} - 10^{12}$ cm^{-3}. Under these conditions, condensation of H_2O, CO_2, S, and SO_2 can occur in cold traps. The lifetime τ_a of such an atmosphere is determined by the photolysis time τ_{ph} of the main compounds in the exosphere, since the products of photolysis escape the Moon's gravity: $\tau_a \sim \tau_{ph} N_a/N_{ex}$, where N_{ex} is the density of neutrals in the exosphere: $N_{ex} \sim (\sigma H_a)^{-1}$, where $\sigma \simeq 10^{-15}$ cm^2, is the characteristic cross section of elastic collisions for the compounds under study, and $H_a \simeq 100$ km is the scale height of the impact-produced atmosphere. In our case $\tau_a \sim 10^{10}$ s.

The condensable compounds of the atmosphere diffuse into cold traps in a characteristic time $\tau_d \sim (S_M/S_c) H \langle d_c \rangle N_a \sigma / v$, where S_M and S_c are, respectively, the area of the lunar surface and the area of cold traps; $\langle d_c \rangle$ is the characteristic crater depth; v is the thermal velocity; and, the ratio $\tau_d/\tau_a \approx (S_M/S_c)\langle d_c \rangle / \tau_{ph} v \leqslant 0.1$.

Therefore the condensable gases of the impact-produced atmosphere essentially all fall into cold traps. It is easy to show that the temperature of the cold traps remains essentially unchanged during the condensation of the compounds under study.

In summary, the impact of a comet with $d_i \simeq 2$ km leads over times $t \sim t_d \simeq 10^9$ s to the accumulation of $\sim 10^{14}$ g H_2O in cold traps with surface density $\delta_{\rm ice}^{\rm comet} \sim 0.3\ {\rm g/cm^2}$, sulfur S $\sim 5\times 10^{13}$ g, $CO_2 \sim 5\times 10^{13}$ g, $SO_2 \leqslant 5\times 10^{12}$ g (for SO_2 $\tau_d/\tau_a \geqslant 1$). The composition of such ''cometary'' ice is determined by the composition of the condensable compounds of the impact-produced atmosphere: The relative concentrations of volatiles in the temporary atmosphere are essentially the same as in the cold traps.

The subsequent redistribution of this ice is determined by the rate of micrometeorite processing of the surface of the cold traps. It is known that the cumulative flux $F(d)$ of particles of size greater than d falling on the Moon follows a power law $F(>d) \propto d^{-\alpha}$, where $\alpha \simeq 3.5$ for particles of size $d \sim 10^{-2} - 10^2$ cm (see, for example, Ref. 6). Since the depth of a microcrater is $\simeq d$ and the area of a microcrater $S_c(d) \propto d^2$, the characteristic time t_g for reprocessing of the surface at a depth h depends on depth as $t_g \propto F(>h)^{-1} S_c(h)^{-1} \propto h^{\alpha-2}$.

According to Apollo data,[12] a layer at depth $d = 1$ cm mixes over a time $t_{1\ {\rm cm}} \simeq 10^6$ yr. A layer at depth 1 m mixes over a time $t_{1\ m} \simeq t_{1\ {\rm cm}} \times 10^{2\alpha-4} \simeq 10^9$ yr. Therefore over the lifetime of the traps ($\simeq 2-3\times 10^9$ yr) their surface becomes well mixed to depth $h_m \simeq 1$ m, while the volume content of the ice in the soil in the case of the impact of one comet $\simeq \delta_{\rm ice}^{\rm comet}/d_m \sim 0.3\%$, which agrees with Lunar Prospector data. We note that the CO_2 and SO_2 content in the soil is very sensitive to soil temperature, so that the concentrations of these compounds can decrease appreciably over the lifetime of the traps.

We tacitly assumed above that micrometeorite bombardment leads only to mixing of the soil. However, this is not the case in reality. Undoubtedly, a definite dissipation of water of cometary origin occurs by means of the mechanisms considered above, but these mechanisms are inefficient, since the velocities of collisions of micrometeorites with the Moon ($v_i \sim 5-15$ km/s) are much lower than that of comets, and the products of the explosion essentially do not escape the Moon's gravity. In the process, a definite dilution of the ''cometary'' water by water of meteoritic origin occurs.

The number of impacts of large celestial bodies that the moon has experienced over the lifetime of the traps can be estimated as follows. By analyzing the cumulative age distribution of the craters (see, for example, Ref. 6), the fraction δ_y of the lunar surface occupied by craters of size $D_c > 4$ km, which are younger than the cold traps, can be determined to be $\delta_y \simeq 10^{-4}$. The mass distribution of large celestial bodies of size $d \sim 0.1-10$ km that have collided with the Moon is also well known: $N(>m) \propto m^{-0.55}$ (Ref. 6).

The crater size d_c depends on the impactor velocity v_i and the impactor size d_i: $d_c \propto d_i^{0.8} v_i^{0.44}$. For impact parameters of interest to us ($d_i \sim 1-3$ km, $v_i \sim 20-50$ km/s) $d_c \simeq 10-20$ km. The number of impacts of large celestial bodies over the entire history of the cold traps, with allowance for the dependence $N(>m)$, is of the order of $\sim \delta_y S_{\rm M}/10\pi D_c^2 \sim 10$. The comet fraction of the celestial bodies which have collided with the Moon remains an open question. We note that to explain the amount of ice observed by Lunar Prospector, a single impact of a comet with $d_i \simeq 2$ km is sufficient.

The amounts of water observed by Lunar Prospector in cold traps can also be

explained by the action of the solar wind and/or micrometeorite bombardment of the lunar surface. However, in these cases the composition of the ice will be different.

An important feature of ice of cometary origin is the anomalous high content of deuterium. The ratio D/H in comets is of the order of $\sim 10^{-3}$ (in the coma of comet Hale–Bopp[13]). Thermodynamic calculations show that as the impact cloud expands, enrichment of the water by deuterium by approximately a factor of 2–2.5 occurs as a result of isotopic exchange.

The isotopic composition of hydrogen formed by the interaction of the solar wind with lunar soil is identical to that of the solar wind ($D/H \leqslant 1.5 \times 10^{-4}$), while D/H in water of meteoritic origin corresponds to the isotopic composition of hydrogen in meteorites ($D/H \sim (2-5) \times 10^{-4}$).

Ices of cometary origin can also be distinguished by analyzing the chemical composition of the ice. In meteorite impacts, sulfur and sulfur-containing compounds enter the lunar exosphere. Cold traps capture only sulfur; sulfur-containing compounds are not captured because of the short photolysis time. The solar wind, however, is an inefficient source of sulfur.

In summary, the falling of comets onto the Moon leads to formation of ice with definite chemical and isotopic composition in cold traps: The ratio D/H is much higher than for other sources of water. Another distinguishing property of ''cometary ice'' is that it could contain the compounds SO_2 and CO_2. The comet hypothesis of the origin of lunar ice can be checked during the next lunar missions, and if confirmed, then definite progress will have been made in understanding the composition and, accordingly, the nature of comets.

We thank B. I. Ivanov for helpful discussions. One of us (B. A. A.) was supported in part by the State Science and Technology Program ''Astronomy'' (Grant No. 1.6.1.1).

[a)]The possibility of water being transported into deep (~ 100 m) layers of the lunar soil by a comet fall was studied in Ref. 2. It was shown that large cracks present in the soil at the impact location can become partially filled by the cometary matter. However, if the impact location is located in an illuminated section of the lunar surface, then the temperature regime in the soil makes retention of volatiles impossible.

[1] K. Watson, B. C. Murray, and H. Brown, J. Geophys. Res. **66**, 3033 (1961).
[2] T. I. Orlova and V. V. Svetsov, Earth, Moon, Planets **71**, 255 (1995).
[3] S. Nozette, C. L. Lichtenberg, P. Spudis *et al.*, Science **275**, 1495 (1996).
[4] http://lunar.arc.nasa.gov/science/results/lunarice.
[5] J. R. Arnold, J. Geophys. Res. **84**, 5659 (1979).
[6] T. Gehrels (Ed.), *Hazards Due to Comets and Asteroids*, Univ. of Arizona Press, 1994.
[7] K. R. Housen and R. M. Schmidt, J. Spacecr. Rockets **32**, 162 (1995).
[8] J. D. O'Keefe and T. J. Ahrens, J. Geophys. Res. **87**, 6668 (1982).
[9] M. J. Cintala, J. Geophys. Res. **97**, 947 (1992).
[10] A. A. Berezhnoĭ, B. A. Klumov, V. E. Fortov, and V. V. Shevchenko, JETP Lett. **63**, 405 (1996).
[11] E. Anders, N. Grevesse, Geochim. Cosmochim. Acta **53**, 197 (1989).
[12] R. V. Morris, *Proceedings of the Lunar and Planetary Science Conference*, 1978, p. 1801.
[13] R. Meier *et al.*, Science **279**, 1707 (1998).

Translated by M. E. Alferieff