JETP Letters, Vol. 82, No. 6, 2005, pp. 317–318. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 355–357.

Appeal to Authors and Readers

V. F. Gantmakher, Editor-in-Chief of the journal *Pis'ma v ZhÉTF*

The year 2005 marks forty years since the foundation of the journal *Pis'ma v ZhÉTF*. It appeared in 1965 and soon after gained a high degree of prestige in the physics community, becoming one of the leading physics journals in the Soviet Union. Its English translation was begun immediately with the foundation of its identical English-language edition *JETP Letters*.

In order to retain its role in the current, rapidly changing world, the journal was reorganized in the 1990s: it became open and international and began to present the achievements of not only Russian physicists but also of scientists from all over the world. Since 1991, articles in English have been accepted and published in the journal, and manuscripts have been received via e-mail since 1994. This means that foreign scientists, as well as Russian scientists working abroad or in remote regions of Russia, can send their manuscripts to *Pis'ma v ZhÉTF* in either English or Russian as easily as Moscow-based authors submit their printed manuscripts.

As a result, we now receive a large number of articles from abroad. The numbers presented below illustrate the source countries of our articles.

DATA ON THE PAST TEN YEARS (1995–2004)[1]

The total number of published articles is 3114

Works were presented by authors from scientific organizations of various countries. The amounts of such presentations from various countries are[2]

Russia 3626

Former USSR republics 231

Other countries 1077

The distribution of the presentations from Russian regions and various countries is as follows.

Russia

Moscow 1423

Moscow region 982

St. Petersburg and Leningrad region 328

Novosibirsk 288

Krasnoyarsk 113

Nizhni Novgorod 109

Yekaterinburg 103

Kazan 78

Tomsk 36

Voronezh 15

Saratov 15

Other cities 136

Former USSR republics

Ukraine 172

Armenia 23

Belarus 19

Others 17

Other countries

Germany 264

France 151

United States 141

United Kingdom 81

Japan 67

Italy 55

Finland 54

Switzerland 34

Sweden 28

The Netherlands 28

Poland 26

Israel 25

Belgium 16

Spain 11

Czech Republic 10

Others (24 countries) 86

The main achievement of the current editorial board is the rapid publication of the articles. To this end and in order to ensure distinct and judicious selection of articles, we follow the tradition of our predecessors, who always ensured that the editorial board was comprised only of actively working scientists possessing their own opinions in their areas of research and who could themselves evaluate the essence and significance of the works without intermediaries. In inviting new members to the editorial board, we examine the creative potential of a person, his ability to understand and

[1] The statistical data were prepared by T.G. Tratas (Institute of Solid State Physics, Russian Academy of Sciences). They are presented more completely on the journal's website http://www.jetpletters.ac.ru.

[2] The numbers of presentations are larger than the respective numbers of works, because the authors of one article can be from different organizations and different states.

0021-3640/05/8206-0317$26.00

evaluate a work, his personal nature, and his dedication to his responsibilities. Received manuscripts are immediately distributed according to the specialization of the members of the editorial board, who assign referees, meet with them or contact them by phone and e-mail, and formulate a preliminary decision. No less than twice a month does the editorial board meet for common discussion of all the works and to make final decisions.

Owing to the procedure developed for refereeing articles, the time interval from the submission of a manuscript to the journal to its publication has been reduced to 40 days (Russian edition). The preparation of the English-language edition takes an additional 40–45 days (this is currently performed by the International Academic Publishing Company Nauka/Interperiodica). Simultaneously with the transfer of an issue to press, we place it on our website, http://jetpletters.ac.ru, where it appears with free access 10–12 days before official publication of the issue. As a result, the real publication time is even shorter.

Confidence in the kindness and respectability of our editors strengthens the position of the editorial board in conflict situations. When conversing with authors who do not agree with our decisions, we always emphasize: {*We may have made a mistake in rejecting your article. If you think that your results are worthy of publication, send your manuscript again. However, please do not write us long letters with explanations. These explanations are needed not for us but for your future readers. We urge you to rewrite the article before sending it to us again, because our first decision means that something was wrong: perhaps your argumentation was weak, the proofs were unclear, or the explanations were doubtful, etc.*}. This suggestion works, and manuscripts submitted repeatedly are often accepted—often, but not always. The percentage of rejected articles is quite high. The numbers of articles submitted and accepted in recent years are as follows.

Year	1995	1997	1999	2001	2004
Submitted	705	604	614	561	550
Accepted	365	305	308	292	308

Several years ago, a new rubric Scientific Summaries appeared in the journal. Small reviews ordered by the editorial board and written by the heads of completed projects supported by the Russian Foundation for Basic Research are published under this rubric. Selection of such projects is also the prerogative of the members of the editorial board.

Thus, the members of the editorial board of *Pis'ma v ZhÉTF* play a more important role than in other journals in the choice of which articles are published in the journal; they also represent a source of indirect influence on the development of various scientific areas. For this reason, in the anniversary year of 2005, the editorial board has decided to give its editors the opportunity to express themselves under the rubric Scientific Summaries. This issue consists entirely of reviews of urgent subjects written by former and current editors of *Pis'ma v ZhÉTF* and thereby represents an original manifesto of the editorial board. We hope for a favorable reaction from our readers. From our authors, we await new articles presenting brilliant results.

Translated by R. Tyapaev

JETP Letters, Vol. 82, No. 6, 2005, pp. 319–324. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 358–363.

Vacuum Energy: Quantum Hydrodynamics vs. Quantum Gravity¶

G. E. Volovik*

Low Temperature Laboratory, Helsinki University of Technology, FIN-02015 HUT, Finland

Landau Institute for Theoretical Physics, Moscow, 119334 Russia

e-mail: volovik@boojum.hut.fi

Received June 6, 2005

We compare quantum hydrodynamics and quantum gravity. They share many common features. In particular, both have quadratic divergences, and both lead to the problem of the vacuum energy, which, in quantum gravity, transforms to the cosmological constant problem. We show that, in quantum liquids, the vacuum energy density is not determined by the quantum zero-point energy of the phonon modes. The energy density of the vacuum is much smaller and is determined by the classical macroscopic parameters of the liquid, including the radius of the liquid droplet. In the same manner, the cosmological constant is not determined by the zero-point energy of quantum fields. It is much smaller and is determined by the classical macroscopic parameters of the Universe dynamics: the Hubble radius, the Newton constant, and the energy density of matter. The same may hold for the Higgs mass problem: the quadratically divergent quantum correction to the Higgs potential mass term is also cancelled by the microscopic (trans-Planckian) degrees of freedom due to the thermodynamic stability of the whole quantum vacuum. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 03.70.+k, 04.90.+e, 67.90.+z, 98.80.Es

1. INTRODUCTION

The problem of quantum hydrodynamics is at least 65 years old (see quantization of the macroscopic dynamics of liquid in the first Landau paper on superfluidity of ^{4}He [1]). It is almost as old as the problem of quantum gravity [2]. Quantum hydrodynamics and quantum gravity share many common features (e.g., both have quadratic divergences), and they probably will have a common destiny. The main message from quantum hydrodynamics to quantum gravity is that, most probably, quantum gravity cannot be constructed, because quantum hydrodynamics cannot be constructed.

Of course, one can quantize sound waves in hydrodynamics to obtain quanta of sound waves—phonons. Similarly, one can quantize gravitational waves in general relativity to obtain gravitons. However, one should not use low-energy quantization for calculation of the radiative corrections that contain Feynman diagrams with integration over high momenta. In particular, the effective field theory is not appropriate for the calculation of the vacuum energy in terms of the zero-point energy of quantum fields. The latter leads to the cosmological constant problem in gravity [3, 4] and to the similar paradox for the vacuum energy in quantum hydrodynamics: in both cases, the vacuum energy estimated using the effective theory is many orders of magnitude too big. We know how this paradox is solved in quantum liquids, and we may expect that the same general arguments based on the thermodynamic stability of the ground state of a quantum liquid are applicable to a quantum vacuum.

There is another big discrepancy between the theory and experiment, which is called the hierarchy problem in the Standard Model [5]. It is believed that the mass of the Higgs boson is on the order of or somewhat larger than the mass of the gauge boson: $m_{\rm H}^2 \sim M_Z^2$. For example, in analogy with the effect of Cooper pairing in superconductivity, the Higgs mass can be equal to $2m_t$, where m_t is the mass for the top quark [6, 7]. However, the radiative correction to the Higgs mass is quadratically diverging, and it is determined by the ultraviolet cutoff. If one chooses the natural GUT or Planck scale as the cutoff energy, one obtains that the Higgs boson mass must be extremely large: $m_{\rm H}^2 \sim \pm 10^{26} M_Z^2$ and $m_{\rm H}^2 \sim \pm 10^{34} M_Z^2$ correspondingly (the sign of the radiative correction is determined by the relevant fermionic and bosonic content of the theory and the cut-off scheme). We argue that this discrepancy is related to the problem of the vacuum energy, and, thus, the same thermodynamic arguments that have been used for the cosmological constant problem can be applicable to the hierarchy problem.

¶ The text was submitted by the author in English.

*A member of the editorial board of the journal *JETP Letters* since 1991.

0021-3640/05/8206-0319$26.00 © 2005 Pleiades Publishing, Inc.

2. CLASSICAL HYDRODYNAMICS OF QUANTUM LIQUID

Let us consider the hydrodynamics of an isotropic superfluid liquid at $T = 0$ (such as the bosonic superfluid ^{4}He and the fermionic superfluid ^{3}He–B [8]). Though the superfluid liquid is essentially quantum, its macroscopic low-frequency dynamics is classical and is represented by classical hydrodynamics. It is background independent; i.e., it does not depend on the details of the underlying microscopic physics, and it is the same for fermionic and bosonic liquids. The equations of the nonrelativistic superfluid hydrodynamics (in the absence of quantized vortices) are the Hamilton equations for canonically conjugated fields:

$$\partial_t \rho = \frac{\delta H}{\delta \phi}, \quad \partial_t \phi = -\frac{\delta H}{\delta \rho}. \tag{1}$$

Here, ρ is the mass density, ϕ is the velocity potential (in the absence of quantized vortices the superfluid velocity is potential: $\mathbf{v} = \nabla\phi$), and the Hamiltonian is the energy functional of the liquid expressed in terms of ρ and ϕ:

$$H = \int d^3r \left[\frac{1}{2}\rho(\nabla\phi)^2 + \tilde{\epsilon}(\rho)\right], \tag{2}$$

where $\tilde{\epsilon}(\rho) = \epsilon(\rho) - \mu\rho$, $\epsilon(\rho)$ is the energy density of the liquid expressed in terms of the liquid density, and μ is the chemical potential—the Lagrange multiplier that takes into account the mass conservation $\int d^3r\rho = \text{const}$. At the fixed chemical potential μ, this functional has a minimum at $\mathbf{v} = 0$ and $\rho = \rho_0(\mu)$, where the equilibrium density ρ_0 is determined by the equation $d\tilde{\epsilon}/d\rho = 0$ (or $d\epsilon/d\rho = \mu$). This is the ground state of the liquid with the relevant energy density $\tilde{\epsilon}(\rho_0)$ and the relevant total energy

$$E_0 = V\tilde{\epsilon}(\rho_0). \tag{3}$$

Note again that this consideration is completely classical and operates with the quantities ρ_0 and $\tilde{\epsilon}(\rho_0)$, which are the classical output of the quantum system: the superfluid ^{4}He and superfluid ^{3}He–B are systems of strongly correlated, strongly interacting, and highly entangled helium atoms governed by the laws of quantum mechanics. The reason for the classicality is the macroscopic character of the collective motion.

3. QUANTIZED HYDRODYNAMICS

What happens if we try to construct the quantum hydrodynamics, i.e., to quantize the hydrodynamic motion of the liquid determined by Eqs. (1). If we only use hydrodynamic variables, we are unable to reconstruct the whole microscopic Hamiltonian for the interacting atoms. This is because, from the big realm of the complicated quantum motion of the ^{4}He atoms, we have chosen only the hydrodynamic modes whose wavelengths are much bigger than the interatomic spacing a, which plays the role of the Planck length: $ka \ll 1$. That is why, what we can do at best is to quantize the sound modes. However, even in this case, there is a danger of double counting, because, starting from the quantum system, we have obtained the classical behavior of the soft variables ρ and ϕ, and now we are trying to quantize them again. In particular, the energy E_0 in Eq. (3) is the whole energy of the quantum liquid, and it already includes (from the very beginning) the energy of those degrees of freedom that are described in terms of phonons. Let us see how this double counting typically occurs.

The conventional quantization procedure for sound waves in the background of the state with $\rho = \rho_0$ and $\phi = 0$ is the introduction of the commutation relations for the canonically conjugated variables:

$$[\hat{\phi}_{\mathbf{k}}, \hat{\rho}_{\mathbf{k}'}] = i\hbar\delta_{\mathbf{k}\mathbf{k}'}, \tag{4}$$

where

$$\hat{\rho}(\mathbf{r}) = \rho_0 + \frac{1}{\sqrt{V}}\sum_{\mathbf{k}}(\hat{\rho}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} + \text{c.c.}), \tag{5}$$

$$\hat{\phi}(\mathbf{r}) = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}}(\hat{\phi}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} + \text{c.c.}). \tag{6}$$

Introducing these perturbations to Eq. (2), one obtains the quantum Hamiltonian as the sum of the ground state energy and the Hamiltonians for the quantum oscillators:

$$\begin{aligned}\hat{H} &= E_0 + \frac{1}{2}\sum_{\mathbf{k}}\left(\rho_0 k^2|\hat{\phi}_{\mathbf{k}}|^2 + \frac{c^2}{\rho_0}|\hat{\rho}_{\mathbf{k}}|^2\right)\\ &= E_0 + \frac{1}{2}\sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}(a_{\mathbf{k}}a_{\mathbf{k}}^\dagger + a_{\mathbf{k}}^\dagger a_{\mathbf{k}}),\end{aligned} \tag{7}$$

where $\omega_{\mathbf{k}} = ck$, and the speed of sound c is given by $c^2 = \rho_0 d^2\epsilon/d\rho^2|_{\rho_0}$.

There is nothing bad about this quantization if we are constrained by the condition $ka \ll 1$. However, if we start to apply this consideration to the diverging quantities such as the vacuum energy, we are in trouble.

4. VACUUM ENERGY IN QUANTUM HYDRODYNAMICS

The vacuum energy—the ground state of the Hamiltonian (7)—contains the zero-point energies of quantum oscillators:

$$\langle \text{vac}|\hat{H}|\text{vac}\rangle = E_0 + \frac{1}{2}\sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}. \tag{8}$$

However, the vacuum expectation value of $\hat{H}$ must be equal to E_0 by definition. This is because E_0 is not the

"bare" energy, but it is the total relevant energy of the liquid, which includes all the quantum degrees of freedom of the liquid. Thus, the naive application of the zero-point energy leads to the paradoxical conclusion that

$$\frac{1}{2}\sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}} = 0. \quad (9)$$

This contradicts our intuition that the zero-point fluctuations give for the vacuum energy the estimate $\tilde{\epsilon}_{zp} \sim \hbar c k_{uv}^4$, where $k_{uv} \sim 1/a$ is the ultraviolet cutoff. However, Eq. (9) simply means that the zero-point energy of the phonons has already been included into the original E_0, together with all the other modes; i.e., by writing Eq. (9) we simply prevent the double counting of the vacuum energy. Thus, the correct form of the Hamiltonian for the phonons must be

$$\hat{H} = E_0 + \sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}. \quad (10)$$

One may ask what the role of the quantum fluctuations of the hydrodynamic field is. Do they provide the main or a substantial part of E_0? To see this, let us consider the vacuum energy of a superfluid liquid in the case when an external pressure P is applied to the liquid. According to the Gibbs–Duhem relation, which is valid for the equilibrium states, one has

$$P = TS + \mu\rho_0 - \epsilon(\rho_0). \quad (11)$$

At $T = 0$, one obtains that the vacuum energy is regulated by external pressure

$$\tilde{\epsilon} = -P. \quad (12)$$

For the positive external pressure $P > 0$, one obtains the negative energy density of the vacuum, $\tilde{\epsilon} < 0$, which certainly cannot be obtained by summation of the positive zero-point energies of the phonons. Thus, the contribution of the zero-point energies of phonons to E_0 gives us no idea regarding the total value of E_0, since it may even give the wrong sign of the vacuum energy.

Furthermore, for the superfluid helium liquid isolated from the environment, the pressure $P = 0$ and, thus, the vacuum energy density and the vacuum energy are zero: $\tilde{\epsilon}(\rho_0) = 0$, $E_0 = 0$. For the finite system—the helium droplet—the vacuum energy density becomes nonzero due to the capillary pressure:

$$\tilde{\epsilon} = -P = -\frac{2\sigma}{R}. \quad (13)$$

It is expressed through the classical parameters of the liquid droplet: its radius R and surface tension σ. When we compare this physical result with the naive estimation that only takes into account the contribution of the phonon zero-point energy $\tilde{\epsilon}_{zp} \sim \hbar c k_{uv}^4$, one finds that their ratio is determined by the ratio of the quantum microscopic scale to the classical macroscopic scales: $\tilde{\epsilon}_{true}/\tilde{\epsilon}_{zp} \sim a/R$. For the macroscopic bodies, the discrepancy is big.

5. APPLICATION TO QUANTUM VACUUM

This demonstrates that the vacuum energy is determined by the macroscopic thermodynamic laws and is not related to the diverging contribution of the zero-point motion of phonons. The result $E_0 = 0$ for the self-sustained homogeneous systems shows that, in this system, the large positive contribution $\tilde{\epsilon}_{zp} \sim \hbar c k_{uv}^4$ of phonons is completely compensated for without any fine tuning by the other quantum degrees of freedom, i.e., by the microscopic (atomic ≡ trans-Planckian) degrees of freedom that cannot be described in terms of the effective (hydrodynamic) field [9].

This compensation, which occurs in the homogeneous vacuum, does not prohibit the Casimir effect in the systems with boundaries. If the energy difference between two vacua comes solely from the long-wavelength physics, it is within the responsibility of the phonon (photon) modes and can be calculated using their zero-point energies.

One can immediately apply this lesson to quantum gravity: the vacuum energy density ϵ_{vac} (or cosmological constant) is not renormalized by the zero-point energies of the quantum fluctuations of the low-energy modes. It is meaningless to represent the vacuum energy as the sum over the zero-point oscillations $\frac{1}{2}\sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}$ and to estimate the vacuum energy density as $\epsilon_{zp} \sim \hbar c k_{uv}^4$. The vacuum energy (and thus the cosmological constant) is the final classical output of the whole quantum vacuum with all its degrees of freedom: sub-Planckian and trans-Planckian. It is regulated by macroscopic physics, and it obeys the macroscopic thermodynamic laws. The thermodynamic Gibbs–Duhem relation (the analog of Eq. (12)) must be satisfied for the equilibrium vacuum, and it does follow from the cosmological term in Einstein action:

$$\Lambda = \epsilon_{vac} = -P_{vac}. \quad (14)$$

For the vacuum isolated from the "environment" (free vacuum), the pressure is zero and thus the vacuum energy density is zero too: $\epsilon_{vac} = -P_{vac} = 0$ [9, 10]. This means that the natural value of the cosmological constant in the equilibrium homogeneous time-independent free vacuum is $\Lambda = 0$ rather than the contribution $\Lambda \sim \hbar c k_{uv}^4$ of the zero point energies of the effective quantum fields, which is completely compensated for.

In the case of the developing Universe polluted by matter, the vacuum energy is disturbed and the compensation is not complete. However, again, the natural value of Λ is determined not by the quantum zero-point energy but (as in quantum liquid in Eq. (13)) by the

classical macroscopic parameters of the Universe dynamics: the Hubble radius R of the Universe, the Newton constant G, and the energy density of matter ρ_{mat}. This implies that, depending on the details of the process, one has $\Lambda \sim \rho_{mat}$ or $\Lambda \sim 1/GR^2$ (or Λ is given by some combination of these factors). Both estimates are comparable to the measured value of Λ and are much smaller than the naive estimation of the zero-point energy of quantum fields: $\Lambda_{true}/\Lambda_{zp} \sim a^2/R^2 \sim 10^{-120}$. As in quantum hydrodynamics, this contains the ratio of the quantum microscopic scale $a = 1/k_{uv} = \hbar c/E_{Planck}$ to the classical macroscopic scale R.

6. HIGGS MASS PROBLEM

As distinct from quantum gravity, the Standard Model of electroweak and strong interactions is the low-energy effective theory, which can exist in the quantum form. The condensed matter experience demonstrates that this effective theory can be emergent. In condensed matter systems with point nodes in the energy spectrum of fermions, the effective gauge fields and chiral fermions gradually emerge in the low-energy corner. The point nodes are protected by topology in momentum space and, thus, are generic [9]; that is why the effective theory of the Standard-Model type is generic. Though this effective theory is the final output of the underlying quantum system, it can be quantized again, in spite of the ultraviolet divergences. The reasons for this is that above the symmetry breaking electroweak scale, i.e., at $k^2 \gg M_Z^2$, the divergences are logarithmic of the type $\ln(k^2/M_Z^2)$. Since the logarithm is concentrated mostly in the sub-Planckian region $k_{uv} \gg k^2 \gg M_Z^2$, it is within the jurisdiction of the low-energy effective theory.

At $k^2 \sim M_Z^2 \sim 10^{-34} E_{Planck}^2$, the symmetry breaking occurs and the nonzero vacuum expectation value of the Higgs field develops. One may expect that, at these extremely low energies as compared to the Planck or GUT scale, the effective theory must work well. Instead the ultraviolet problem arises. In particular, if one tries to calculate the quantum radiative corrections to the mass of the Higgs boson, the quadratic divergence occurs, which is outside the jurisdiction of the Standard Model. What can one say about this problem using the condensed matter experience?

To describe the electroweak transition, the action for the gauge fields and quarks and leptons is supplemented by the action for the Higgs field,

$$S_{Higgs} = \int d^4x \left(\frac{1}{4}\lambda(\phi^2 - \phi_0^2)^2 + (c^2\nabla\phi)^2 - (\partial_t\phi)^2 \right), \quad (15)$$

and the terms describing the interaction of the Higgs field with fermions and gauge fields. Here, the Higgs field ϕ is a weak doublet. The Hamilton function for the Higgs field is:

$$H_{Higgs} = E_0 + \int d^3x \left(\frac{1}{4}\lambda(\phi^2 - \phi_0^2)^2 + (c^2\nabla\phi)^2 + (\partial_t\phi)^2 \right). \quad (16)$$

The Higgs field is massive with the mass

$$m_H^2 = \lambda\phi_0^2. \quad (17)$$

From the condensed matter point of view, the mass m_H^2 and the vacuum energy E_0 are the final classical output of the whole underlying quantum vacuum. We know that the energy of the whole vacuum must be zero according to the thermodynamic stability of the free vacuum: $E_0 = H(\phi = \phi_0) = 0$. If one does not take into account the possible infrared anomalies, Eq. (16) with $E_0 = 0$ is the general form satisfying the stability condition at $T = 0$ in the absence of the external environment. In the effective theory, the information from the underlying physics is thus lost, and the type of the effective theory only depends on the symmetry and topology of the system.

Two parameters of the effective theory—λ and the equilibrium value ϕ_0 of the Higgs field—can be considered as phenomenological. Though these parameters are determined by microscopic (Planck) physics, their values may essentially differ from the microscopic scales. Examples are provided by superconductivity of metals and superfluidity of Fermi liquids, whose energy scale is exponentially small compared to the corresponding microscopic energy scale E_F (the Fermi energy): $\lambda\phi_0^2 \sim E_F^2 \exp(-1/g)$. Since the effective coupling g is typically small, $g \ll 1$, this leads to the macroscopic energy and length scales for the effective Ginzburg–Landau theory of superconductivity. In Rhodium metal, which has the lowest transition temperature observed in metals, $T_c \sim 0.3$ mK [11], one has $\lambda\phi_0^2 \sim 10^{-14} E_F^2$. In the fermionic superfluid liquid ^{3}He, the coupling g is small due to the many-body effects and one has $\lambda\phi_0^2 \sim 10^{-6} E_F^2$; the superfluidity of the ^{3}He atoms in a dilute ^{3}He–^{4}He mixture has not yet been found, which suggests that, in this system, $\lambda\phi_0^2 < 10^{-8} E_F^2$.

Let us see what happens if we take into account the zero-point energy of quantum fields, including the zero-point energy of the Higgs field ϕ. In condensed mater, this means that we quantize the system again; however, now, instead of the whole system of atoms, we are only dealing with the collective bosonic and fermionic fields that enter the effective theory.

Introducing the quantum operators for the Higgs field

$$\hat{\phi} = \phi_0 + \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} (\hat{\phi}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \text{c.c.}), \quad (18)$$

one obtains the Hamiltonian for the Higgs bosons:

$$\hat{H}_{\rm Higgs} = E_0 + \frac{1}{2}\sum_{\mathbf{k}} \hbar\omega_{\mathbf{k}}(a_{\mathbf{k}}a^{\dagger}_{\mathbf{k}} + a^{\dagger}_{\mathbf{k}}a_{\mathbf{k}}). \qquad (19)$$

The vacuum energy now contains the zero-point energy of the ϕ field, and, for completeness, one must also add the zero-point energy of other bosonic fields of the Standard Model—the gauge fields—and the negative energy of the Dirac vacuum of fermions (quarks and leptons):

$$E_{\rm zp} = \frac{1}{2}\sum_{\mathbf{k}b} \hbar\omega_{\mathbf{k}b} - \sum_{\mathbf{k}f} \hbar\omega_{\mathbf{k}f}. \qquad (20)$$

However, from the condensed matter experience, it follows that we must forget about the zero-point energy of the effective fields. They have already been included into the vacuum energy E_0, together with the energies of microscopic degrees of freedom. Moreover, the total vacuum energy must be zero in equilibrium, which means that the trans-Planckian physics fully compensates for the zero-point energy contribution of the sub-Planckian modes without any fine tuning.

Let us suppose that we are not aware of this thermodynamic principle of perfect compensation. Then, we must seriously consider the zero-point energy of bosonic and fermionic quantum fields (massless and massive) and estimate its contribution to the radiative correction to the Higgs mass. Due to the interaction of the Higgs field with the gauge bosons and fermions, all the fermions and also the *W*- and *Z*-bosons become massive:

$$\omega^2_{\mathbf{k}f} = m_f^2 + c^2k^2, \quad \omega^2_{\mathbf{k}b} = m_b^2 + c^2k^2, \qquad (21)$$

and the masses of the fermions and bosons depend on the Higgs field

$$m^2_{f,b} = \lambda_{f,b}\phi_0^2 \qquad (22)$$

(for the Higgs field itself, the corresponding $\lambda_{\rm H} = \lambda$ in Eq. (17), while the photon remains massless, $\lambda_A = 0$). This dependence leads to the ϕ_0^2 term in the zero-point energy, which must be identified with the additional mass term for the Higgs field coming from the quantum effects:

$$\delta(m^2_{\rm H}) = \frac{1}{2}\frac{d^2E_{\rm zp}}{d\phi_0^2} \approx \frac{1}{2}\sum_{\mathbf{k}b}\frac{\lambda_b}{\omega_{\mathbf{k}b}} - \sum_{\mathbf{k}f}\frac{\lambda_f}{\omega_{\mathbf{k}f}}. \qquad (23)$$

The sum in Eq. (23) diverges quadratically. Neglecting the fermion masses except for that of the heaviest fermion—the top quark with the mass m_t—one obtains for the radiative correction to the Higgs mass:

$$\delta(m^2_{\rm H}) \sim \lambda k^2_{\rm uv}\frac{m^2_{\rm H} + 2M_Z^2 + 4M_W^2 - 12m_t^2}{m^2_{\rm H}}. \qquad (24)$$

With $M_W \sim M_Z \sim m_t \sim m_{\rm H} \sim 10^2$–$10^3$ GeV, $\lambda \sim 1$ and $k_{\rm uv} \sim 10^{16}$–10^{19} GeV, the quantum correction $\delta(m^2_{\rm H})$ is many orders of magnitude bigger than $m^2_{\rm H}$ itself; this is the hierarchy problem in the Standard Model (see review paper [5]). From the condensed-matter point of view, the origin of this huge discrepancy is the same as in the cosmological constant problem: it is the diverging zero-point energy of quantum fields. The general form of the zero point energy in Eq. (20) is

$$E_{\rm zp} = a_4k^4_{\rm uv} + a_2k^2_{\rm uv}m^2_{\rm H} + a_0m^4_{\rm H}\ln(k^2_{\rm uv}/m^2_{\rm H}), \qquad (25)$$

where $|a_4|$, $|a_2|$, and $|a_0|$ are of order unity. The main contribution of zero-point energy to the vacuum energy and cosmological constant diverges quartically, while its contribution to the Higgs mass diverges quadratically. The term with a_4 gives too large a value for the vacuum energy, which in systems with gravity transforms to the large cosmological constant $\Lambda \sim k^4_{\rm uv}$ leading to the main cosmological constant problem [3, 4]. The term with a_2 leads to the hierarchy problem in the Standard Model, since it gives $\delta(m^2_{\rm H}) = (1/2)d^2E_{\rm zp}/d\phi_0^2 \sim k^2_{\rm uv}$, where we used Eq. (17). Both problems come from the same estimation of the vacuum energy as the zero point energy.

The exact calculation and also the thermodynamic analysis demonstrate that, in equilibrium, the total vacuum energy is zero, $E_{\rm vac} = 0$, if the system is isolated from the environment. The zero-point energy of quantum fields, which is only a part of the whole energy of the quantum vacuum, is thus fully compensated for by the microscopic degrees of freedom. This suggests the solution to the cosmological constant problem, and it also makes doubtful the use of the zero-point energy for the estimation of the mass of the Higgs boson, since the quadratically diverging term in the vacuum energy is also absorbed by microscopic physics to nullify the energy density of the equilibrium vacuum. The same thermodynamic principle that leads to the full compensation of the $k^4_{\rm uv}$ term leads to the full compensation of the $k^2_{\rm uv}$ term. Thus, the condensed matter suggests that the thermodynamic principle of the stability of the free vacuum provides the solution of both the cosmological constant problem and the hierarchy problem.

On the other hand, the logarithmically diverging term in Eq. (25) is within the responsibility of the effective theory, and it may play an important role when the deviations from the equilibrium are considered.

7. DISCUSSION

There are some lessons from condensed matter, and from quantum hydrodynamics in particular, for the relativistic quantum fields and gravity. Because of the power-law divergences, the quantum hydrodynamics

cannot be constructed. One can quantize the acoustic field to obtain its quanta—phonons—and use this only at the tree level. All the other diagrams are not within the responsibility of the low-energy effective theory. Similarly, the quantum gravity can be used only at the tree level. The classical energy functional (2) in hydrodynamics, as well as the classical Einstein action for gravity, and Eq. (16) for the Higgs field represent the final classical output of the whole quantum vacuum. They cannot be renormalized by the zero-point energy of the effective quantum fields, which represent only a part of all the degrees of freedom of the quantum vacuum. The quadratically and quartically divergent terms are not within the jurisdiction of the effective low-energy theory alone. They must be considered using the microscopic theory of the vacuum state, which is known in condensed matter but is not known in particle physics.

However, the condensed matter suggests that the whole quantum vacuum, which contains the zero-point motion of the effective quantum fields as well as the trans-Planckian degrees of freedom, obeys the thermodynamic laws. The latter state that the energy of the equilibrium free vacuum must be zero. This means that the huge energy of the zero-point motion of the effective fields is cancelled without any fine tuning by the trans-Planckian degrees of freedom. The exact cancellation of the quartic terms $k_{\rm uv}^4$ in the vacuum energy provides a possible solution of the cosmological constant problem. Thus, when the whole vacuum is considered, the natural value of the cosmological constant Λ is zero in the free equilibrium vacuum. In the perturbed vacuum, Λ is determined not by the quantum zero-point energy but by the classical macroscopic parameters of the Universe dynamics: the Hubble radius R of the Universe, the Newton constant G, and the energy density of the matter $\rho_{\rm mat}$.

Similar exact cancellation of the quadratic terms $k_{\rm uv}^2$, which are responsible for the radiative corrections to the Higgs mass, provides the possible solution of the hierarchy problem in the Standard Model. The superconductivity in metals and superfluidity in Fermi liquids demonstrate that the analog of the Higgs mass in these systems is typically much smaller than the relevant ultraviolet energy scale.

The main condensed-matter argument against the quantum gravity is that, in condensed matter, the effective metric field is the low energy phenomenon, which naturally emerges together with gauge fields and chiral fermions in the low-energy corner [9]. At high energy, the metric modes can no longer be separated from all the other microscopic degrees of freedom of the quantum vacuum, and, thus, pure quantum gravity cannot exist. Nevertheless, there were attempts to construct the quantum gravity in terms of only the metric field; see, e.g., [12], where the ultraviolet fixed point for quantum gravity has been derived. This makes sense under the following conditions: the ultraviolet fixed point must occur much below the real microscopic energy scale, i.e., in the region where the metric field is still well determined; in the infrared limit, the cosmological constant must be zero to match the thermodynamic requirement. One may look for a similar ultraviolet fixed point in quantum hydrodynamics; i.e., one can try to construct such a liquid in which the fixed point occurs at the intermediate length scale much bigger than the interatomic distance, where the macroscopic hydrodynamic description is still valid.

The possibility of such an intermediate energy cutoff scale in the Standard Model has been discussed in [13]. It was suggested that the intermediate scale is close to the Planck scale and that the Newton constant is determined by this scale. In this scenario, the real microscopic ("atomic") energy scale, where, say, the Lorentz invariance is violated, is much bigger. This scenario has many advantages. In particular, the merging of the running coupling constants of the weak, strong, and electromagnetic fields occurs naturally and does not require the unification of these gauge fields at high energy.

This work was supported in part by the Russian Ministry of Education and Science through the "Leading Scientific Schools" Program (project no. 2338.2003.2) and by the European Science Foundation COSLAB Program.

REFERENCES

1. L. D. Landau, Zh. Éksp. Teor. Fiz. **11**, 592 (1941); J. Phys. (Moscow) **5**, 71 (1941).
2. J. Stachel, in *Black Holes, Gravitational Radiation and the Universe*, Ed. by B. R. Iyer and B. Bhawal (Kluwer Academic, Dordrecht, 1999).
3. S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
4. T. Padmanabhan, Phys. Rep. **380**, 235 (2003).
5. S. Pokorski, *Phenomenological Guide to Physics Beyond the Standard Model*, Lectures given at *NATO Advanced Study Institute and EC Summer School on String Theory: From Gauge Interactions to Cosmology, Cargése, France, 2004*; hep-ph/0502132.
6. M. Veltman, *Facts and Mysteries in Elementary Particle Physics* (World Sci., River Edge, NJ, 2003).
7. Y. Nambu, Physica D (Amsterdam) **15**, 147 (1985).
8. D. Vollhardt and P. Wölfle, *The Superfluid Phases of Helium 3* (Taylor and Francis, London, 1990), Chap. 9.
9. G. E. Volovik, *The Universe in a Helium Droplet* (Clarendon, Oxford, 2003).
10. G. E. Volovik, Ann. Phys. (Leipzig) **14**, 165 (2005); gr-qc/0405012.
11. T. A. Knuuttila, J. T. Tuoriniemi, and K. Lefmann, Phys. Rev. Lett. **85**, 2573 (2000).
12. D. F. Litim, Phys. Rev. Lett. **92**, 201 301 (2004).
13. F. R. Klinkhamer and G. E. Volovik, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 683 (2005) [JETP Lett. **81**, 551 (2005)].

JETP Letters, Vol. 82, No. 6, 2005, pp. 325–334. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 364–372.

Nonlinear $k_\perp$ Factorization: A New Paradigm for Hard QCD Processes in a Nuclear Environment¶

N. N. Nikolaev[a, b, *], W. Schäfer[a], B. G. Zakharov[b], and V. R. Zoller[c, **]

[a] *Institut für Kernphysik, Forschungszentrum Jülich, D-52425 Jülich, Germany*

[b] *Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow region, 117334 Russia*

[c] *Institute for Theoretical and Experimental Physics, Russian Academy of Sciences, Moscow, 117218 Russia*

e-mail: zoller@itep.ru

Received June 2, 2005

A large thickness of the heavy nuclei brings a new hard scale into the pQCD description of hard processes in a nuclear environment. This new scale breaks the conventional linear $k_\perp$ factorization which must be replaced by a new concept of the nonlinear nuclear $k_\perp$ factorization. Here, we review the recent progress in the theory of nonlinear $k_\perp$ factorization. Our focus is on the role of diffractive interactions, the variation of the pattern of $k_\perp$ factorization for single-jet processes from deep inelastic scattering to hadron–nucleus collisions and universality classes for dijet production off nuclei. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 11.80.La, 12.38.Bx, 13.85.–t, 13.97.–a

1. INTRODUCTION

Linear $k_\perp$ factorization is a fundamental ingredient of the pQCD description of high-energy hard processes off free nucleons. A large thickness of a target nucleus introduces a new scale—the so-called saturation scale Q_A^2, which breaks the linear $k_\perp$ factorization theorems for hard scattering in a nuclear environment. In this paper, we review the recent work by the ITEP–Jülich–Landau Collaboration, in which a new concept of nonlinear $k_\perp$ factorization has been introduced [1]. We illustrate this new concept using several examples of single-jet to dijet production in deep inelastic scattering (DIS) off heavy nuclei and proton (deuteron)–nucleus collisions studied at the Relativistic Heavy Ion Collider (RHIC). The nonlinear $k_\perp$ factorization emerges as a generic feature of the pQCD approach to hard processes in a nuclear environment [2–5]. The concrete realizations depend strongly on the relevant pQCD subprocesses, and we define the universality classes of nonlinear $k_\perp$ factorization for production of hard dijets. Our approach is based on the equivalence between the parton fusion description of the shadowing introduced in 1975 [6] and the unitarization on the color dipole–nucleus interaction [7]. The major technical problem in the unitarization program is the non-Abelian evolution of color dipoles in a nuclear environment, and we present a closed-form solution based on the multiple-scattering theory [1, 8, 9].

2. THE $k_\perp$ FACTORIZATION FOR DIS OFF FREE NUCLEONS

Our starting point is the color dipole factorization for DIS at small x:

$$\sigma_T(x, Q^2) = \langle\gamma^*|\sigma(x, \mathbf{r})|\gamma^*\rangle = \int_0^1 dz \int d^2\mathbf{r}\Psi^*_{\gamma^*}(z, \mathbf{r})\sigma(x, \mathbf{r})\Psi_{\gamma^*}(z, \mathbf{r}). \quad (1)$$

Here, z and $(1 - z)$ is the energy partition between q & $\bar{q}$, and $\mathbf{r}$ is the size of the color dipole. There is an exact equivalence between the color dipole and the $k_\perp$ factorization [7, 10, 11]:

$$\sigma(x, \mathbf{r}) = \alpha_S(r)\int \frac{d^2\boldsymbol{\kappa} 4\pi[1 - \exp(i\boldsymbol{\kappa}\mathbf{r})]}{N_c\kappa^4}\frac{\partial G_N}{\partial \log\kappa^2}, \quad (2)$$

$$f(x, \boldsymbol{\kappa}) = \frac{4\pi}{N_c\sigma_0(x)}\frac{1}{\kappa^4}\frac{\partial G_N(x, \boldsymbol{\kappa})}{\partial \log\kappa^2}, \quad (3)$$

where $\sigma_0(x) = \sigma(x, \mathbf{r})|_{r\to\infty}$. The x dependence of $\sigma(x, \mathbf{r})$ is governed by the color dipole BFKL equation [12]. The unintegrated gluon density $f(x, \boldsymbol{\kappa})$ furnishes a universal description of $F_{2p}(x, Q^2)$ and of the final states. For instance, the linear $k_\perp$ factorization for a forward dijet cross section reads [13]

$$\frac{(2\pi)^2 d\sigma_N}{dz d^2\mathbf{p}_+ d^2\boldsymbol{\Delta}} = \frac{\alpha_S(\mathbf{p}^2)}{2} f(x, \boldsymbol{\Delta}) \times \left|\Psi(z, \mathbf{p}_+) - \Psi(z, \mathbf{p}_+ - \boldsymbol{\Delta})\right|^2, \quad (4)$$

¶ The text was submitted by the authors in English.

* A member of the editorial board of the journal *JETP Letters* from 1991 to 1993.

** A member of the editorial board of the journal *JETP Letters* since 1994.

0021-3640/05/8206-0325$26.00 © 2005 Pleiades Publishing, Inc.

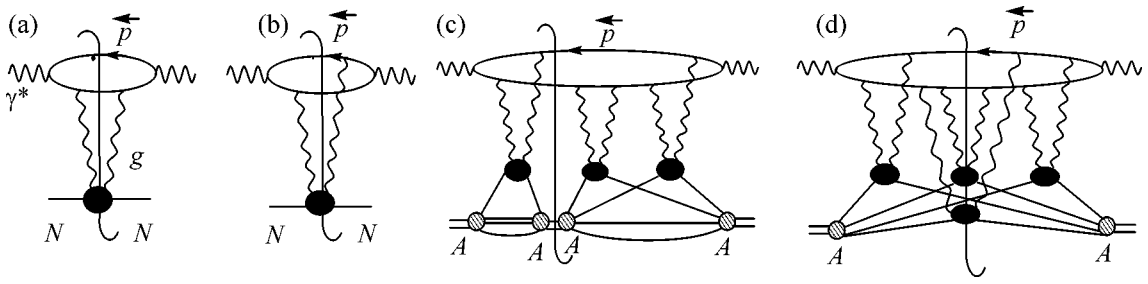


Fig. 1. The typical unitarity cuts and dijet final states in DIS: (a, b) free-nucleon target, (c) coherent diffractive DIS off a nucleus, (d) truly inelastic DIS with multiple color excitation of the nucleus.

where $\Psi(z, \mathbf{p})$ is the $q\bar{q}$ wave function of the photon and $\mathbf{\Delta} = \mathbf{p}_+ + \mathbf{p}_-$ is the jet–jet decorrelation momentum.

From the unitarity point of view, Eq. (4) describes the unitarity cuts of diagrams (Figs. 1a, 1b) for the forward Compton scattering amplitude. The point that the acoplanarity momentum $\mathbf{\Delta}$ comes from the transverse momentum of the exchanged gluon is obvious.

3. COLLECTIVE UNINTEGRATED NUCLEAR GLUE

DIS off a nucleus at $x \lesssim x_A$ can be described in terms of the color dipole–nucleus cross section [7]:

$$\sigma_A(\mathbf{r}) = 2\int d^2\mathbf{b}\left\{1 - \exp\left[-\frac{1}{2}\sigma(\mathbf{r})T(\mathbf{b})\right]\right\}, \qquad (5)$$

where $T(\mathbf{b}) = \int dr_z\, n_A(r_z, \mathbf{b})$ is the optical thickness of a nucleus at the impact parameter $\mathbf{b}$ and $n_A(r_z, \mathbf{b})$ is the nuclear matter density. This dipole cross section sums the compact form of the Glauber–Gribov multiple-scattering diagrams and is a basis for the quantitative description of nuclear shadowing in DIS [14]. On the other hand, nuclear shadowing can be understood in terms of the fusion of parton fields of nucleons, which spatially overlap in ultrarelativistic nuclei [6], and one needs to quantify the idea of a fusion of partons.

Taking Eq. (4) for guidance, we would like to define the collective unintegrated nuclear glue in terms of the final-state observables. This requires an understanding of the unitarity properties of the dipole–nucleus cross section (5). In Fig. 1, we show the two typical unitarity cuts for a nuclear target: the diagram of Fig. 1c corresponds to the so-called coherent diffractive final state in which the target nucleus remains in the ground state. It is remarkable that, although the deposition of a dozen MeV energy will break any heavy nucleus, at $x \lesssim x_A$, such a coherent diffraction makes up ≈50% of the total cross section of DIS off a heavy nucleus [15]. To the lowest order in pQCD, the coherent diffractive final state consists of a back-to-back dijet with vanishing transverse momentum transfer to the target nucleus, and the large transverse momentum of the dijets comes entirely from gluons exchanged with the target nucleons [16]. Consequently, one can take the partial wave of the diffraction amplitude, i.e., the nuclear profile function, for the definition of the collective nuclear glue per unit area in the impact parameter space, $\phi(\mathbf{b}, x, \mathbf{\kappa})$ [1, 6]:

$$\begin{aligned}\Gamma_A(\mathbf{b}, \mathbf{r}) &= \left[1 - \exp\left(-\frac{1}{2}\sigma(\mathbf{r})T(\mathbf{b})\right)\right] \\ &= \int d^2\mathbf{\kappa}\phi(\mathbf{b}, x, \mathbf{\kappa})\{1 - \exp[i\mathbf{\kappa}\mathbf{r}]\}.\end{aligned} \qquad (6)$$

A useful expansion is

$$\begin{aligned}\phi(\mathbf{b}, x, \mathbf{\kappa}) &= \sum_{j=1}^{\infty} w_j(\mathbf{b}) f^{(j)}(x, \mathbf{\kappa}), \\ w_j(\mathbf{b}) &= \frac{1}{j!}\left[\frac{1}{2}\sigma_0(x)T(\mathbf{b})\right]^j \exp[-\nu_A(x, \mathbf{b})],\end{aligned} \qquad (7)$$

where $\nu_A(x, \mathbf{b}) = \frac{1}{2}\sigma_0(x)T(\mathbf{b})$, w_j is the probability to find j overlapping nucleons at the impact parameter $\mathbf{b}$ in a Lorentz-contracted nucleus, and $f^{(j)}(x, \mathbf{\kappa})$ is a collective glue of j overlapping nucleons:

$$\begin{aligned}f^{(j)}(x, \mathbf{\kappa}) &= \int\prod_{i=1}^{j} d^2\mathbf{\kappa}_i f(x, \mathbf{\kappa}_i)\delta\left(\mathbf{\kappa} - \sum_{i=1}^{j}\mathbf{\kappa}_i\right), \\ f^{(0)}(x, \mathbf{\kappa}) &= \delta(\mathbf{\kappa}).\end{aligned} \qquad (8)$$

The plateau at small momenta of gluons,

$$\begin{aligned}\phi(\mathbf{b}, x, \mathbf{\kappa}) &\approx \frac{1}{\pi}\frac{Q_A^2(\mathbf{b})}{(\mathbf{\kappa}^2 + Q_A^2(\mathbf{b}))^2}, \\ Q_A^2(\mathbf{b}, x) &\approx \frac{4\pi^2}{N_c}\alpha_S(Q_A^2)G(x, Q_A^2)T(\mathbf{b}),\end{aligned} \qquad (9)$$

is a signal of the saturation effect.

The above-defined collective nuclear glue $\phi(\mathbf{b}, x, \boldsymbol{\kappa})$ gives precisely the same description of the amplitude of the diffraction off a nucleus as $f(\boldsymbol{\kappa})$ does for the free-nucleon target [16–18]. However, the diffraction cross section that makes up ≈50% of the DIS off nucleus is a quadratic, nonlinear functional of the collective nuclear glue. Specifically, the diffractive single-jet cross section reads

$$\frac{d\sigma_D}{d^2\mathbf{b}d^2\mathbf{p}dz} = \frac{1}{(2\pi)^2} \times \left|\int d^2\boldsymbol{\kappa}\phi(\mathbf{b}, x, \boldsymbol{\kappa})[\Psi(z, \mathbf{p}) - \Psi(z, \mathbf{p}-\boldsymbol{\kappa})]\right|^2. \quad (10)$$

4. THE NON-ABELIAN INTRANUCLEAR EVOLUTION OF COLOR DIPOLES AND TRULY INELASTIC DIS

One must then study the factorization properties of truly inelastic DIS of nuclei, which leaves several target nucleons in the color-excited state (see Fig. 1d). The salient feature of such processes is a non-Abelian intranuclear evolution of color dipoles [1, 8, 9]. We start directly with the dijet spectrum. The *ab initio* calculation of the nuclear distortion of the two-parton density matrix, the Fourier transform of which gives the spectrum of dijets, can be reduced upon the closure over the nuclear excitations to the problem of intranuclear propagation of the color-singlet 4-parton states as illustrated in Fig. 2:

$$\frac{(2\pi)^4 d\sigma_{\rm in}}{dzd^2\mathbf{p}_+d^2\mathbf{p}_-} = \int d^2\mathbf{b}'_+d^2\mathbf{b}'_-d^2\mathbf{b}_+d^2\mathbf{b}_- \times \exp[-i\mathbf{p}_+(\mathbf{b}_+ - \mathbf{b}'_+) - i\mathbf{p}_-(\mathbf{b}_- - \mathbf{b}'_-)] \times \Psi^*(z, \mathbf{b}'_+ - \mathbf{b}'_-)\Psi(z, \mathbf{b}_+ - \mathbf{b}_-) \times \{S_{4A}(\mathbf{b}'_+, \mathbf{b}'_-, \mathbf{b}_+, \mathbf{b}_-) - S_{4A}^{(\rm Diffr)}(\mathbf{b}'_+, \mathbf{b}'_-, \mathbf{b}_+, \mathbf{b}_-)\}, \quad (11)$$

where we subtracted the diffractive contribution. To the standard dilute-gas nucleus approximation, the Glauber–Gribov theory gives

$$S_{4A}(\mathbf{b}'_+, \mathbf{b}'_-, \mathbf{b}_+, \mathbf{b}_-) = \exp\left\{-\frac{1}{2}\sigma_4(\mathbf{b}'_+, \mathbf{b}'_-, \mathbf{b}_+, \mathbf{b}_-)T(\mathbf{b})\right\}, \quad (12)$$

where $\sigma_4(\mathbf{b}'_+, \mathbf{b}'_-, \mathbf{b}_+, \mathbf{b}_-)$ is the coupled-channel operator in the space of singlet–singlet $|11\rangle$ or octet–octet

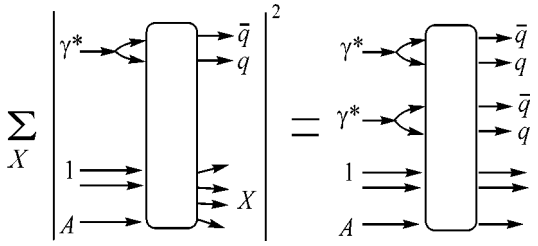


Fig. 2. Unitarity diagram for the dijet spectrum in terms of the 4-parton scattering amplitude.

$|88\rangle$ 4-body dipoles. The derivation of this operator is a major technical task (for more details, see [1]).

The integration over the transverse momenta of the antiquark jet gives the following single-jet spectrum in truly inelastic DIS [1]:

$$(2\pi)^2\frac{d\sigma_{\rm in}}{d^2\mathbf{b}d^2\mathbf{p}dz} = \int d^2\boldsymbol{\kappa}\phi(\mathbf{b}, x, \boldsymbol{\kappa})|\Psi(\mathbf{p}) - \Psi(\mathbf{p}-\boldsymbol{\kappa})|^2 - \left|\int d^2\boldsymbol{\kappa}\phi(\mathbf{b}, x, \boldsymbol{\kappa})[\Psi(\mathbf{p}) - \Psi(\mathbf{p}-\boldsymbol{\kappa})]\right|^2. \quad (13)$$

It is also the nonlinear functional of the collective nuclear glue. However, in the total inclusive single-jet spectrum, the diffractive and truly inelastic nonlinear terms exactly cancel each other, and the single-particle spectrum takes the linear $k_\perp$-factorizable form given by the integral form of Eq. (4) [19].

It is a highly nontrivial finding: the whole multitude of diffractive and truly inelastic unitarity cuts for DIS off nuclei sums up to the same unitarity cuts as shown in Figs. 1a and 1b, in which the unintegrated glue is replaced by the collective nuclear glue as defined in [1, 16]. All distortions of the transverse momentum distribution of the struck quark can exactly be reabsorbed into the collective nuclear glue, which by itself is a highly nonlinear functional of the free nucleon glue. This is not a universal feature of hard scattering off nuclei, though such a linear $k_\perp$ factorization for a single-jet spectrum in DIS is a special consequence of the incident photon being the color-singlet parton. Even in DIS, the property of linear $k_\perp$ factorization shall break for dijets.

5. THE FATE OF $k_\perp$ FACTORIZATION FOR DIJETS IN NUCLEAR DIS

The couple-channel non-Abelian evolution problem for the dijet spectrum in nuclear DIS is readily solvable to the large-N_c approximation. Here, the incident color-singlet dipole first propagates the slice $[0, \beta]$ from the front face of a nucleus, then, at some depth $0 < \beta < 1$, excites into the color-octet state, and the further non-Abelian evolution in the remaining slice $[\beta, 1]$ consists

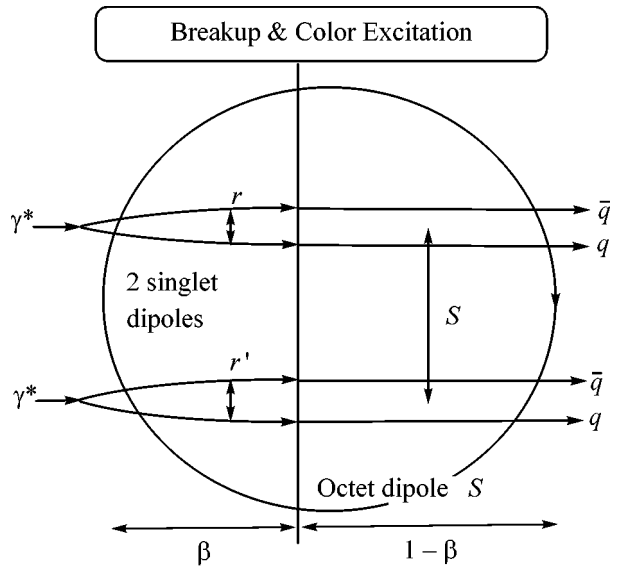


Fig. 3. The color excitation of the dipole in the large-N_c approximation.

of color rotations within the space of octet dipoles (Fig. 3).

The resulting nuclear dijet spectrum is a manifestly nonlinear functional of the collective nuclear glue and here emerges as a new concept of the nonlinear $k_\perp$ factorization:

$$\frac{(2\pi)^2 d\sigma_A(\gamma^* \longrightarrow Q\bar{Q})}{d^2\mathbf{b}dzd^2\mathbf{p}_- d^2\mathbf{\Delta}} = \frac{1}{2}T(\mathbf{b})\int_0^1 d\beta\int d^2\boldsymbol{\kappa}_1 d^2\boldsymbol{\kappa} \times f(x,\boldsymbol{\kappa})\Phi(1-\beta,\mathbf{b},x,\mathbf{\Delta}-\boldsymbol{\kappa}_1-\boldsymbol{\kappa})\Phi(1-\beta,\mathbf{b},x,\boldsymbol{\kappa}_1) \times\left|\Psi(\beta;z,\mathbf{p}_- -\boldsymbol{\kappa}_1)-\Psi(\beta;z,\mathbf{p}_- -\boldsymbol{\kappa}_1-\boldsymbol{\kappa})\right|^2 + \delta^{(2)}(\mathbf{\Delta})\left|\Psi(1;z_g,\mathbf{p}_-)-\Psi(z_g,\mathbf{p}_-)\right|^2, \quad (14)$$

where $\Phi(\beta,\mathbf{b},x,\boldsymbol{\kappa})$ is the collective nuclear glue for the slice $[0,\beta]$ of a nucleus defined by

$$\exp\left[-\frac{1}{2}\beta\sigma(x,\mathbf{r})T(\mathbf{b})\right] = \int d^2\boldsymbol{\kappa}\Phi(\beta,\mathbf{b},x,\boldsymbol{\kappa})\exp(i\boldsymbol{\kappa}\mathbf{r}) \quad (15)$$

and

$$\Psi(\beta;z,\mathbf{p}) = \int d^2\boldsymbol{\kappa}\Phi(\beta,\mathbf{b},x,\boldsymbol{\kappa})\Psi(z,\mathbf{p}+\boldsymbol{\kappa}) \quad (16)$$

is the wave function of the incident dipole distorted by the coherent initial-state interaction (ISI) in the slice $[0,\beta]$ of a nucleus. The diffractive component, $\propto\delta^{(2)}(\mathbf{\Delta})$, gives exactly back-to-back dijets (for finite-size nuclei, the $\mathbf{\Delta}$ dependence is controlled by a slightly modified nuclear form factor with the width $\mathbf{\Delta}^2 \lesssim R_A^{-2}$; see [16]). The first component in (14) describes truly inelastic DIS. Here, the slice $[\beta, 1]$, in which the dipole is in the color-octet state, gives the final-state interactions (FSI). The nuclear dijet spectrum is of fifth order in gluon field densities: a quartic functional of the collective nuclear glue for the two slices of a nucleus and a linear one of the free-nucleon glue $f(x,\boldsymbol{\kappa})$. The role of different glues is noteworthy: the latter one describes the hard singlet-to-octet transition; the former ones, the coherent ISI and incoherent FSIs. The distinct physics of ISI and FSI can require invoking the collective glue for slices of the nucleus; the $k_\perp$ factorization for the truly inelastic DIS cannot be described by the classical gluon field of the whole nucleus. The nonlinear $k_\perp$-factorization result (14) must be contrasted with the free-nucleon spectrum (4); it entails a nuclear enhancement of the decorrelation of dijets from truly inelastic DIS (the semihard dijets, $|\mathbf{p}_\pm|^2 \lesssim Q_A^2(\mathbf{b},x)$, are completely decorrelated).

6. THE MASTER FORMULA FOR NUCLEAR DIJETS

The above discussion of leading quark–antiquark dijets in DIS, $\gamma^* \longrightarrow Q\bar{Q}$, can readily be extended to quark–antiquark and quark–gluon dijets in subprocesses $g^*g \longrightarrow Q\bar{Q}$, $q^*g \longrightarrow qg$. Here, we discuss the case when the beam and final-state partons interact coherently over the whole nucleus, which at RHIC amounts to dijets in the largest rapidity bins of the proton fragmentation region, $x = (Q^2 + M_{JJ}^2)/2mE_a \lesssim x_A = 1/R_A m_p \approx 0.1A^{-1/3}$, where R_A is the radius of the target nucleus of mass number A, E_a is energy of the beam parton a in the target rest frame, and m_p is the proton mass [6, 14]. To the lowest order in pQCD, all the above processes are of the form $ag \longrightarrow bc$ and, in the laboratory frame, can be viewed as an excitation of the perturbative $|bc\rangle$ Fock state of the physical projectile $|a\rangle$ by one-gluon exchange with the target nucleon. Our focus on excitation of the lowest Fock states at $x \lesssim x_A$ is justified by the kinematical constraints in pA collisions at RHIC; understanding the very rich pattern of nonlinear $k_\perp$ factorization found in this regime is a must for the further studies of the small-x evolution of jet phenomena.

The derivation of the master formula for the dijet spectrum, based on the technique developed in [1, 8, 9], is found in [3]:

$$\frac{d\sigma(a^* \longrightarrow bc)}{dz_b d^2\mathbf{p}_b d^2\mathbf{p}_c} = \frac{1}{(2\pi)^4}\int d^2\mathbf{b}_b d^2\mathbf{b}_c d^2\mathbf{b}'_b d^2\mathbf{b}'_c \times \exp[-i\mathbf{p}_b(\mathbf{b}_b - \mathbf{b}'_b) - i\mathbf{p}_c(\mathbf{b}_c - \mathbf{b}'_c)] \times \Psi(z_b, \mathbf{b}_b - \mathbf{b}_c)\Psi^*(z_b, \mathbf{b}'_b - \mathbf{b}'_c) \times \{S^{(4)}_{\bar{b}\bar{c}cb}(\mathbf{b}'_b, \mathbf{b}'_c, \mathbf{b}_b, \mathbf{b}_c) + S^{(2)}_{\bar{a}a}(\mathbf{b}', \mathbf{b}) - S^{(3)}_{\bar{b}\bar{c}a}(\mathbf{b}, \mathbf{b}'_b, \mathbf{b}'_c) - S^{(3)}_{\bar{a}bc}(\mathbf{b}', \mathbf{b}_b, \mathbf{b}_c)\}. \quad (17)$$

It generalizes the DIS equation (11). If $\mathbf{b}_a = \mathbf{b}$ is the projectile's impact parameter, then $\mathbf{b}_b = \mathbf{b} + z_c\mathbf{r}$, $\mathbf{b}_c = \mathbf{b} - z_b\mathbf{r}$, where $z_{b,c}$ stand for the fraction of the light-cone momentum of the projectile a carried by partons b and c; $\Psi(z, \mathbf{r})$ stands for the light-cone wave function of the $|bc\rangle$ Fock state of the projectile; and its connection to the parton-splitting functions is found in [3]. All $S^{(n)}$ describe a scattering of color-singlet systems of n partons, as indicated in Fig. 4. This is the crucial point: in the course of our derivation of the dijet spectra and single-jet spectra, we only deal with infrared-safe observables. $S^{(2)}$ and $S^{(3)}$ are readily calculated in terms of the 2-parton and 3-parton dipole cross sections [7, 9, 11]; general rules for the multiple scattering theory calculation of the coupled-channel $S^{(4)}$ are found in [1, 5].

7. THE FATE OF $k_\perp$ FACTORIZATION FOR SINGLE-JET SPECTRA IN pA COLLISIONS

A short digression from dijets to single jets—the integration over the transverse momentum of the unobserved parton in the master formula (17)—is straightforward. The unobserved parton and its antiparton will enter at the same impact parameter, and multiparton color singlet states simplify to the two-parton ones. The non-Abelian evolution simplifies to the Abelian one for color-singlet dipoles made out of partons in the relevant color representations. Still, the non-Abelian features of QCD manifest themselves in the breaking of linear $k_\perp$ factorization.

As we mentioned above, a remarkable recovery of linear $k_\perp$ factorization (9) for the single-jet spectrum in DIS is rather an exception due to the abelianization in the case of a colorless projectile—the photon. The radiation of gluons from quarks, $q^* \longrightarrow qg$, illustrates nicely the salient features of breaking of linear $k_\perp$ factorization for the single-jet spectrum [3]. It is directly relevant to jet production in the proton hemisphere of pA collisions at RHIC [20, 21].

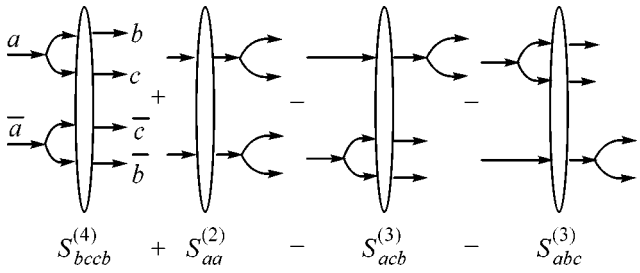


Fig. 4. The S-matrix structure of the two-body density matrix for the excitation $a \longrightarrow bc$.

Here, we again show the large N_c results. The spectrum of gluons for the free-nucleon target reads

$$\frac{(2\pi)^2 d\sigma_A(q^* \longrightarrow gq)}{dz_g d^2\mathbf{p}_g} = \frac{1}{2}\int d^2\boldsymbol{\kappa} f(x, \boldsymbol{\kappa}) \times \{|\Psi(z_g, \mathbf{p}_g) - \Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa})|^2 + |\Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa}) - \Psi(z_g, \mathbf{p}_g + z_g\boldsymbol{\kappa})|^2\}, \quad (18)$$

where $\Psi(z_g, \mathbf{p}_g)$ is the wave function of the qg Fock state of the photon; its explicit form in terms of the parton splitting functions is found in [3]. The same spectrum for the nuclear target is of the two-component form

$$\frac{(2\pi)^2 d\sigma_A(q^* \longrightarrow gq)}{dz_g d^2\mathbf{p}_g d^2\mathbf{b}} = S\left[\mathbf{b}, \frac{C_A}{C_F}\sigma_0(x)\right]\int d^2\boldsymbol{\kappa}\phi(\mathbf{b}, x, \boldsymbol{\kappa}) \times \{|\Psi(z_g, \mathbf{p}_g) - \Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa})|^2 + |\Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa}) - \Psi(z_g, \mathbf{p}_g + z_g\boldsymbol{\kappa})|^2\} + \int d^2\boldsymbol{\kappa}_1 d^2\boldsymbol{\kappa}_2\phi(\mathbf{b}, x_A, \boldsymbol{\kappa}_1)\phi(\mathbf{b}, x, \boldsymbol{\kappa}_2) \times |\Psi(z_g, \mathbf{p}_g + z_g\boldsymbol{\kappa}_1) - \Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa}_1 + \boldsymbol{\kappa}_2)|^2. \quad (19)$$

The first component is an exact counterpart of the free-nucleon spectrum: it is linear $k_\perp$ factorizable but is suppressed by the nuclear absorption factor $S[\mathbf{b}, \sigma_0(x)] = \exp\left[-\frac{1}{2}\sigma_0(x)T(\mathbf{b})\right]$. For central interactions of the main experimental interest, the gluon spectrum is entirely dominated by the second component, which is a nonlinear functional of the collective nuclear glue. This illustrates clearly a breaking of linear $k_\perp$ factorization for single jets; a full compendium of nonlinear $k_\perp$-factorization results for single jets from all possible pQCD subprocesses is found in [3].

For soft gluons, $z_g \ll 1$, and the result (19) simplifies to

$$\frac{(2\pi)^2 d\sigma_A(q^* \longrightarrow gq)}{dz_g d^2\mathbf{p}_g d^2\mathbf{b}} = \int d^2\boldsymbol{\kappa}\,\phi_{gg}(\mathbf{b}, x_A, \boldsymbol{\kappa})\left|\Psi(z_g, \mathbf{p}_g) - \Psi(z_g, \mathbf{p}_g + \boldsymbol{\kappa})\right|^2. \quad (20)$$

It takes the linear $k_\perp$-factorization form in terms of $\phi_{gg}(\mathbf{b}, x_A, \boldsymbol{\kappa}) = (\phi \otimes \phi)(\mathbf{b}, x_A, \boldsymbol{\kappa})$, which has the meaning of the collective nuclear glue defined in terms of the intranuclear propagation of the gluon–gluon color dipole. This illustrates nicely the important point that the collective nuclear glue is a density matrix in the color space rather than a single scalar function [1].

One more point is noteworthy: there is a conspicuous difference between the z_g dependence of the free-nucleon and nuclear spectra. This amounts to the $\mathbf{p}_g$ dependence of the Landau–Pomeranchuk–Migdal effect; the same applies to the spectrum of leading quarks and nuclear quenching of forward jets in *pA* collisions [3].

8. THE FATE OF $k_\perp$ FACTORIZATION FOR NUCLEAR DIJETS FROM *pA* COLLISIONS

A comparison of the $k_\perp$-factorization properties of single-jet spectra in various pQCD processes suggests a very rich pattern of nonlinear $k_\perp$ factorization. Here, we report closed-form analytic results for nuclear dijet spectra from $q^* \longrightarrow gq$ and $q^* \longrightarrow Q\bar{Q}$ subprocesses. We show the leading-order terms of the $1/N_c$ expansion; the higher-order terms can be derived following the technique of [1]. The large-N_c properties of the excitation $q^* \longrightarrow q\bar{g}$ are similar to those of the excitation $\gamma^* \longrightarrow q\bar{q}$ in DIS, while those of $g^* \longrightarrow Q\bar{Q}$ are quite different. The free-nucleon quark–gluon dijet cross section,

$$\frac{2(2\pi)^2 d\sigma_N(q^* \longrightarrow gq)}{dz_g d^2\mathbf{p}_g d^2\boldsymbol{\Delta}} = f(x, \boldsymbol{\Delta}) \times \left[\left|\Psi(z_g, \mathbf{p}_g) - \Psi(z_g, \mathbf{p}_g - \boldsymbol{\Delta})\right|^2 + \left|\Psi(z_g, \mathbf{p}_g - \boldsymbol{\Delta}) - \Psi(z_g, \mathbf{p}_g - z_g\boldsymbol{\Delta})\right|^2\right], \quad (21)$$

is simply the differential form of the single-jet spectrum (Eq. (82) in [3]).

The extension to nuclear targets is straightforward. The set of color singlet 4-parton states $qg\bar{q}'g'$ that enter the master formula (17) includes $|3\bar{3}\rangle$, $|6\bar{6}\rangle$, and $|15\overline{15}\rangle$ states. The amplitude of the excitation of the $|6\bar{6}\rangle$ and $|15\overline{15}\rangle$ states from the initial state $|3\bar{3}\rangle$ is suppressed $\propto 1/N_c$, which is compensated for in the dijet cross section by the number of color states in $|6\bar{6}\rangle$ and $|15\overline{15}\rangle$. At large N_c, one of the $|6\bar{6}\rangle \pm |15\overline{15}\rangle$ states decouples from the initial state $|3\bar{3}\rangle$ [5]. The nuclear dijet spectrum takes the form

$$\frac{(2\pi)^2 d\sigma_A(q^* \longrightarrow qg)}{d^2\mathbf{b}dzd^2\mathbf{p}_g d^2\boldsymbol{\Delta}} = \frac{1}{2}T(\mathbf{b}) \times \int_0^1 d\beta \int d^2\boldsymbol{\kappa}_1 d^2\boldsymbol{\kappa}_2 d^2\boldsymbol{\kappa} f(x, \boldsymbol{\kappa}) \times \Phi(\beta, \mathbf{b}, x, \boldsymbol{\kappa}_2)\Phi(1-\beta, \mathbf{b}, x, \boldsymbol{\Delta} - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}) \times \Phi\left(\frac{C_A}{C_F}(1-\beta), \mathbf{b}, x, \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2\right) \times \left|\Psi(\beta; z_g, \mathbf{p}_g - \boldsymbol{\kappa}_1) - \Psi(\beta; z_g, \mathbf{p}_g - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa})\right|^2 + \phi(\mathbf{b}, x, \boldsymbol{\Delta})\left|\Psi(1; z_g, \mathbf{p}_g - \boldsymbol{\Delta}) - \Psi(z_g, \mathbf{p}_g - z_g\boldsymbol{\Delta})\right|^2 + \delta^{(2)}(\boldsymbol{\Delta})S[\mathbf{b}, \sigma_0(x)]\left|\Psi(1; z_g, \mathbf{p}_g) - \Psi)(z_g, \mathbf{p}_g)\right|^2. \quad (22)$$

The contribution from the coherent diffractive excitation of color-triplet qg dipoles, $\propto\delta^{(2)}(\boldsymbol{\Delta})$, is suppressed by the nuclear attenuation because of the initial parton q^* being a colored one. The second term in (22) can be associated with excitation of the color-triplet qg states. It looks like satisfying linear $k_\perp$ factorization in terms of $\phi(\mathbf{b}, x, \boldsymbol{\Delta})$, but it does not: one of the wave functions, $\Psi(1; z_g, \mathbf{p}_g)$, is coherently distorted over the whole thickness of the nucleus (see Eq. (15)). Finally, the first component of the nuclear spectrum (22) describes excitation of the color sextet and 15-plet qg states. Notice the emergence of the ratio of Casimir operators C_A/C_F in one of the collective nuclear densities, which is still another demonstration that the collective nuclear glue is a density matrix in the color space [1]. The free-nucleon Eq. (21) is recovered to the impulse approximation.

Excitation of open charm is driven by gluon–gluon collisions. The free-nucleon dijet cross section from $g^* \longrightarrow Q\bar{Q}$ is simply the differential form of the single-jet spectrum derived in [3]:

$$\frac{2(2\pi)^2 d\sigma_N(g^* \longrightarrow Q\bar{Q})}{dzd^2\mathbf{p}_- d^2\boldsymbol{\Delta}} = f(x, \boldsymbol{\Delta}) \times \left[\left|\Psi(z, \mathbf{p}_-) - \Psi(z, \mathbf{p}_- - z\boldsymbol{\Delta})\right|^2 + \left|\Psi(z, \mathbf{p}_- - \boldsymbol{\Delta}) - \Psi(z, \mathbf{p}_- - z\boldsymbol{\Delta})\right|^2\right]. \quad (23)$$

Here, one starts with the color-octet $Q\bar{Q}$ dipole; the intranuclear interactions are color rotations in the space of the octet states, and the transitions to the color-singlet $Q\bar{Q}$ dipoles are $1/N_c$ suppressed [1]. The same suppression holds for coherent diffraction. The non-Abelian evolution of the $Q\bar{Q}Q'\bar{Q}'$ state becomes the

single-channel problem, and the resulting nuclear dijet cross section equals

$$\begin{aligned}
&\frac{(2\pi)^2 d\sigma_A(g^* \longrightarrow Q\bar{Q})}{dz d^2\mathbf{p}_- d^2\mathbf{b} d^2\boldsymbol{\Delta}} \\
&= \int d^2\boldsymbol{\kappa}\Phi(1;\mathbf{b},x,\boldsymbol{\kappa})\Phi(1;\mathbf{b},x,\boldsymbol{\Delta}-\boldsymbol{\kappa}) \\
&\times\left|\Psi(z,\mathbf{p}_- -\boldsymbol{\kappa})-\Psi(z,\mathbf{p}_- - z\boldsymbol{\Delta})\right|^2 \\
&= S[\mathbf{b},\sigma_0(x)]\phi(\mathbf{b},x,\boldsymbol{\Delta}) \\
&\times\{\left|\Psi(z,\mathbf{p}_-)-\Psi(z,\mathbf{p}_- - z\boldsymbol{\Delta})\right|^2 \\
&+\left|\Psi(z,\mathbf{p}_- -\boldsymbol{\Delta})-\Psi(z,\mathbf{p}_- - z\boldsymbol{\Delta})\right|^2\} \\
&+\int d^2\boldsymbol{\kappa}\phi(\mathbf{b},x,\boldsymbol{\kappa})\phi(\mathbf{b},x,\boldsymbol{\Delta}-\boldsymbol{\kappa}) \\
&\times\left|\Psi(z,\mathbf{p}_- -\boldsymbol{\kappa})-\Psi(z,\mathbf{p}_- - z\boldsymbol{\Delta})\right|^2.
\end{aligned} \tag{24}$$

Interestingly, the nuclear dijet spectrum (24) is precisely the differential version of the single-quark spectrum (Eq. (31) in [3]) if, in the nonlinear term, one makes the identification $\boldsymbol{\Delta} = \boldsymbol{\kappa}_1 + \boldsymbol{\kappa}_2$. It satisfies the quadratic nonlinear $k_\perp$ factorization in terms of the collective glue defined for the whole nucleus, which must be contrasted with the fifth-order nonlinearity for the leading quark–antiquark dijets in the DIS and the sixth-order nonlinearity for qg dijets from $q^* \longrightarrow qg$. Kinematically, it looks like the subprocess $g^* g_{1A} g_{2A} \longrightarrow Q\bar{Q}$ with two uncorrelated collective nuclear gluons g_A, but it cannot readily be associated with specific Feynman diagrams in terms of g_A.

Now, we demonstrate that how the diverse nonlinear $k_\perp$-factorization results for different pQCD processes fall into universality classes.

9. NONLINEAR $k_\perp$ FACTORIZATION FOR DIJETS: THE UNIVERSALITY CLASSES

9.1. Excitation of higher-color representations from partons in the lower representations. Excitation of color-octet states in DIS, as well as sextet and 15-plet states in qA interactions, belongs to this universality class. The two reactions have much similarity. In both cases, the nonlinear $k_\perp$-factorization formulas contain the free-nucleon gluon density $f(x, \boldsymbol{\kappa})$, which describes the transition from the qg color dipole from the lower (triplet for qg and singlet for DIS) to higher (sextet and 15-plet for qg and octet in DIS) color dipoles. In both cases, the number of states in higher representations is larger by a factor of N_c^2 than in the lower representation. In $q\bar{q}$ excitation in DIS, the corresponding contribution to the dijet spectrum is the fifth-order functional of gluon densities. In the qg case, it is the sixth-order functional of gluon densities. Of these, two powers of the collective nuclear glue enter implicitly via the coherent ISI distortions of the wave function $\Psi(\beta; z, \mathbf{p})$ in the slice of the nuclear matter before excitation of color dipoles in the higher representation.

The principal difference between DIS and qA interactions is in the nuclear thickness dependence of the distortion factors. Namely, the factor

$$\Phi((1-\beta),\mathbf{b},\boldsymbol{\kappa}_2)\Phi((1-\beta),\mathbf{b},\boldsymbol{\kappa}_1)$$

in DIS (Eq. (14)) is the symmetric function of the collective nuclear gluon momenta $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2 = \boldsymbol{\Delta} - \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}$, which flow from the nucleus to the quark and antiquark (or vice versa), respectively. It describes equal and uncorrelated distortion of the outgoing quark and antiquark waves by pure FSI. The independence of the two distortion factors is a feature of the large-N_c approximation.

For qg dijets in qA collisions, the overall distortion factor in (22) is of the form

$$\Phi(\beta;\mathbf{b},\boldsymbol{\kappa}_3)\Phi\left(\frac{C_A}{C_F}(1-\beta);\mathbf{b},\boldsymbol{\kappa}_2\right)\Phi(1-\beta;\mathbf{b},\boldsymbol{\kappa}_1).$$

The FSI distortions in the slice $(1-\beta)$ of the nucleus are given by the two last factors, of which $\Phi(1-\beta;\mathbf{b},\boldsymbol{\kappa}_1)$ is broadening due to final-state rescatterings of the quark, while the second FSI factor $\Phi\left(\frac{C_A}{C_F}(1-\beta);\mathbf{b},\boldsymbol{\kappa}_2\right)$ describes the FSI distortion of the outgoing gluon wave. To the large-N_c approximation, the rescatterings of the quark and gluon are uncorrelated.

The coherent ISI distortion of the wave functions in the slice $[0, \beta]$ of the nucleus in DIS and qA collisions is identical. However, in qA collisions, this coherent distortion is accompanied by incoherent ISI distortions of the incident quark wave described by $\Phi(\beta;\mathbf{b},\boldsymbol{\kappa}_3)$. In DIS, the incoherent ISI distortions are absent, because the photon is a color-singlet particle. We can anticipate that gluon–nucleus collisions with excitation of gluon–gluon dijets in higher color representations will belong to this universality class.

9.2. Excitation of final-state dipoles in exactly the same color state as the incident parton: coherent diffraction. To this universality class belong the exactly back-to-back dijets. Another experimental signature of the coherent diffraction is a retention of the target nucleus in the ground state and a large rapidity gap between the hadronic debris of the diffractive dijet and the recoil nucleus. It is most important for DIS, where coherent diffraction dissociation of the photon into $q\bar{q}$ dijets makes for heavy nuclei (≈50% of the total DIS rate [15]). The origin of the coherent diffraction is a coherent nuclear distortion of the wave function of the $q\bar{q}$ Fock state over the whole thickness of the nucleus.

In the coherent diffractive excitation of qg dipoles in qA collisions, the qg dipole must propagate in exactly the same color state as the incident quark. Here, the

nuclear suppression factor $S[\mathbf{b}, \sigma_0(x)]$ has the meaning of

$$S[\mathbf{b}, \sigma_0(x)] = \left(S\left[\mathbf{b}, \frac{1}{2}\sigma_0(x)\right]\right)^2 \quad (25)$$

and the factor $S\left[\mathbf{b}, \frac{1}{2}\sigma_0(x)\right]$ in the diffractive amplitude corresponds to the intranuclear attenuation of the quark wave with the total cross section

$$\sigma_{qN} = \frac{1}{2}\sigma_0(x). \quad (26)$$

Coherent diffractive excitation of color-octet gluon–gluon dijets in gluon–nucleus collisions is expected to exhibit similar properties.

Coherent diffractive excitation of $Q\bar{Q}$ dipoles in gA collisions is allowed, but it is suppressed at large N_c by the condition that the $Q\bar{Q}$ dipole must propagate in exactly the same color state as the incident gluon.

9.3. Incoherent excitation of final-state dipoles in the same lower color representation as the incident parton. An example of this universality class is an inelastic excitation of color-triplet qg states in qA collisions followed by a color excitation of the target. Here, both the incident parton and dijet belong to the fundamental, i.e., lower, representation of $SU(N_c)$. The intranuclear evolution of such a dipole is confined to rotations within the color-triplet state. This contribution is not suppressed at large N_c. The dijet cross section for this universality class looks like satisfying the linear $k_\perp$ factorization in terms of $\phi(\mathbf{b}, x, \mathbf{\Delta})$. But this is not the case: one of the wave functions, $\Psi(1; z, \mathbf{p}_g)$, is coherently distorted over the whole thickness of the nucleus, so that this contribution is a cubic functional of the collective nuclear glue.

We can anticipate that gluon–nucleus collisions with excitation of color-octet gluon–gluon dijets will belong to this universality class, although one has to account for the existence of the two F-coupled and D-coupled octet states.

Although superficially it looks like a subclass of this universality class, the coherent diffraction is a distinct class for the property of the exact back-to-back dijets and the rapidity gap between the dijet and the recoil nucleus in the ground state.

9.4. Excitation of final-state dipoles in the same higher color representation as the incident parton. In the realm of QCD with gluons in the adjoint representation and quarks in the fundamental representation, this universality class consists of the quark–antiquark dijets in gluon–nucleus collisions. Only in this case does the initial parton (gluon) belong to the higher (octet) color multiplet of the final $Q\bar{Q}$ state. At large N_c, the intranuclear evolution of $Q\bar{Q}$ will consist of color rotations within the space of color-octet states. The deexcitation from the color-octet to color-singlet $Q\bar{Q}$ dipoles is suppressed at large N_c. Consequently, the non-Abelian evolution of the $Q\bar{Q}Q'\bar{Q}'$ state becomes the single-channel problem. The coherent diffraction excitation, in which the initial and final color states must be identical, is likewise suppressed. The emerging pattern of quadratic nonlinearity can be related to the large N_c gluon behaving as a color-uncorrelated quark and antiquark.

The above classification exhausts reactions caused by incident photons, quarks, and gluons. However, technically, all the universality classes have a much broader basis. Indeed, instead of an incident gluon, one can think of the projectile, which is a compact lump of many partons in the highest possible color representation. For instance, in the presence of extra gluons, compact diquarks in the proton can be viewed as sextet partons.

10. SUMMARY AND OUTLOOK

Despite the manifest breaking of the linear $k_\perp$ factorization, the collective nuclear glue remains a useful concept and is an important ingredient of nonlinear $k_\perp$ factorization, which is a generic feature of the pQCD description of single-jet and dijet production in a nuclear environment. The pattern of nuclear $k_\perp$ factorization changes dramatically depending on the color properties of the specific pQCD subprocess. A unique case is a leading quark jet in DIS off nuclei: here, the collective nuclear glue (by itself, a highly nonlinear functional of the free-nucleon glue) furnishes a linear $k_\perp$-factorization description of a nuclear single-jet cross section that is an exact counterpart of the conventional linear $k_\perp$ factorization for a free-nucleon target. The pQCD bremsstrahlung of gluon jets, which carry a very small fraction of energy of the incident quark or gluon, possesses the same property in terms of the collective nuclear glue defined in terms of octet–octet color dipoles. However, in the generic case of single jets in the hadron-induced reactions, the linear $k_\perp$ factorization is badly broken.

We presented the closed-form analytic results for nuclear dijets and identified four major universality classes for the dijet cross sections. The variation of the degree of nonlinearity from one universality class to another is clearly related to the color properties of the pQCD excitation processes. However, the four universality classes found differ by more than the degree of the nonlinearity. The coherent diffractive mechanism and the excitation of quark–gluon dijets in the same color representation as the incident quark are explicitly calculable in terms of the collective nuclear glue of Eq. (5), which is defined for the whole nucleus. This is not the case for the excitation of leading quark–antiquark dijets in DIS and quark–gluon dijets in higher

color multiplets. Here, the hard excitation is described by the unintegrated gluon density in the free nucleon. The coherent initial-state interaction, before the excitation of higher color multiplets at the depth β of the nucleus, must be described in terms of the unintegrated collective glue (15) defined for the slice $[0, \beta]$ of the nucleus. Coherent distortions of the qg wave function are complemented by incoherent broadening of the incident quark transverse momentum distribution in the same slice of the nucleus. Likewise, the final-state interactions after the excitation of higher multiplets must be described in terms of the unintegrated collective glue defined for the slice $[\beta, 1]$ of the nucleus. This reinforces the point [1] that hard processes in a nuclear environment cannot be described in terms of a nuclear gluon density defined for the whole nucleus, as it was advocated, for instance, within the color glass condensate approach [22]. Furthermore, besides the collective nuclear glue defined for color-singlet quark–antiquark dipoles, there emerges a new nuclear gluon density that depends on the Casimir operators of higher quark–gluon color representations; i.e., the gluon field of the nucleus must be described by a density matrix in the space of the color representations.

The representation for the dijet cross section similar to our master formula (17) has been discussed recently by several authors [23–25], but these works stopped short of the solution of the coupled-channel intranuclear evolution for the 4-parton state.

The nuclear coherency condition, $x \lesssim x_A \approx 0.1A^{-1/3}$, restricts the applicability domain of the reviewed formalism to the forward part of the proton hemisphere of pA collisions at RHIC. These predictions could be tested after the detectors at RHIC II are upgraded to cover the proton fragmentation region [26]. This restriction is only technical; however, it can be lifted, and we would like to conclude with the statement that our formalism can readily be extended to the midrapidity dijets studied so far at the RHIC [20]. Specifically, in close similarity to the incoherent diffraction production of heavy quarkonia [27], one must distinguish coherency in the excitation of the dijet from the incident parton and the coherency in the intranuclear propagation of produced partons. Only the former will be broken for the midrapidity jets; hopefully, it will be tractable within the technique of light-cone evolution developed in application to the LPM effect [28]; this issue is currently being studied. Finally, one can extend the reported technique to a derivation of nonlinear $k_\perp$ factorization for correlation of properties of forward jets with excitation of the target nucleus; this analysis is in progress.

This work was partly supported by Deutsche Forschungsgemeinschaft (German Research Foundation), grant no. 436 RUS 17/101/04.

REFERENCES

1. N. N. Nikolaev, W. Schäfer, B. G. Zakharov, and V. R. Zoller, Zh. Éksp. Teor. Fiz. **124**, 491 (2003) [JETP **97**, 441 (2003)].
2. N. N. Nikolaev, W. Schäfer, B. G. Zakharov, and V. R. Zoller, Yad. Fiz. **68**, 692 (2005) [Phys. At. Nucl. **68**, 661 (2005)].
3. N. N. Nikolaev and W. Schäfer, Phys. Rev. D **71**, 014023 (2005).
4. N. N. Nikolaev, W. Schäfer, and B. G. Zakharov, hep-ph/0502018.
5. N. N. Nikolaev, W. Schäfer, B. G. Zakharov, and V. R. Zoller, Phys. Rev. D **72**, 034033 (2005).
6. N. N. Nikolaev and V. I. Zakharov, Yad. Fiz. **21**, 434 (1975) [Sov. J. Nucl. Phys. **21**, 227 (1975)]; Phys. Lett. B **55B**, 397 (1975).
7. N. N. Nikolaev and B. G. Zakharov, Z. Phys. C **49**, 607 (1991).
8. B. G. Zakharov, Yad. Fiz. **46**, 148 (1987) [Sov. J. Nucl. Phys. **46**, 92 (1987)].
9. N. N. Nikolaev, G. Piller, and B. G. Zakharov, Zh. Éksp. Teor. Fiz. **108**, 1554 (1995) [JETP **81**, 851 (1995)]; Z. Phys. A **354**, 99 (1996).
10. N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **332**, 184 (1994).
11. N. N. Nikolaev and B. G. Zakharov, Zh. Éksp. Teor. Fiz. **105**, 1117 (1994) [JETP **78**, 598 (1994)]; Z. Phys. C **64**, 631 (1994).
12. N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, JETP Lett. **59**, 6 (1994).
13. A. Szczurek, N. N. Nikolaev, W. Schäfer, and J. Speth, Phys. Lett. B **500**, 254 (2001).
14. V. Barone, M. Genovese, N. N. Nikolaev, *et al.*, Z. Phys. C **58**, 541 (1993).
15. N. N. Nikolaev, B. G. Zakharov, and V. R. Zoller, Z. Phys. A **351**, 435 (1995).
16. N. N. Nikolaev, W. Schäfer, and G. Schwiete, Phys. Rev. D **63**, 014 020 (2001); Pis'ma Zh. Éksp. Teor. Fiz. **72**, 583 (2000) [JETP Lett. **72**, 405 (2000)].
17. N. N. Nikolaev and B. G. Zakharov, Z. Phys. C **53**, 331 (1992).
18. N. N. Nikolaev and B. G. Zakharov, Phys. Lett. B **332**, 177 (1994).
19. N. N. Nikolaev, W. Schäfer, B. G. Zakharov, and V. R. Zoller, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 231 (2002) [JETP Lett. **76**, 195 (2002)].
20. C. Adler *et al.* (STAR Collab.), Phys. Rev. Lett. **90**, 082302 (2003).
21. I. Arsene *et al.* (BRAHMS Collab.), Phys. Rev. Lett. **93**, 242303 (2004).

22. L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 2233 (1994); J. Jalilian-Marian, A. Kovner, L. D. McLerran, and H. Weigert, Phys. Rev. D **55**, 5414 (1997); A. H. Mueller, in *Proceedings of QCD Perspectives on Hot and Dense Matter, Cargese, France, 2001*, Ed. by J.-P. Blaizot and E. Iancu (Kluwer, Dordrecht, 2002); hep-ph/0111244; E. Iancu, A. Leonidov, and L. McLerran, in *Proceedings of QCD Perspectives on Hot and Dense Matter, Cargese, France, 2001*, Ed. by J.-P. Blaizot and E. Iancu (Kluwer, Dordrecht, 2002); hep-ph/0202270; E. Iancu and R. Venugopalan, in *Quark Gluon Plasma 3*, Ed. by R. C. Hwa and X. N. Wang (World Sci., Singapore, 2004); hep-ph/0303204.
23. J. P. Blaizot, F. Gelis, and R. Venugopalan, Nucl. Phys. A **743**, 57 (2004).
24. J. Jalilian-Marian and Y. V. Kovchegov, Phys. Rev. D **70**, 114017 (2004).
25. R. Venugopalan, hep-ph/0502190; H. Fujii, F. Gelis, and R. Venugopalan, hep-ph/0502204.
26. L. C. Bland *et al.*, hep-ex/0502040; P. Steinberg *et al.*, nucl-ex/0503002.
27. N. N. Nikolaev, Comments Nucl. Part. Phys. **21**, 41 (1992).
28. B. G. Zakharov, JETP Lett. **63**, 952 (1996); JETP Lett. **65**, 615 (1997); Yad. Fiz. **61**, 924 (1998) [Phys. At. Nucl. **61**, 838 (1998)]; R. Baier, D. Schiff, and B. G. Zakharov, Annu. Rev. Nucl. Part. Sci. **50**, 37 (2000).

JETP Letters, Vol. 82, No. 6, 2005, pp. 335–342. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 373–380.

Topological Quantum Mechanics for Physicists¶

A. Losev[a] and I. Polyubin[a, b, *]

[a] *Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia*
[b] *Landau Institute for Theoretical Physics, Moscow, 117334 Russia*
e-mail: polyub@itp.ac.ru
Received August 4, 2005

This text is an attempt to write an introduction and outline of main results of topological quantum mechanics for readers with a physical background. Instead of presenting rigorous mathematical formulations, we concentrate on explanation of the physical ideas that underline most of the constructions. We review here topological quantum mechanics, since it is the simplest in the diverse family of topological theories, which contains most of their common properties. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 03.65.Vf, 11.90.+t

1. INTRODUCTION

1.1. Topological theories in modern physics. Topological theories are almost 25 years old. They appeared as the main tool in the study of spontaneous supersymmetry breaking (Witten's index; see [1]). Another example of topological theories appeared in the nonperturbative treatment of quantum field theory in the language of vacuum condensates. It was observed [5] that, in the supersymmetric theories, the correlator of the chiral observables (whose condensates determine the nonperturbative behavior of the theory) does not depend on the distance between the observables and has a simple scaling behavior with respect to the dilatation of the metric (later, this was interpreted by Witten as the metric independence of the correlators in the so-called twisted theories [4]).

However, it was only the beginning. Topological theories served as world sheet theories in the so-called topological strings. These constructions appeared in the alternative description of the noncritical strings (for $c \leq 1$) and matrix models [2], but, later on, they were found as tractable subsectors inside superstrings [13]. Moreover, they turned out to be rather interesting theories themselves and finally turned out to be in the core of the most recent attempt [6] to get a deeper understanding of the quantization of the very heart of the string theory—its majesty the superstring theory!

It appears to be an insolent conjecture, but we cannot ignore that there are breathtaking indications [7] that topological theories can shed light on the main mystery of the last decade—the fundamental degrees of freedom of the M-theory, that is, the nonperturbative theory underlying the string theory.

It looks even more intriguing in the light of the recent advances regarding topological strings and the indication of the existence of the topological M-theory [8, 9], which plays the same role for topological strings as the M-theory plays for superstrings.

It would be considered a great surprise at any time (but not in the recent 30 years) that topological theories in different dimensions can also be considered as a clever way to pack together and relate to each other interesting numbers appearing in different areas of modern mathematics. From this point of view, mathematics appears as an experimental device testing the validity of the field theoretical reasoning in topological field theories. We should mention that mathematicians cannot cope effectively with the group of conjectures coming from physics that they have to prove in their conventional fashion.

1.2. Topological quantum mechanics as the simplest model. It is well known that any phenomena in theoretical physics should be studied as a function of parameters of the problem with special attention being paid to the critical values of these parameters. It is also well known that dimension is one of the most important parameters, and, by studying phenomena in the minimal dimension, we can obtain a quantitative understanding of effects that, in higher dimensions, may be described only qualitatively. Here, the most famous example is the two-dimensional field theory. Its physical significance is not only in the description of line objects such as polymers or the behavior of surfaces but also in the creation of the tractable world of nonperturbative quantum field theories; it is clear that the dimensional transmutation in the two-dimensional sigma model is a baby version of the much harder problem of confinement.

The known world of topological theories contains theories of four dimensions (the Donaldson theory or

¶ The text was submitted by the authors in English.

* A member of the editorial board of the journal *JETP Letters* since 1994.

the topological Yang–Mills theory), three dimensions (the Chern–Simons theory), and two dimensions (the topological sigma models), but most of the phenomena already appear in one dimension of space–time, i.e., in topological quantum mechanics. In this paper, we will concentrate on the dependence on what physicists call the gravitational background of the field theory; in simple terms, this is just the distances between points of insertions of operators (in higher dimensions, this includes the full geometry of the worldsheet, such as the complex structure on the two-dimensional surface).

Before we proceed with the explanation of topological quantum mechanics, we would like to clarify several points that traditionally cause confusion in the study of topological theories.

The first cause of confusion is the demand that the worldsheet of the one-dimensional field theory (quantum mechanics) has to be a smooth one-dimensional manifold such as an interval, a circle, or a straight line. Somehow, even classically, we do not study mechanics on graphs (by the way, it is an interesting problem to compute the shape of a net carrying a melon). However, all these problems are rather familiar to particle physicists, since they arise in the study of… Feynman diagrams. In particular, one may assign different Lagrangians to different links of the graph. In a mechanical analogy, this corresponds to the problem of the melon in the net with links of different tension, and such a problem looks rather artificial and useless. However, it is what particle physicists study when particles have different masses and charges.[1]

The second cause of confusion is the widely spread point of view that correlators in topological theories should be metric-independent. In reality, these are the simplest correlators, and they appeared in the earliest examples. Mathematically speaking, such correlators are constant functions on the space of gravitational backgrounds. However, as it was later realized, general correlators should be considered not as functions but rather as the differential form on the space of gravitational backgrounds. It is not hard to guess that the condition of a function being constant is replaced by the condition of a differential form being annihilated by the action of $d = t^i \frac{\partial}{\partial t^i}$ (here, t^i stands for coordinates on the space of gravitational backgrounds). Mathematicians call such differential forms closed.

The use of forms on graphs in one-dimensional theory has been well known for many years. The parameters t^i are known as Schwinger parameters, or proper times, in the one-dimensional theory on Feynman graphs. In order to get the contribution of such a graph in the target space amplitude or the correlator, we should integrate differential forms over the space of Schwinger parameters. It is amazing that, in the case of amplitudes in topological quantum mechanics, we may promote top degree forms to forms of lower degree (and even to functions) and study all these objects simultaneously. Furthermore, the prize—the tree level amplitudes—satisfy quadratic relations known in mathematics as homotopical algebra relations. In addition, relations including graphs with loops were not studied by mathematics—in this case, physicist may predict new mathematical objects and call them quantum homotopical algebra structures, since the loops come from quantum corrections.

In most examples in quantum field theory, the relations among Feynman diagrams are just reflections of some symmetries of the action, and these relations become the Ward identities. It turns out that the same happens in quantum homological algebra. There is an action, called the Batalin–Vilkovisky (BV) action, such that its exponent solves the famous master equation [10]. This may be considered as the set of symmetries of the action and leads to equations on amplitudes that may also be found as relations on contributions of particular graphs.

2. SIMPLEST PROPERTIES OF TOPOLOGICAL QUANTUM MECHANICS

Quantum mechanics, being the simplest quantum field theory, can be described purely in terms of linear algebra: the space of states V is a vector space with a hermitian scalar product, and the Hamiltonian H is just a hermitian linear operator.[2]

The symmetries of the system correspond to operators that commute with H. We will assume that the theory contains odd or fermionic symmetry. This means that the theory contains an operator of fermionic parity that physicists traditionally denote as $(-1)^F$ that squares to 1 and commutes with the following Hamiltonian:

$$(-1)^F H = H(-1)^F; \quad ((-1)^F)^2 = 1. \tag{1}$$

States that are eigenfunctions of $(-1)^F$ with eigenvalues +1 (–1) are called bosonic (fermionic) states, respectively, and the operator is called odd or fermionic if it changes the fermionic parity (turns bosonic states into fermionic and vice versa). In particular, 1 means that the Hamiltonian itself is an even operator.

Symmetry is called fermionic if its operator is odd. Such symmetry is traditionally denoted by Q, and we have

$$QH = HQ, \quad (-1)^F Q = -Q(-1)^F. \tag{2}$$

From this algebra, it is clear that Q^2 is also an even symmetry, but we demand that it just equals zero:

$$Q^2 = 0. \tag{3}$$

[1] We cannot ignore the following troubling remark: from this point of view, the proper analogue of the graph is not a smooth worldsheet of a string but rather a piecewise smooth shape of a foam!

[2] For the sake of simplicity, we will ignore the subtleties that can appear if the vector space turns out to be infinite-dimensional.

In mathematics, this property means that Q is a differential.

In what follows, we introduce the notion of a supercommutator $[,]_{\pm}$:

$$[A, B]_{\pm} = AB + (-1)^{\text{parity } A \cdot \text{ parity } B} BA. \quad (4)$$

Note that, from equation (3), it follows that any operator of the form $[Q, A]_{\pm}$ supercommutes with Q. Such operators are called exact in mathematics.

Now, we are ready to give the definition of topological quantum mechanics: topological quantum mechanics is a quantum mechanics with an odd symmetry that squares to zero such that its Hamiltonian is exact:

$$H = [Q, G]_{pm} = QG + GQ. \quad (5)$$

The operator G is called a superpartner of the Hamiltonian. This operator is not uniquely defined; one can show that one can always take it to be odd with[3] $G^2 = 0$.

The simplest example of topological quantum mechanics is the supersymmetric oscillator. Let a (a^+) be bosonic and ψ (ψ^+) be fermionic annihilation (creation) operators. The Hamiltonian of the supersymmetric oscillator and the operators Q and G are equal to

$$H = \frac{\omega}{2}(a^+ a + \psi^+ \psi); \quad Q = \psi^+ a; \quad G = \frac{\omega}{2} a^+ \psi. \quad (6)$$

It is quite remarkable that such a simple property (5) could cause such interesting consequences.

The first consequence is that

$$I(T) = \mathrm{Tr}(-1)^F \exp(-TH) \quad (7)$$

is T-independent. Therefore, $I(T) = I$, and I is called Witten's index of the system.

The easiest way to see this is to consider the spectrum of the system. It consists of doublets at nonzero values of the Hamiltonian with opposite fermionic parity, $(a^+)^k|0\rangle$ and $(a^+)^{k-1}\psi^+|0\rangle$ (here, $|0\rangle$ is the vacuum). The contribution of these doublets drops out of the expression for the index. Only vacuum contributes, and its contribution is equal to 1.

Since T could be considered as the volume of the one-dimensional worldsheet, we may say that I is independent of the metric and, therefore, depends only on the topology of the worldsheet; this explains the name of the theory.

Despite the simplicity of the reasoning above, it does not provide a conceptual explanation of the metric independence of I. The real reason for the metric independence is that the derivative of $I(T)$ with respect to the change of the metric (rescaling of T in the simplest example) is the correlator of H, and its exactness guarantees that this correlator equals zero:

$$\frac{d}{dT} I(T) = -\mathrm{Tr}(-1)^F \exp(-TH) H = -\mathrm{Tr}(-1)^F \exp(-TH)(QG + GQ) = 0. \quad (8)$$

If we move the Q in the second term, all the way around under the trace we will get the first term but with the opposite sign.

Being inspired by such success with the index, we may try to study the correlators of other operators. The one-point correlator of Φ (on the Euclidean periodic worldsheet, circle S_1 of the length T) would look as follows:

$$\langle \Phi \rangle_{S_1;\, T} = \mathrm{Tr}(-1)^F \exp(-TH)\Phi. \quad (9)$$

The attempt to prove its T independence (just like we did for an index) fails,

$$\frac{d}{dT}\langle \Phi \rangle_{S_1;\, T} = \mathrm{Tr}(-1)^F \exp(-TH)[Q, \Phi]_{\pm} G \quad (10)$$

and differs from zero for general Φ. However, Φ, which is a zero supercommutator with Q saves the game, and its correlator is again T-independent. Such Φ are called Q-closed and are also called zero-topological observables.

Reasoning in this way, one can easily show that the correlator of zero-topological observables on the circle

$$\langle \Phi_1(T_1) \ldots \Phi_k(T_k) \rangle_{S_1;\, T} = \mathrm{Tr}(-1)^F \Phi_1 \times \exp(-(T_1 - T_2)H)\Phi_2 \ldots \Phi_k \exp(-(T - T_1)H) \quad (11)$$

is independent of the distances (in time) between the observables. Really, taking a derivative with respect to T_1 replaces $\Phi_1(T_1)$ with $[H, \Phi_1(T_1)]_{\pm}$, and we get zero as above. Naively, one may conclude that (11) is T_i-independent, but this is not so. The correlator jumps when we are trying to interchange observables; therefore, it depends only on their cyclic order. This is a one-dimensional version of the famous link invariants that were interpreted by Witten as correlators of the Wilson line observables [3].

These properties of correlators on a circle can be generalized to transition amplitudes, i.e., correlators on an interval. The novelty here is boundary conditions that take the form of initial and final states:

$$\langle a|\Phi_1(T_1) \ldots \Phi_k(T_k)|b\rangle = \langle a|\Phi_1 \exp(-(T_1 - T_2)H)\Phi_2 \ldots \Phi_k|b\rangle. \quad (12)$$

Such an amplitude depends only on the order of the events happening at the times $T_1, \ldots, T_k$ if the initial and final states are annihilated by Q:

$$\langle b|Q = Q|a\rangle = 0. \quad (13)$$

Now, let us suppose that H has a nonnegative spectrum and that all vacua are annihilated not only by H but

[3] Another name for topological quantum mechanics is $N = 1$ supersymmetric quantum mechanics. The latter is mostly known as a quantum mechanics with two symmetries Q_a with $Q_1^2 = Q_2^2 = 0$. However, after substituting $Q = Q_1 + iQ_2$ and $G = Q_1 - iQ_2$, we come back to (5).

also by Q (as in the example of the supersymmetric oscillator); then, one can show that k-point correlators (12) can be rewritten in terms of one-point correlators. Note that these conditions are automatically satisfied if there is a hermitian metric on the space of states such that $G = Q^+$ (that is how it happens in supersymmetric quantum mechanics, where the operators Q_a are supposed to be real) and if the Hamiltonian has a discrete spectrum.

Really, consider the two-point correlator and tend $T_1 - T_2$ to infinity. In this limit,

$$\exp(-(T_1 - T_2)H) \longrightarrow \Pi = \sum_c |c\rangle\langle c| \tag{14}$$

where $|c\rangle$ denote the basis in the space of vacua. Therefore,

$$\begin{aligned}&\langle a|\Phi_1(T_1)\Phi_2(T_2)|b\rangle\\ &= \langle a|\Phi_1\exp(-(T_1 - T_2)H)\Phi_2|b\rangle \longrightarrow \Pi\\ &= \sum_c \langle a|\Phi_1|c\rangle\langle c|\Phi_2|b\rangle.\end{aligned} \tag{15}$$

Moreover, even one-point functions are not independent. Consider an algebra of zero observables:

$$\Phi_i\Phi_j = C_{ij}^k\Phi_k. \tag{16}$$

In the limit $(T_1 - T_2) \longrightarrow 0$, we get

$$\sum_c \langle a|\Phi_i|c\rangle\langle c|\Phi_j|b\rangle = C_{ij}^k\langle a|\Phi_k|b\rangle. \tag{17}$$

This means that the one-point vacuum transition amplitudes form a representation of the algebra with the structure constants C_{ij}^k.

Life becomes even more interesting when we include one-observables in the game. We will do it in the next section.

3. DEFORMATION THEORY AND ONE-OBSERVABLES

Not only the generic physical system but also a system of a very special type generally depends on some parameters.[4] Therefore, consider topological quantum mechanics that depends on parameters that we call t_i. In particular, consider the equation $Q(t)^2 = 0$, and let us see what happens to the first order in t. We get

$$\left[Q, \frac{\partial Q}{\partial t_i}\right]_\pm = 0. \tag{18}$$

This equation means that $\partial Q/\partial t_i$ can be considered as a zero-observable Φ_i, which we have just studied (note that the parity of t_i and Φ_i are opposite[5]).

Let us assume, for simplicity, that, when we change the parameter t_i the superpartner G does not change. Therefore, we are going to study the variation in t of the correlators of the zero-observables. In what follows, we will use the remarkable integral representation for the variation of the evolution operator:

$$\begin{aligned}\frac{\partial e^{-TH}}{\partial t_i} &= -\int_0^T dT_1 e^{-(T-T_1)H}\frac{\partial H}{\partial t_i}e^{-T_1H}\\ &= -\int_0^T dT_1 e^{-(T-T_1)H}[G, \Phi_i]_\pm e^{-T_1H}.\end{aligned} \tag{19}$$

Let us consider, for simplicity, variation of a one-point function on the circle:

$$\begin{aligned}&\frac{\partial\langle\Phi_j\rangle_{S_1;\,T}}{\partial t_i}\\ &= \int_0^T dT_1 \mathrm{Tr}(-1)^F e^{-(T-T_1)H}[G, \Phi_i]_\pm e^{-T_1H}\Phi_j\\ &= \int_0^T \langle\Phi_i^{(1)}(T_1)\Phi_j(0)\rangle_{S_1;\,T}\end{aligned} \tag{20}$$

where we have introduced the one-observable $\Phi_i^{(1)}$ that is a one-differential on the worldsheet and that, in Heisenberg representation, takes the form

$$\Phi_i^{(1)}(T_1) = [G, \Phi_i]_\pm(T_1)dT_1. \tag{21}$$

Originally, the notion of one-observables (and higher observables in multidimensional theories) was introduced by Witten based on the following considerations. He tried to construct a nonlocal observable ω that would be integrated under the correlator and that would preserve the symmetry Q [4]. His idea was that if there is such an observable Φ that solves what he called the topological descend equation

$$[Q, \omega(T_1)] = dT_1\frac{\partial}{\partial T_1}\Phi(T_1), \tag{22}$$

then, under a correlator,

$$\int[Q, \omega(T_1)] = \int dT_1\frac{\partial}{\partial T_1}\Phi(T_1) = 0, \tag{23}$$

[4] Only perfect bodies look rigid, but even they have parameters such as an overall scale.

[5] One may be surprised to see a system with an odd parameter t and even claim that such a system is unphysical. However, the role of this parameter may be played either by some fermionic background field or by a differential form. Here, we use that odd parameters anticommute with any odd objects with odd fields, as well as with odd parameters.

since the right-hand side is the integral of the total derivative. However, this reasoning is only the first approximation to the correct construction. Really, the expression above is not zero in general since the integrand has well-known jumps due to supercommutators of Φ with other operators involved. Taking these supercommutators into account leads to the modified statement

$$\begin{aligned}&\langle [Q, \omega(T_1)]\Phi_1 \dots \Phi_k\rangle \\ &= \langle [\Phi, \Phi_1]_\pm \dots \Phi_k\rangle + \dots \langle \Phi_1 \dots [\Phi, \Phi_k]_\pm\rangle\end{aligned} \tag{24}$$

which is nothing but the Ward identities for the correlators $\langle \Phi_1 \dots \Phi_k\rangle_t$ in the theory with the modified symmetry $Q + t\Phi$ in the first order in t (just as we discussed above). In order to completely demystify this issue, we note that, for Q-closed Φ, the one-observable ω that solves the descent equation (22) is just given up to the sign by formula (21).

4. PROBLEMS IN DEFORMATION THEORY

4.1. Obstructions. Interpretation of correlators of one-observables as deformations of the theory is good as the first approximation; however, it has some problems.

The first problem is that deformed $Q(t) = Q + t_i\Phi_i$ generically does not square to zero at the second order in t. Really, for Q-closed Φ_i, we have

$$Q^2(t) = t_i t_j [\Phi_i, \Phi_j]_\pm. \tag{25}$$

This would mean that $Q(t)$ is not a symmetry of the deformed Hamiltonian $H(t) = Q(t)G + GQ(t)$. We may try to cure the problem by adding terms quadratic in t:

$$Q_{\mathrm{mod}(2)}(t) = Q + t_i\Phi_i + t_i t_j \Phi_{ij}. \tag{26}$$

Therefore, the modified $Q_{\mathrm{mod}(2)}$ squares to

$$Q^2_{\mathrm{mod}(2)}(t) = t_i t_j ([Q, \Phi_{ij}]_\pm + [\Phi_i, \Phi_j]_\pm). \tag{27}$$

We would like to find such Φ_{ij} that it would take the right-hand side of (27) to zero. However, it cannot always be done. The obstruction to doing this is hidden in the fact that, while $[\Phi_i, \Phi_j]_\pm$ always (anti)commutes with Q (i.e., it is closed), it is not (generically) an (anti)commutator; i.e., it is not exact. The space of closed operators' moduli (exact ones) is called the space of cohomology of operators, and the condition is that the (anti)commutator of the deforming operators has to vanish as an element of this space.

This situation is not so novel in quantum field theory. Let us recall the gauge theory in lower dimensions obtained from the gauge theory in higher dimensions by dimensionally reducing (or compacting on a circle) k dimensions. At the point with unbroken non-Abelian symmetry, we get k massless scalar fields ϕ_i in the adjoint representation. We know that massless fields are often the sign of a valley in the potential, and we may try to look here for such a valley. However, as we know, the (classical) potential is quartic:

$$V = \mathrm{Tr}([\phi_i, \phi_j])^2 \tag{28}$$

and a naive valley is lifted to higher order in the deformation ϕ. Here, the analogue of $Q(t)$ is the BRST operator in the ϕ background, and the vanishing of the potential means that the obstruction is zero.[6]

After improving $Q(t)$ at the second order, we may meet a problem in the third order in the deformation parameter. Namely, in the third order in t, we would need to solve

$$t_i t_j t_k [Q, \Phi_{ijk}]_\pm = -t_i t_j t_k [\Phi_i \Phi_{jk}]_\pm \tag{29}$$

and, once again, the obstruction to solving this equation would be the image of the right-hand side of (29) in cohomologies.

Problems of this kind are well known in mathematics; they are called Torelli-like problems. In the old days, people studied situations when an obstruction vanishes (the valleys of the potential, in physical language); now, they are studying the obstructions themselves (the shape of the potential, in physical language).

4.2. Connection. We have seen that one-observables are related to deformations of the theory. Naively, one may think that the correlator of one one-observable and some number of zero-observables should be expressed as the first derivative of the correlator with respect to the parameters of the background. However, life is not so simple. Topological zero-observables are different for different backgrounds (since they have to (anti)commute with $Q + t_i\Phi_i$, and the latter depends on t). In order to take the derivative, we need to identify them somehow for different t. This problem is well known in gauge theory, where one has to take derivatives of the charged field and one has to pick up some identification of its phase at different points. In this case, we say that we need a connection.

When we look at a correlator of one-observables itself, we find the same problems. Recall that, in performing a deformation of Q, we had to solve the equations $[Q, \Phi_{i_1 \dots i_m}] = \textit{something}$. Note that we can always modify the solution by replacing

$$\Phi_{i_1 \dots i_m} \longrightarrow \Phi_{i_1 \dots i_m} + V^i_{i_1 \dots i_m} \Phi_i. \tag{30}$$

One can check that this corresponds to the local change of coordinates

$$t^i \longrightarrow t^i + V^i_{i_1 \dots i_m} t^{i_1} \dots t^{i_m}. \tag{31}$$

Therefore, in order to study the correlators of just one observable in the case when the deformation is not obstructed, we have to pick up a special coordinate on the base of the deformation that is equivalent to taking

[6] However, we do not know how to make this analogy complete, since we do not know what should correspond to the correcting term ij in the gauge theories.

a flat torsion-free connection on the tangent space to the base of the deformation.

However, if we pick (anti)commuting Φ_i there is an obvious solution to all the deformation equations; the base of the deformation gets equipped with the distinguished coordinate system, and there is also a distinguished connection on the bundle of zero-observables.

5. COUPLING TO ONE-DIMENSIONAL TOPOLOGICAL GRAVITY AND QUADRATIC RELATIONS AMONG CORRELATORS

We have already observed the appearance of gravity in the topological theory in the form of the distance $T_i - T_{i+1}$ between the observables. However, as we will see in this section, in order to put together correlators of different observables and to see quadratic relations among them, we need to study topological gravity.

We will promote the bosonic distance T to a superdistance that is a pair (T, Ψ). In other terms, we will consider Ψ_i as dT_i, and the correlators (previously considered as functions of T_i) will become functions of T_i and dT_i; namely, they will become differential forms. The operator Q acting on the matter field will be promoted to the total operator

$$Q^{(\mathrm{tot})} = Q + Q_{\mathrm{gravity}} = Q + \Psi_i \frac{\partial}{\partial T_i} = Q + d. \quad (32)$$

The evolution operator is promoted to the superrevolution operator

$$\begin{aligned} U(T, dT) &= \exp(-\{Q^{\mathrm{tot}}, TG\}) \\ &= \exp(-TH - dTG). \end{aligned} \quad (33)$$

This presentation shows that the evolution operator depends on the supermetric in the Q-exact manner, i.e., through the Q-exact term. In particular, it means that U is Q^{tot} closed (i.e., that Q^{tot} is really an odd symmetry of the theory coupled to gravity):

$$\{Q + d, U(T, dT)\} = 0. \quad (34)$$

Therefore, the topological correlators that we have already studied are promoted to supertopological correlators. They are differential forms of indefinite degree. The zero degree part of the supertopological correlator is just a topological correlator. The condition of T-independence of the topological correlator is promoted to d-closeness of the corresponding differential form. In order to save space, we will explicitly write down the supertopological correlator for three points on a line:

$$\begin{aligned} &\langle a|\Phi_1(T_1, dT_1)\Phi_2(T_2, dT_2)\Phi_3(T_3, dT_3)|b\rangle \\ &\quad = \langle a|\Phi_1 U(T_1 - T_2, d(T_1 - T_2)) \\ &\quad \times \Phi_2 U(T_2 - T_3, d(T_2 - T_3))\Phi_3|b\rangle \end{aligned} \quad (35)$$

and

$$d\langle a|\Phi_1(T_1, dT_1)\Phi_2(T_2, dT_2)\Phi_3(T_3, dT_3)|b\rangle = 0. \quad (36)$$

It is clear that the supertopological correlator is independent of the overall shifts in T_i, so it is a function on the coset M_3 (mathematicians call such coset moduli spaces), which is nothing else than R_+^2, and it is parameterized by two positive coordinates $T_1 - T_2$ and $T_2 - T_3$. Moreover, it is clear that, on M_3, the supertopological correlator is also closed.

This allows us to write down relations on the integrals of supertopological correlators by integrating them along the boundary that we will denote as follows:

$$\begin{aligned} &A^a_{i_1, \ldots, i_m; b} \\ &= \int\limits_{R_+^{(m-1)}} \langle a|\Phi_{i_1}(T_1, dT_1)\ldots\Phi_{i_m}(T_m, dT_m)|b\rangle. \end{aligned} \quad (37)$$

Since supertopological correlators are closed, their integral along the boundary gives zero. So, it is the only thing that is needed to restrict the supertopological correlator to the boundary. There are two types of boundaries when the coordinate (say, $T_1 - T_2$) goes to zero (ultraviolet) or to $+\infty$ (infrared). On the ultraviolet boundary, the operator is just replaced by a product of operators. On the infrared boundary, the evolution operator is replaced by a projector to vacuum states. So, we get

$$A^a_{1,2;c}A^c_{3;b} + A^a_{1;c}A^c_{2,3,b} = f^i_{12}A^a_{i,3;b} + f^i_{23}A^a_{3,i;b}. \quad (38)$$

It is clear how to extend this to the case of an arbitrary number of points on the line.

After we get this remarkable quadratic equation (and its m-point generalizations, which are known in mathematics as structures of the A_∞-module of the algebra with structure constants of f on the space of vacua), we may deduce some consequences. First of all, we may introduce the generating parameters t_i and study

$$C^a_b(t) = A^a_{i_1, \ldots, i_m; b}t_{i_1}\ldots t_{i_m}. \quad (39)$$

From the quadratic equations of (38), we will see that it satisfies

$$(D + C)^2 = 0 \quad (40)$$

where

$$D = c^k_{ij}t^i t^j \frac{\partial}{\partial t^k}; \quad c^k_{ij} = f^k_{ij} - f^k_{ji} \quad (41)$$

is what physicists used to call the BRST operator for the Lie algebra with the structure constant c, and mathematicians call the Chevalley differential. Equations (40) are known in mathematics as a structure of the L_∞ module (for a Lie algebra with the structure constants c) on the space of vacua (in the physical literature, this equation was first obtained in [12]).

Note that, due to symmetrization (due to permutation of t), one can interpret C^a_b as a generating function for the correlator of many one-observables in the presence of only one zero-observable:

$$C^a_{i_1,\ldots,i_m;\,b} = \langle a|\Phi^{(0)}_m \int_R \Phi^{(1)}_{i_1} \ldots \int_R \Phi^{(1)}_{i_{m-1}}|b\rangle. \quad (42)$$

Therefore, we can conclude that the proper treatment of the correlator of higher observables as a correlator in a system coupled to gravity allows us to obtain interesting equations on them.

6. TOPOLOGICAL THEORY ON GRAPHS AND FEYNMAN DIAGRAMS IN BV THEORY

In the previous section, we developed the topological gravity approach to topological correlators. We may interpret a line with points on it as the simplest tree with two-valent vertices, one leaf and one root.

We will start our generalization with rooted trees with polyvalent vertices; these vertices have k inputs and one output and may be considered as operators from $V^{\otimes k}$ to V. The simplest example is the three-valent vertex that generates the bilinear operation. Surely, we will connect the vertices by the propagators U described above, so, in order to preserve the Q^{tot} symmetry, we have to assume that Q commutes with the vertex (considered as an operation); for the three-valent vertex, this is nothing but the Leibnitz rule.

As in the previous section, given a tree with K edges, we may construct from the vertices and propagators U a differential form on R^K_+, and this form would be closed. Integrating this form along the boundary, we will get two types of contributions. The infrared contributions would correspond to decoupling of a tree into two, and the ultraviolet, to shrinking of an edge and multiplication of the K-valent and L-valent vertices into a $(K + L - 1)$-valent vertex. Similarly, one can write down a quadratic relation and equate the contribution of the infrared and ultraviolet boundaries. This equation is known in mathematics as an equation of A_∞-morphism, and we are not going to present it here due to lack of space. We already presented an outline, so an interested reader can reconstruct either this equation or some of its particular examples.

However, why should we restrict our attention only to rooted trees? We may wish to consider any graph. However, graphs can be oriented or not oriented.[7]

For the oriented graph, we need to introduce vertices with several inputs and outputs (in mathematics, they are called PROPs). Then, we construct the differential form as above, integrate it along the boundary, and observe that we now have a new infrared component corresponding to lowering of the number of the loops in the graph by one.

For nonoriented graphs, we may introduce the bilinear pairing on the space V that is preserved by Q, G, such that it makes the K-valent vertices into K-valent cyclically symmetric tensors.

Therefore, just by considering graphs with real numbers on edges, we can construct many operations that generalize in a highly nontrivial way ∞-structures that were introduced by mathematicians.

Note that merely this would be a great success. However, physics can go further and interpret the graphs introduced above as Feynman graphs in the theory of Batalin–Vilkovisky integrals [15].

We will start with the case of rooted oriented trees. In this case, we may consider $V \oplus V^*$ [1] as a space of fields in the theory. Here is just a Hilbert space of our quantum mechanics, V^* is the dual space, and [1] is a fancy way to say that we are inverting statistics (dual to the bosonic field, which is fermionic, and vice versa). Let Z^A be coordinates in V and P_A be coordinates in V^* [1]. We will introduce an action

$$S_{BV}(P, Z) = P_A W^A(Z) \quad (43)$$

where

$$W^A(Z) = Q^A_B Z^B + t_i \Phi^A_{i,B} + F^A_{BC} Z^B Z^C + \ldots, \quad (44)$$

and Q^A_B and $\Phi^A_{i,B}$ are matrix elements of the operators Q and Φ_i.

Then, we decompose spaces V and V^* [1] into infrared and ultraviolet parts. The infrared parts would consist of vacua, and the ultraviolet parts would be the rest.

We will integrate $\exp S/\hbar$ along the submanifold of ultraviolet variables:

$$G^A_B Z^B = 0; \quad P_B G^B_A = 0.$$

The resulting induced action is $S_{\text{ind}}(P_{IR}, Z_{IR}, \hbar)$, which is a sum over the graphs, with h being a genus counting parameter.

Due to the well-known theorem of the BV integral [16]

$$\frac{\partial}{\partial_{P_{IR,A}}} \frac{\partial}{\partial_{Z^A_{IR}}} \exp(S_{\text{ind}}(P_{IR}, Z_{IR}, \hbar)/\hbar) = 0,$$

this is the physical meaning of the quantum generalizations of the infinity structures.

7. CONCLUDING REMARKS

It is clear that what we have covered here is just an introduction to the rapidly developing subject of topological theories. In particular, just by replacing intervals by cylinders, one can get many new structures such as commutativity equations and their relations to

[7] Just like trees. There are rooted trees with orientation on their edges from their leaves to their roots and abstract trees; the latter are not oriented.

WDVV equations [14]. The string-motivated generalization of graphs with numbers on their edges to the moduli space of the Riemann surface leads to the Zwiebach invariants introduced in [17].

However, this subject still contains a lot of open questions. We will mention some of them.

The Donaldson theory is still waiting for the proper treatment, including a four-dimensional version of topological gravity.

The relation between Virasoro constraints in topological strings and BV master equations is not clear at all.

It is even not clear that topological theories have no divergences.

Furthermore, the most intriguing question is whether the four-dimensional gravity theory in the Palatini formulation (which is also a topological theory) is divergent.

The work of A.L. was supported by the Russian Foundation for Basic Research (project no. 04-01-00637) and by the Council of the President of the Russian Federation for Support of Young Scientists and Leading Scientific Schools (project no. NSh-1999.2003.2). The work of I.P. was supported in part by the Russian Foundation for Basic Research (project no. 01-02-16027) and by the Council of the President of the Russian Federation for Support of Young Scientists and Leading Scientific Schools (project no. NSh-2044.2003.2).

REFERENCES

1. E. Witten, Nucl. Phys. B **188**, 513 (1981).
2. E. Witten, Nucl. Phys. B **340**, 281 (1990).
3. E. Witten, Commun. Math. Phys. **121**, 351 (1989).
4. E. Witten, Commun. Math. Phys. **117**, 353 (1988).
5. V. A. Novikov, M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **223**, 445 (1983).
6. N. Berkovits, J. High Energy Phys. **0009**, 046 (2000).
7. N. Berkovits, J. High Energy Phys. **0209**, 051 (2002).
8. R. Dijkgraaf, S. Gukov, A. Neitzke, and C. Vafa, hep-th/0411073.
9. N. Nekrasov, hep-th/0412021.
10. I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **102B**, 27 (1981); Phys. Rev. D **28**, 2567 (1983).
11. E. Witten, Commun. Math. Phys. **118**, 411 (1988).
12. V. Lysov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 855 (2002) [JETP Lett. **76**, 724 (2002)].
13. M. Bershadsky, S. Cecotti, H. Ooguri, and C. Vafa, Commun. Math. Phys. **165**, 311 (1994).
14. A. Losev and I. Polyubin, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 59 (2003) [JETP Lett. **77**, 53 (2003)].
15. M. Movshev and A. Schwarz, Nucl. Phys. B **681**, 324 (2004).
16. A. Schwarz, Commun. Math. Phys. **158**, 373 (1993).
17. A. S. Losev and S. V. Shadrin, math.QA/0506039.

JETP Letters, Vol. 82, No. 6, 2005, pp. 343–349. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 381–387.

Evidence for Fine Tuning of Fermionic Modes in Lattice Gluodynamics¶

F. V. Gubarev[1], S. M. Morozov[1], M. I. Polikarpov[1], and V. I. Zakharov[2]

[1] *Max-Planck Institut für Physik, 80805 München, Germany*

[2] *Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia*

e-mail: gubarev@itep.ru

Received August 9, 2005

We consider properties of zero and near-zero fermionic modes in lattice gluodynamics. The modes are known to be sensitive to the topology of the underlying gluonic fields in the quantum vacuum state of the gluodynamics. We find evidence that these modes are fine-tuned; that is, they exhibit sensitivity to both physical (one can say, hadronic) scale and to the ultraviolet cutoff. Namely, the density of the states is in physical units, while the localization volume of the modes tends to zero in physical units with the lattice spacing tending to zero. We briefly discuss the possible theoretical implications and also include some general, review-type remarks.
© *2005 Pleiades Publishing, Inc.*

PACS numbers: 11.15.Ha, 11.30.Rd, 12.38.Gc

1. TWO SCALES OF QCD

In this letter, we will consider properties of vacuum fluctuations within lattice formulation of Yang–Mills theories. For simplicity, we concentrate on the case of pure gluodynamics with no dynamical fermions. What is specific for the lattice formulation (see, e.g., [1]) is that it is a field theory in Euclidean space–time. The equation reads

$$S = \frac{1}{4g^2}\int d^4x G^a_{\mu\nu}(x)G^a_{\mu\nu}(x), \qquad (1)$$

where $G^a_{\mu\nu}(x)$ is the non-Abelian field strength tensor, a is the color index, and g is the coupling. We will actually consider the $SU(2)$ case, $a = 1, 2, 3$. Using action (1), one generates vacuum field configurations, $\{A^a_\mu\}$, where A^a_μ is the gauge potential and it performs further measurements on these fields.

As any renormalizable theory, quantum gluodynamics exhibits two scales, infrared and ultraviolet. Moreover, the ultraviolet cutoff is introduced explicitly, through a finite lattice spacing a. The infrared scale, Λ_{QCD}, on the other hand, emerges dynamically:

$$\Lambda^2_{QCD} \approx \frac{1}{a^2}\exp(-b_0/g^2(a)), \qquad (2)$$

where b_0 is a constant, $g(a)$ is the bare coupling constant normalized at the lattice spacing, and Λ_{QCD} characterizes the scale where the running coupling is of order unit. As $a \longrightarrow 0$, the bare coupling squared, $g^2(a)$, tends to zero as an inverse log of $(a \cdot \Lambda_{QCD})$. If one changes the lattice spacing a and modifies $g(a)$ according to the rules of the renorm group, the scale Λ_{QCD} does not depend on a.

Lattice simulations allow us to directly study the vacuum fluctuations and both the scales; Λ_{QCD} and $1/a$ get manifested in the vacuum fluctuations. In particular, the zero-point fluctuations are sensitive to the ultraviolet cutoff. One can measure them, for example, by studying the gluon condensate or the vacuum expectation value of the gluonic field strength tensor squared:

$$\langle 0|\frac{g^2}{32\pi^2}(G^a_{\mu\nu})^2|0\rangle \approx \frac{\text{const}}{a^4}\{1 + \Sigma_k a_k g^{2k}(a)\}, \qquad (3)$$

where a_k are coefficients of the perturbative series. The matrix element (3) is divergent as the fourth power of the ultraviolet cutoff. This is the well-known divergence of the density of the vacuum energy in field theory, which arises due to the zero-point fluctuations. High gluonic frequencies dominate the matrix element (3) because of the phase space associated with these fluctuations. One can say that the zero-point fluctuations represent an example of entropy-dominated fluctuations.

The situation looks absolutely different if one considers the topological charge. The density of the topological charge is given by

$$Q(x) = \frac{g^2}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}G^a_{\mu\nu}(x)G^a_{\rho\sigma}(x), \qquad (4)$$

where $\epsilon_{\mu\nu\rho\sigma}$ is the totally antisymmetric tensor. One can readily show that $Q(x)$ is a full derivative. As a result,

¶ The text was submitted by the authors in English.

all the perturbative fluctuations do not contribute to the topological charge. On the other hand, the probability of finding a nonperturbative fluctuation of size ρ is suppressed for small ρ as $\exp(-\text{const}/g^2(\rho))$, where $g(\rho)$ is the running constant. This factor grows fast with growing ρ and is somehow smoothened out at $\rho \sim \Lambda_{\rm QCD}^{-1}$, where the running of the coupling cannot be calculated reliably. As a result, topologically nontrivial fluctuations have a typical size of order $\Lambda_{\rm QCD}^{-1}$ and are absolutely negligible on the scale of a. In particular, instantons are topologically nontrivial, and all the factors mentioned above are known explicitly. Instantons represent fluctuations whose probability is determined primarily by action and not entropy (for a review of instantons in the context of the lattice measurements, see, e.g., [2]).

A common viewpoint is that confinement is due to vacuum fluctuations on the scale $\Lambda_{\rm QCD}$, i.e., that confinement is determined by soft, semiclassical fields. The expectations can be confronted with lattice measurements on the fully quantum, vacuum state of gluodynamics. Most recently, there has been emerging evidence [3] that the confining fields are actually of a third type exhibiting both infrared and ultraviolet scales. Both the action and entropy are very large, but they balance each other almost exactly. One can call them fine-tuned fluctuations (for a review, see [4]).

2. FINE TUNING

Imagine that a relativistic system has the size r_0. Then, the typical momenta should be of the order $p \sim 1/r_0$. Respectively, one expects that the typical masses are of the order $m^2 \sim 1/r_0^2$. If, on the other hand, the observed masses are much smaller than r_0^{-2}, one calls such a case fine tuning.

The notion of fine tuning has been discussed most thoroughly in connection with Higgs physics. The problem here is that to fulfill its role as a part of a renormalizable theory of weak interactions, the Higgs particle should have a mass of the order of 100 GeV, as intermediate bosons of weak interactions. On the other hand, if Higgses are pointlike particles down to the scale r_0, then radiative corrections to the Higgs mass are of the order

$$\Delta m_H^2 \sim \alpha r_0^{-2}, \tag{5}$$

where α is the electromagnetic coupling. If r_0 is, say, of the order of the inverse Planck mass, fine tuning of the radiative correction and of the bare mass is required.

To resolve the puzzle of fine tuning, if it is observed experimentally, one usually invokes hidden symmetry. In the Higgs case, for example, one of the favorite candidates for such a symmetry now is supersymmetry.

In the case of Yang–Mills theories, one usually does not expect to confront the problem of fine tuning. However, recently, it was found that the field configurations, which are responsible for the confinement, appear fine-tuned. In particular, there is ample evidence that so-called central vortices (for a review, see, e.g., [5]) are responsible for the confinement. The central vortices represent closed two-dimensional surfaces whose total area scales in physical units are the following:

$$A_{\rm tot} \approx 24 V_{\rm tot}\ \text{fm}^{-2}, \tag{6}$$

where $V_{\rm tot}$ is the total volume of the lattice. On the other hand, the non-Abelian action associated with the surfaces is ultraviolet divergent [3]:

$$S_{\rm tot} \approx 0.55 \frac{A_{\rm tot}}{a^2}. \tag{7}$$

Combining observations (6) and (7), one concludes that the suppression of the fluctuations due to the action of (7) is nearly compensated for by the enhancement due to the entropy. That is, the fluctuations are fine-tuned.

The observations above are purely empirical. Theoretically, it is natural to again speculate that there is a hidden symmetry that ensures the observed fine tuning. Moreover, the only symmetry that can come into consideration is the conjectured duality between the Yang–Mills theories and string theories [6]. However, such a connection is purely speculative at the moment [4]. We mention this possibility just to emphasize that further studies of the fine tuning in Yang–Mills theories are of great interest. In this paper, we address the probable manifestation of the fine tuning in fermionic zero modes.

3. WITTEN–VENEZIANO AND BANKS–CASHER RELATIONS

We will study the properties of low-lying modes of the Dirac operator,

$$D_\mu \gamma_\mu \psi_n = \lambda_n \psi_n, \tag{8}$$

where γ_μ are the Dirac matrices and the covariant derivative D_μ is constructed on the gauge potential A_μ^a generated as a vacuum field configuration. The Dirac equation is solved numerically. Moreover, one generates many configurations and studies the properties of the modes with low values of λ_n. Note that both λ_n and the volume occupied by the nth wave function ψ_n are gauge-invariant quantities.

There exists rich literature on the low-lying Dirac eigenmodes (LDEs). Most commonly, one uses the instanton model of the vacuum (for a review and further references, see, e.g., [2]). The instantons are localized solutions with nontrivial topological charge, and there are exact zero fermionic modes associated with them:

$$n_+ - n_- = N_f Q_{\rm top}, \tag{9}$$

where $n_\pm$ are the number of zero modes with positive or negative chirality, N_f is the number of quark flavors, and the topological charge for an instanton is $Q_{\rm top} = 1$.

In the physical vacuum, the instanton solutions are distorted. Indeed, there are neighboring instantons (or anti-instantons), and the instanton fields are modified because of that. The corresponding fermionic zero modes are becoming near-zero modes.

There are some generic features that are predicted to survive the modifications. First, consider the exact zero modes for a given lattice volume $V_{\rm tot}$. The exact zero modes are related to the total topological charge $Q_{\rm top}$. On average, $Q_{\rm top} = 0$. The value of $Q_{\rm top}^2$ fluctuates, however. The instanton-like picture presumes that there are lumps of topological charge close to $Q_{\rm top} = \pm 1$ occupying subvolumes of order $\Lambda_{\rm QCD}^{-4}$. There are no fluctuations of the topological charge on the scale of the lattice spacing a. Indeed, instantons are topological excitations with the smallest action possible, and the probability of finding instantons of the size of order a is proportional to a high power of a. Quantitatively, the strength of the topological fluctuations is related to the η' mass [7]:

$$\chi_t \equiv \frac{\langle Q_{\rm top}^2\rangle}{V_{\rm tot}} \approx \frac{m_{\eta'}^2 f_\pi}{2N_f} \approx (213\ {\rm MeV})^4, \qquad (10)$$

where the numerical value is borrowed from [15].

There are also near-zero fermionic modes whose total number is proportional to the total volume and which are associated, in the zero approximation, with the original lumps of the topological charge (see above). Since the modes are nearly degenerate, the interaction between instantons results in the delocalization of these zero modes (for details, see, e.g., [2]). Quantitatively, this generic feature of the instantonic vacuum is manifested through the Banks–Casher relation [8]:

$$\langle\bar\psi\psi\rangle = -\pi\rho(\lambda_n \longrightarrow 0), \qquad (11)$$

where $\rho(\lambda_n \longrightarrow 0)$ is the density of the (delocalized) zero modes in the limit of the infinite volume $V_{\rm tot}$. For finite lattice volumes, these modes are near-zero.

Thus, Eqs. (10) and (11) predict that the number of exact zero modes is proportional to $\sqrt{V_{\rm tot}}$, while the number of near-zero modes is proportional to the total volume, $V_{\rm tot}$. Moreover, the instantonic picture fixes the scaling laws according to which the corresponding coefficients of the proportionality are in physical units and not dependent on the lattice spacing a.

4. MEASUREMENTS

We are using the overlap Dirac operator [9]. The advantage of the overlap operator is that it preserves the chiral symmetry and allows us to study the properties of the Dirac modes from the first principles [10]. There were actually many studies of the Witten–Veneziano and Banks–Casher relations on the lattice. However, they mostly used such versions of the lattice fermions such that the topological charge has intrinsic ultraviolet noise. To avoid this, one has to modify and smoothen the original gauge fields that fluctuate on the lattice spacing a (for a review, see, e.g., [11]). As a result, no dependence on a could actually be measured. The use of the overlap fermions allows us to study the configuration of the original gauge fields (see, in particular, [12]). Study of the dependence on the lattice spacing a of various observables is one of the main objectives of the present paper.

More explicitly, the massless overlap Dirac operator is given by [9]:

$$D = \frac{\rho}{a}\left(1 + \frac{A}{\sqrt{AA^\dagger}}\right), \quad A = D_W - \frac{\rho}{a}, \qquad (12)$$

where A is the Wilson Dirac operator with a negative mass term. Antiperiodic (periodic) boundary conditions in time (space) directions were employed. It turns out that the optimal value of the parameter ρ is 1.4. Furthermore, we have used the minimax polynomial approximation [13] to compute the sign function $\mathrm{sgn}(A) = A/\sqrt{AA^\dagger} \equiv \gamma_5\mathrm{sgn}(H)$, where $H = \gamma_5 A$ is the hermitian Wilson Dirac operator. In order to improve the accuracy and performance, about one hundred of the lowest eigenmodes of H were projected out. Note that the eigenvalues of (12) lie on the circle of radius ρ centered at $(\rho, 0)$ in the complex plane. In order to relate them to the continuous eigenvalues of the Dirac operator, the circle was stereographically projected onto the imaginary axis [14].

The calculations were performed on two subsets (table) of statistically independent $SU(2)$ quenched configurations generated with standard Wilson action. For the subset A, the gauge coupling was chosen in such a way that the physical volume remains the same, and the corresponding lattice spacings were determined by interpolating the data of [15]. The subset B was used to determine the IPR volume dependence at the fixed spacing $a = 0.1394(8)$ fm.

5. RESULTS FOR TOPOLOGICAL SUSCEPTIBILITY

First, let us address the issue of the scaling of the topological susceptibility $\chi_t = \langle Q_{\rm tot}^2\rangle/V_{\rm tot}$, where the topological charge is defined by the index of the overlap Dirac operator $Q_{\rm top} = n_+ - n_-$, and $n_\pm$ is the number of exact zero modes with positive (negative) chirality. The topological susceptibility calculated on the subset A scales up to the a^2 corrections (Fig. 1) and equals $\chi_0 = (225(3)\ {\rm MeV})^4$, which is in good agreement with the conventional value $(213(3)\ {\rm MeV})^4$ [15].

Simulation parameters

β	a, fm	L_s	L_t	V^{phys}, fm^4	N_{conf}
Subset A					
2.3493	0.1397(15)	10	10	3.8(2)	300
2.3772	0.1284(15)	10	14	3.8(2)	91
2.3877	0.1242(15)	10	16	3.8(2)	198
2.4071	0.1164(15)	12	12	3.8(2)	179
2.4180	0.1120(15)	12	14	3.8(2)	149
2.4273	0.1083(15)	12	16	3.8(2)	198
2.4500	0.0996(22)	14	14	3.8(3)	200
2.5000	0.0854(4)	16	18	3.92(7)	196
Subset B					
2.3500	0.1394(8)	10	10	3.8(1)	100
2.3500	0.1394(8)	10	14	5.3(1)	100
2.3500	0.1394(8)	12	12	7.8(2)	100
2.3500	0.1394(8)	12	18	11.7(2)	100
2.3500	0.1394(8)	14	14	14.5(3)	94
2.3500	0.1394(8)	14	16	16.5(4)	42

The observed scaling of the topological susceptibility confirms that the discretization errors and the finite volume effects are small. Note that the Wilson gauge action is plagued by lattice dislocations, and one might have speculated that spurious fermionic zero modes are generated. However, Fig. 1 confirms that the dislocations are inessential for an overlap Dirac operator, which is insensitive to the ultraviolet noise.

Moreover, we have measured the density of the near-zero modes (see Fig. 2). The value of the condensate $\langle\bar{\psi}\psi\rangle$ determined using the Banks–Casher relation turns out to be close to the value obtained earlier with the Wilson fermions (see, for instance, [16]).

6. SHRINKING OF THE VOLUME OCCUPIED BY FERMIONIC MODES

A natural measure of the localization of the eigenmodes is provided by the inverse participation ratio (IPR) I_λ, which is defined as follows (for a review and applications, see, for example, [17]). Let $\rho_\lambda(x)$ be the normalized bilinear

$$\rho_\lambda(x) = \psi_\lambda^\dagger(x)\psi_\lambda(x), \quad \sum_x \rho_\lambda(x) = 1,$$

where $\psi_\lambda(x)$ is the eigenmode of the overlap Dirac operator in the given gauge field background with the virtuality λ, $D\psi_\lambda = \lambda\psi_\lambda$. Then, for any finite volume V, the IPR I_λ is defined by

$$I_\lambda = V\sum_x \rho_\lambda^2(x), \tag{13}$$

and characterizes the inverse fraction of sites contributing to the support of $\rho_\lambda(x)$. Note that, for delocalized modes, $\rho_\lambda(x) = 1/V$ and, hence, $I_\lambda = 1$, while an extremely localized mode, $\rho_\lambda(x) = \delta_{x,x_0}$, is characterized by $I_\lambda = V$. Moreover, for eigenmodes localized on the fraction f of sites (so that, $\sup\rho_\lambda = V_f = fV$), we have $I_\lambda = V/V_f = 1/f$. If we allow for a mixture of both localized and extended modes, the average value of the IPR is given by

$$\langle I_\lambda\rangle = c_0 = c_1 V/V_f. \tag{14}$$

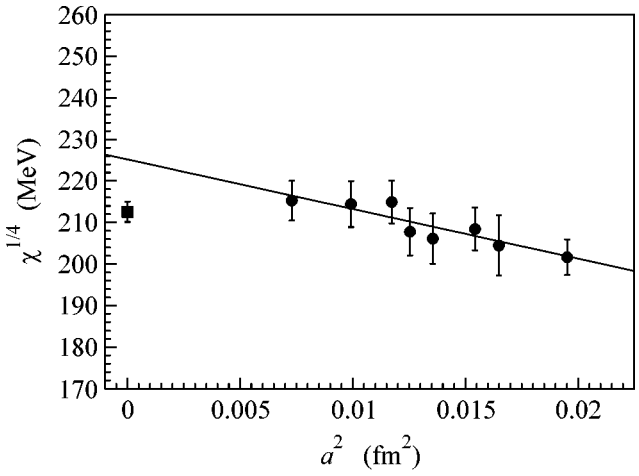


Fig. 1. Scaling of topological susceptibility with lattice spacing; the solid curve is the best fit $\chi^{1/4} = \chi_0^{1/4} + ca^2$. The square corresponds to the conventional value (213(3) MeV)^4 [15].

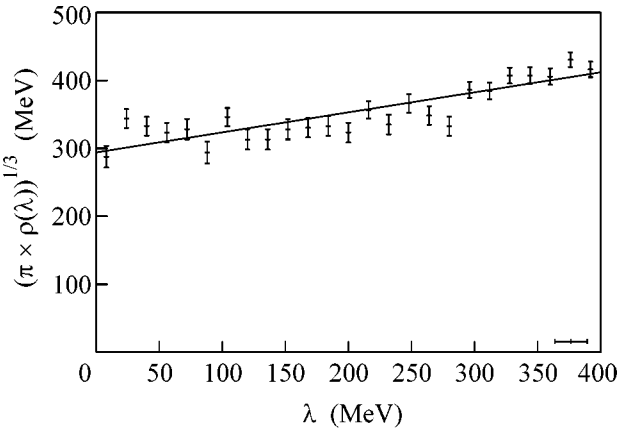


Fig. 2. The spectral density of eigenmodes of the overlap Dirac operator and its interpolation to $\lambda_n \longrightarrow 0$ at $\beta = 2.5$ on an 18×16^3 lattice. The value of the condensate $\langle\bar{\psi}\psi\rangle$ (Eq. (11)) is perfectly consistent with the results of [16].

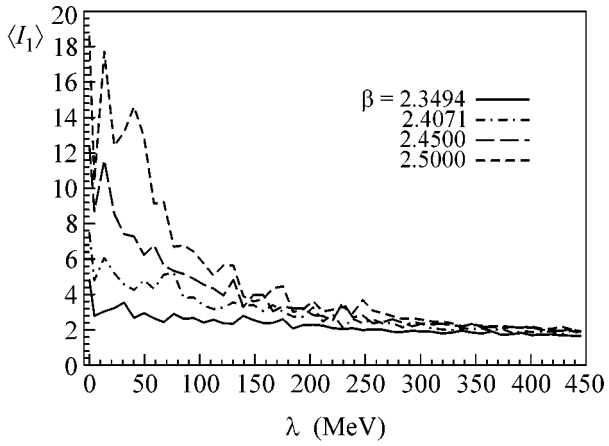


Fig. 3. IPRs for LDEs at various lattice spacings and fixed physical volume. The "mobility edge" $\lambda_{cr} \approx 200$ MeV is clearly seen.

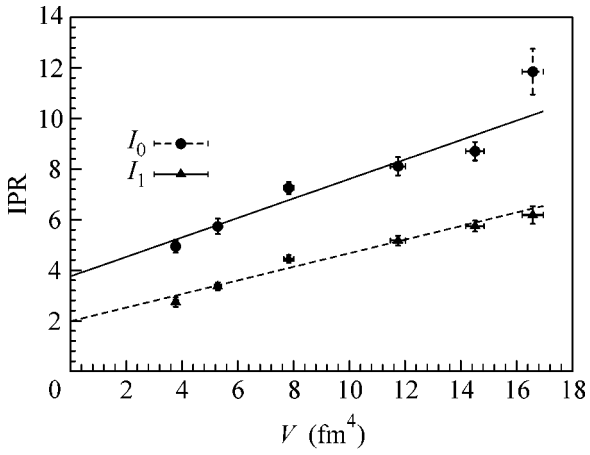


Fig. 4. Volume dependence of I_0 and I_l at fixed lattice spacing.

Usually [17], one considers localization only in terms of the total volume. Very recently, there appeared data such that the localization volume can depend on the lattice spacing as well. Namely, such an effect was observed first in [18] for fermionic modes (in a particular version of lattice fermions, so-called Asqtad fermions). Most recently, the localization properties of scalar probe particles were investigated [19], and, again, a strong dependence on a was observed. The observed dependence on the lattice spacing seems specific for the dynamics of the Yang–Mills theories.

The localization properties of the LDEs in our measurements are illustrated in Fig. 3, where we plot the inverse participation ratios for the modes $0 \le \lambda \le 450$ MeV at various spacings and fixed physical volume (subset A, table). One can see that there is a critical value, $\lambda_{cr} \approx 200$ MeV, above which all the states are, in fact, delocalized and their IPRs are lattice spacing-independent. However, for small eigenvalues below λ_{cr}, the value of the IPRs grow and one cannot exclude the presence of localized states. The mixture of localized and extended modes could be quantitatively characterized by the volume dependence of the IPR at fixed lattice spacing (see Eq. (14)). We have computed the average value $I_l = \sum_\lambda \langle I_\lambda \rangle / N_l$ of the IPRs for the modes $0 < \lambda \le \Lambda = 50$ MeV, where N_l is the total number of modes in this interval, and the analogous quantity I_0 for the exact zero modes. Fitting the data to Eq. (14), we found for the exact zero modes $c_0 = 3.7(5)$, $c_1 = 0.38(5)$ with $\chi/\mathrm{NDF} = 3$, and for the low-lying modes $c_0 = 1.9(2)$, $c_1 = 2.26(3)$ with $\chi/\mathrm{NDF} = 2$.

Moreover, one can see from Fig. 3 that the LDE's localization degree depends nontrivially on the lattice resolution. To quantitatively investigate this dependence, we have plotted I_0 and I_l against the lattice spacing in Fig. 5. It turns out that our results are highly robust with respect to the actual value of Λ as long as $\Lambda \lesssim 50$ MeV, so that we can take $\Lambda = 50$ MeV to improve the statistics. It follows from Fig. 5 that the inverse participation ratios for both zero and low-lying eigenmodes seem to be divergent in the limit $a \longrightarrow 0$. Note that the average values of the IPRs are higher than those of [18].

Furthermore, it is straightforward to estimate the dimensionality d of the objects that localize the low-lying eigenmodes. Indeed, at a fixed physical volume, we have

$$I_\lambda = V/V_f = Va^4/(V_f a^4) = a^{d-4} V^{\mathrm{phys}}/V_f^{\mathrm{phys}}$$

and, therefore, the lattice spacing dependence is given by

$$\langle I_\lambda \rangle = b_0 + b_1 a^{d-4}. \qquad (15)$$

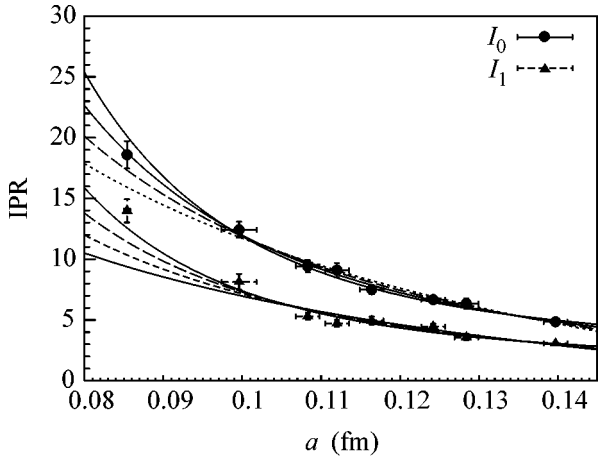


Fig. 5. Scaling of I_0 and I_l with $a \longrightarrow 0$. The lines correspond to the fitting curves (15) for $d = 0, 1, 2, 3$.

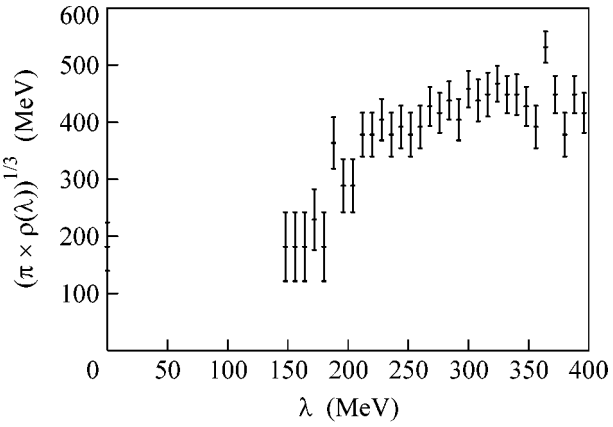


Fig. 6. The spectral density of overlap Dirac operator eigenmodes for vortex-removed configurations. The lattices used are 18×16^3 at $\beta = 2.5$.

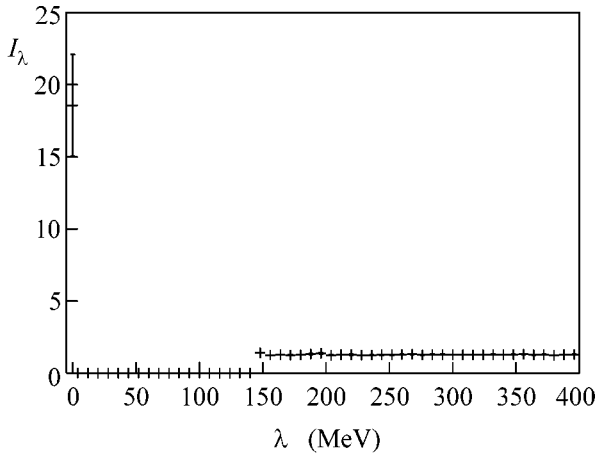


Fig. 7. IPR for LDEs for vortex-removed configurations obtained at $\beta = 2.5$ on 18×16^3 lattices.

We fitted our data to Eq. (15) for $d = 0, 1, 2, 3$; the fitting curves are shown in Fig. 5. As a matter of fact, the IPR data for low-lying modes strongly suggest that the dimensionality of the underlying objects is zero, $d = 0$; the relevant χ^2/n.d.f. is 2.3, which should be compared with its values of 3.4 ($d = 1$) and 4.9 ($d = 2$). As far as the exact zero modes are concerned, our data favor one-dimensional ($d = 1$) localization regions; χ^2/n.d.f., in this case, is 0.38, while it is 0.95 for $d = 0$ and 0.85 for $d = 2$. The error bars are not small enough, however, to rule out reliably that the exact zero modes in the limit $a \longrightarrow 0$ are also localized on pointlike objects.

7. EFFECT OF THE REMOVAL OF THE VORTICES

Thus, our results indicate fine tuning of the fermionic modes. Namely, their volume shrinks without affecting the eigenvalues. We have already mentioned (see Eqs. (6), (7)) that fine tuning was observed first for the central vortices. A natural question arises, whether the fine tuning of the fermionic modes is related to the existence of the fine-tuned vortices. A standard way to probe such a relation is the so-called removal of the vortices [20, 21]. Namely, one multiplies the original link matrices $U_\mu(x)$ by their Z_2 projected values:

$$U_\mu(x) \longrightarrow U_\mu(x)(Z_2)_\mu(x). \qquad (16)$$

One can show that the modification (16) affects (up to a gauge transformation) only a 3D fraction of the total 4D volume of the lattice [22]. In other words, this fraction is proportional to $(a \cdot \Lambda_{QCD})$ and vanishes in the limit of $a \longrightarrow 0$. On the other hand, the vortices are removed.

The impact of the change (16) on the properties of the LDEs is illustrated in Figs. 6 and 7. Namely, all the low-lying modes, with the exception of approximately 4% of the exact zero modes, disappear.[1] For the modes remaining in the spectrum, the value of the IPR is of unit order. In other words, removal of the vortices destroys the topology-related Dirac modes.

As is mentioned above, the change (16) affects the 3D subvolume of the whole lattice. Further information on the manifold crucial for the chirality breaking can be provided by observations of the shrinking of the localization volume as a function of the lattice spacing. If, indeed, the volume shrinks to points, $d = 0$, as favored by the data above, then it is only a small subspace of the 3D volume that is related to the chiral symmetry. Purely geometrically, such points could well be the points of self-intersection of the vortices. Theoretically, a possible connection between the self-intersections of the vortices and the zero modes was considered within a particular model in [23]. In principle, this conjecture based on the relation between the self-intersections of the vortices and the localization of the (near) zero modes could be investigated through direct observation. At the moment, however, such data are not available.

8. CONCLUSIONS

To summarize, we have studied the properties of low-lying Dirac modes in the vacuum state of $SU(2)$ gluodynamics using overlap fermions. We worked with the original gauge fields, which fluctuate on the scale of the lattice spacing, without any smoothening. The Witten–Veneziano and Banks–Casher relations predict scaling laws for densities for exact and near zero modes, respectively. Namely, these densities should depend only on the physical scale, Λ_{QCD}. From the point of view of the lattice measurements, these are

[1] Note that a similar effect was observed in [21], where a chirally improved lattice Dirac operator was studied.

strong constraints. The results of the measurements do agree with the theoretical expectations.

A novel feature uncovered by the lattice simulations is that the low-lying modes are localized on the volumes that shrink as a power of the lattice spacing a. For exact zero modes, the power is close to three, $V_{\lambda=0} \sim a^3$, while, for near-zero modes, it is rather four, $V_{\lambda\to 0} \sim a^4$ (for the error bars, see Fig. 5).

Moreover, the removal of the central vortices from the original configurations eliminates all the topological fermionic modes. This observation favors speculations that the lumps of the topological charge are associated with self-intersections of the vortices.

Note that, even if this picture is correct, this does not automatically imply that the instantonic picture is wrong. There could be two dual pictures of the same effect that are valid for measurements with various resolutions. In terms of soft fields, instanton-like fluctuations could dominate. In terms of fields measured with fine resolution of order a, the fine-tuned surfaces could be responsible for the lumps of the topological charge. It goes without saying that the hypothesis on the crucial role of the surfaces is to be scrutinized much further in the lattice measurements. If existence of the dual pictures is confirmed by measurements, one could speculate that this empirical duality reflects duality on the fundamental level. The vortices themselves could be identified with dual strings [4], while their self-intersections could be basic element in the mechanism of the chiral symmetry breaking in the dual picture.

One cannot rule out that the lattice-spacing dependence we found will change on finer lattices. In principle, it could happen that the $\sim a^{-4}$ dependence of the IPR shown in Fig. 5 will eventually flatten in the limit $a \longrightarrow 0$, so that the localization of the LDEs will identify some four-dimensional objects of finite physical extent. Note, however, that the characteristic size of these objects is definitely less than 0.1 fm and likely to be much smaller than that. Although this scenario is not excluded a priori and requires further investigations, it seems artificial to our mind.

We are grateful to J. Greensite, S. Olejnik, G. Schierholz, and T. Suzuki and to the members of the ITEP lattice group for stimulating discussions. The invaluable assistance of G. Schierholz and T. Streuer in the overlap operator implementation is kindly acknowledged. This work was partially supported by the EU Integrated Infrastructure Initiative Hadron Physics (I3HP) Grant (contract no. RII3-CT-2004-506078) and by the Russian Foundation for Basic Research (project nos. 05-02-16306a, 05-02-17642, 04-02-16079, and 03-02-16941). F.V.G. was partially supported by INTAS YS (grant no. 04-83-3943).

REFERENCES

1. M. Creutz, *Quarks, Gluons and Lattices* (Cambridge Univ. Press, Cambridge, 1983; Mir, Moscow, 1987).
2. T. Schafer and E. V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998); D. Diakonov, Prog. Part. Nucl. Phys. **51**, 173 (2003).
3. V. G. Bornyakov, M. N. Chernodub, F. V. Gubarev, *et al.*, Phys. Lett. B **537**, 291 (2002); F. V. Gubarev, A. V. Kovalenko, M. I. Polikarpov, *et al.*, Phys. Lett. B **574**, 136 (2003); A. V. Kovalenko, M. I. Polikarpov, S. N. Syritsyn, and V. I. Zakharov, Phys. Rev. D **71**, 054511 (2005); V. G. Bornyakov, P. Y. Boyko, M. I. Polikarpov, and V. I. Zakharov, Nucl. Phys. B **672**, 222 (2003); B. L. G. Bakker, A. I. Veselov, and M. A. Zubkov, Phys. Lett. B **544**, 374 (2002).
4. V. I. Zakharov, AIP Conf. Proc. **756**, 182 (2005); hep-ph/0312210.
5. J. Greensite, Prog. Part. Nucl. Phys. **51**, 1 (2003).
6. J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); A. M. Polyakov, Int. J. Mod. Phys. A **14**, 645 (1999).
7. E. Witten, Nucl. Phys. B **156**, 269 (1979); G. Veneziano, Nucl. Phys. B **159**, 213 (1979).
8. T. Banks and A. Casher, Nucl. Phys. B **169**, 103 (1980).
9. H. Neuberger, Phys. Lett. B **417**, 141 (1998); Phys. Lett. B **427**, 353 (1998).
10. M. Luscher, Phys. Lett. B **428**, 342 (1998).
11. M. Teper, Nucl. Phys. Proc. Suppl. **83**, 146 (2000).
12. I. Horvath, S. J. Dong, T. Drape, *et al.*, Phys. Rev. D **66**, 034501 (2002); I. Horvath *et al.*, Phys. Rev. D **68**, 114505 (2003); I. Horvath, S. J. Dong, T. Draper, *et al.*, Phys. Rev. D **67**, 011501 (2003).
13. L. Giusti, C. Hoelbling, M. Luscher, and H. Wittig, Comput. Phys. Commun. **153**, 31 (2003).
14. S. Capitani, M. Gockeler, R. Horsleyet, *et al.*, Phys. Lett. B **468**, 150 (1999).
15. B. Lucini and M. Teper, J. High Energy Phys. **0106**, 050 (2001).
16. S. J. Hands and M. Teper, Nucl. Phys. B **347**, 819 (1990).
17. B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993); C. Gattringer *et al.*, Nucl. Phys. B **617**, 101 (2001); T. Kovacs, Phys. Rev. D **67**, 094501 (2003).
18. C. Aubin *et al.* (MILC Collab.), hep-lat/0410024.
19. J. Greensite, S. Olejnik, M. Polikarpov, *et al.*, Phys. Rev. D **71**, 114 507 (2005).
20. P. de Forcrand and M. D'Elia, Phys. Rev. Lett. **82**, 4582 (1999).
21. J. Gattnar *et al.*, Nucl. Phys. B **716**, 105 (2005).
22. A. V. Kovalenko, M. I. Polikarpov, S. N. Syritsyn, and V. I. Zakharov, Phys. Lett. B **613**, 52 (2005).
23. M. Engelhardt and H. Reinhardt, Nucl. Phys. B **567**, 249 (2000).

JETP Letters, Vol. 82, No. 6, 2005, pp. 350–355. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 388–394.

Spontaneous Parametric Down-Conversion

G. Kh. Kitaeva and A. N. Penin*

Department of Quantum Electronics, Faculty of Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia

e-mail: penin@qopt.phys.msu.su

Received July 19, 2005

Spontaneous parametric down-conversion (SPDC) in media with no inversion center and the use of this phenomenon in the spectroscopy of natural oscillation states of a crystal lattice (i.e., optical phonons) are retrospectively described. We think that the SPDC spectroscopy method is estimated inappropriately and hope to again attract the attention of readers to one of the most interesting quantum phenomena of nonlinear optics that has no classical analog. The capabilities of SPDC spectroscopy will certainly be used in both fundamental science and technology of new materials. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 42.65.Lm, 42.70.Mp, 77.84.–s

The first report on the theoretical possibility of spontaneous parametric down-conversion (at that time, the term parametric fluorescence was used) was made by D.N. Klyshko in 1966 at the All-Union Conference on Nonlinear Properties of Media in Chernogolovka. The first publication appeared in *JETP Letters* in 1967 [1]. In the same year, this phenomenon was experimentally observed simultaneously by three research groups [2–4]. The history of the discovery of this scattering is dramatic; unfortunately, it is impossible to describe this history in detail in this paper. In 1983, Klyshko and two of his disciples were awarded a USSR State Prize for the discovery and investigation of the phenomenon of spontaneous parametric down-conversion and its application in spectroscopy and metrology. Almost since the time of its discovery, SPDC began to be efficiently used, first, in spectroscopy and, then, in quantum metrology and quantum optics. At present, most works on quantum optics and quantum informatics that are associated with investigation and application of quantum features of the electromagnetic field are based on SPDC. The properties of the process allow the generation of specific field states that provide the basis of a number of protocols of quantum calculations and protection of quantum information [5].

The use of SPDC in quantum optics is widely known. However, the prospects of its application for crystal spectroscopy, especially in the region of low-frequency optical phonons and polaritons in the wavelength range from tens of microns to millimeters, are much less known. In the frequency range of optical phonons, SPDC spectra smoothly join the spectra of Raman scattering on phonon polaritons [6, 7]. Although these processes are similar at first glance, they are fundamentally different. Spontaneous parametric down-conversion can be treated as the scattering of light by zero-point electromagnetic vacuum fluctuations in a medium due to its quadratic optical susceptibility. At the same time, the spontaneous Raman scattering on phonon polaritons is scattering due to the imaginary part of cubic optical susceptibility. As a rule, the components of the Raman scattering tensor decrease rapidly to zero when the polariton frequency is detuned from the optical-phonon frequency and, therefore, the basic contribution to the intensity of light scattered on polaritons comes from SPDC. In centosymmetric media, the parametric contribution is absent and the spontaneous scattering of light is pure Raman. In this case, Raman scattering can occur only directly on phonons or on very narrow adjacent sections of the polariton branches.

As was shown by Klyshko [8, 9], SPDC can be described only in the framework of consistent quantum theory. For an observer, SPDC is an optical parametric process of the spontaneous decay of monochromatic-radiation (pumping) photons incident on the medium with frequency ω_0 into a pair of signal (frequency ω_1) and idler (frequency ω_2) photons:

$$\hbar\omega_0 \longrightarrow \hbar\omega_1 + \hbar\omega_2. \quad (1)$$

The sum of the frequencies of the produced photons is equal to the pumping frequency. For nonstationary or quasimonochromatic pumping, this condition is evidently valid with an accuracy of the spectral width of the pumping. The frequencies of the produced photons lie in broad ranges from zero to the pumping frequency. The signal-photon frequency ω_1 and idler-photon frequency ω_2 are traditionally associated with the ranges $\omega_0 > \omega_1 \geq \omega_0/2$ and $0 < \omega_2 \leq \omega_0/2$. The maximum decay efficiency is achieved when the momentum conserva-

* A member of the editorial board of the journal *JETP Letters* since 1998.

tion law is valid, which has the form of the condition of spatial (phase) synchronism in this process

$$\Delta\mathbf{k} \equiv \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_0 = 0. \quad (2)$$

Here, $\mathbf{k}_0$, $\mathbf{k}_1$, and $\mathbf{k}_2$ are the wave vectors of the pumping, signal, and idler waves, respectively, and $\Delta\mathbf{k}$ is the phase detuning.

When the frequencies of both photons lie in the transparency range of the scattering medium, a biphoton field, which is the flux of correlated photon pairs, is generated at the output of the medium. A pair of photons, i.e., a biphoton, is an indivisible quantum object that constitutes the so-called entangled state. This state is described by a united wave function and has a number of statistical properties. The photons in the pair are strictly coupled with each other by the place and time of their production, frequency, and scattering direction. The biphoton flux intensity is directly related to the radiance of zero-point electromagnetic vacuum fluctuations. When $\omega_1 = \omega_2 = \omega_0/2$ and $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}_0/2$, a situation unique for experimental quantum optics is realized: the signal and idler photons become undistinguishable.

If the frequency of the idler photon lies in the range of strong phonon absorption, only a signal photon of each pair remains at the output of the scattering object. In this case, the conditions for the scattering of light on polaritons are realized. The intensity of the signal photon flux can be determined in this case not only by zero-point electromagnetic vacuum fluctuations but also by thermal fluctuations of the field at the idler frequency, which occur in the absorbing medium at sufficiently high temperature. Owing to the strong correlation between the frequencies and directions of photons produced in SPDC, the spatial and frequency distributions of the intensity of the signal radiation that are detected in the visible range contain information on the optical and dynamic parameters of the eigenstates of the scattering medium in the infrared, far infrared, and terahertz ranges, which are poorly accessible for direct observation. It is the process of heterodyning of low-frequency fluctuation fields of the long-wavelength range into the visible range. In this case, the pumping radiation serves as the heterodyne radiation. The two-dimensional spectrum of SPDC, which is the distribution of the signal-radiation intensity in the (frequency, scattering angle) plane, can be detected by means of a spectrographic system with crossed frequency and angular dispersion. The Fourier image of the signal-radiation flux is plotted in the input plane of the spectrograph [10].

In the phenomenological consideration, the radiance of the signal radiation is determined by the macroscopic parameters of the medium and the pumping intensity. In particular, if the thermal fluctuations are negligibly small, the signal radiance (measured in photons per signal-wave mode at the output of the scattering crystal) is given by the expression [9]

$$N_1 = F(\mathbf{k}_1)(N_2 + 1). \quad (3)$$

Here, N_2 is the number of free-field photons in the idler wave mode at the input of the crystal, 1 in the parentheses is the effective radiance of zero-point electromagnetic vacuum fluctuations measured in photons per the idler radiation mode, and F is the parametric transformation coefficient. In a transparent crystal, this coefficient is given by the formula

$$F = 4\pi^2 c^{-2} \omega_1 \omega_2 |\chi(\omega_1 = \omega_0 - \omega_2)|^2 |E_0|^2 l_\Delta^2,$$

where $\chi(\omega_1 = \omega_0 - \omega_2)$ is the convolution of the quadratic susceptibility tensor of the crystal with the polarization orts of the pumping, signal, and idler waves; E_0 is the pumping-wave amplitude; and l_Δ is the coherent interaction length (coherent length). The radiance of the signal radiation of SPDC that is observed only due to zero-point electromagnetic vacuum fluctuations (i.e., at $N_2 = 0$) can be sufficiently high. In particular, the effective radiance temperature reaches approximately 1500 K for a pumping radiation intensity of 1 W, a $LiNbO_3$ single crystal 1-cm long, a pumping wavelength of 0.51 μm, and a signal wavelength of 0.6 μm. It is worth noting that the radiance of a usual 100-W filament bulb does not exceed this value. For $N_2 \neq 0$, formula (3) also describes an induced process—the linear (in intensity and radiance) parametric transformation of the input idler radiation. As was experimentally shown in [11], absorption at the idler wave does not affect the integral signal intensity but changes the frequency and angular (in the **k** space) lineshape. This evidence finished the discussion on the basic properties of SPDC and Raman scattering on polaritons (see, e.g., [12]) and led to the development of SPDC spectroscopy. More recently, it was proved experimentally that the effective radiance of zero-point electromagnetic vacuum fluctuations in parametric processes is equal to unity even if the nonlinear medium absorbs and partially reflects radiation incident from the outside [13].

SPDC spectroscopy is of interest because a number of linear and nonlinear optical parameters of the medium, as well as the dynamic parameters of a crystal lattice, can be determined from the frequency–angular spectra. In particular, the reflective index and absorption coefficient can be measured over the entire scattering range. To solve this problem, it is necessary to know only the refractive index in the visible range at the pumping and signal frequencies. The behavior of the tuning curve (the frequency dependence of the scattering angle signal for zero wave detuning) enables one to determine the contribution of particular phonons to the optical parameters of crystals, as well as the frequencies, damping constants, and even oscillator strengths of these phonons. The measurement of the scattering intensity allows the determination of the quadratic susceptibility and the imaginary part of the cubic suscepti-

bility of the medium. The frequency and angular widths of the tuning curve are unambiguously related to the absorption coefficient at the idler wave frequency (under the assumption of small absorption at the signal and pumping frequencies). We note that SPDC spectra exhibit oscillations that are active only in infrared absorption and oscillations that are active only in Raman scattering, as well as oscillations for which alternative prohibition is removed. Owing to such an indifference of scattering to the activity type of the oscillations, SPDC spectra are very sensitive to the presence of even very weak (in their contribution to the crystal susceptibility) oscillation states of the lattice. In particular, higher order oscillations were detected that have no Raman scattering activity (the nonlinearity of the third order in the field is close to zero) and have an extremely small oscillation strength (characterizing the dipole moment and oscillation activity in infrared absorption) on the order of 10^{-8} [14].

In addition to the practically important information on the dispersion of the optical parameters of crystals in the infrared range, the change in the dispersion curves in the presence of doping and the variation in the composition and the structure of crystal defects (see, e.g., [15–17]), SPDC spectroscopy, during its existence, has provided a number of nontrivial and often unique results interesting for fundamental science. We remind some of them.

Ferroelectric phase transition in the potassium dihydrophosphate crystal (KDP, chemical formula KH_2PO_4). The SPDC spectra enabled one to find that, near the point of the ferroelectric phase transition in KDP ($T_c \sim 123$ K), a change occurs in the oscillator strengths of not only the soft mode whose oscillator strength increases sharply and determines the change in the specific heats at zero frequency but also in a number of other phonons [18]. This statement primarily concerns PO_4-group oscillations in the frequency range 900–950 cm^{-1} whose oscillator strength decreases abruptly when approaching the Curie temperature in the ferroelectric phase. This fact is evidently of fundamental importance for the development of the theory of phase transitions. It was found that the domain structure formed in the ferroelectric phase leads to a significant change in the oscillation spectrum, namely, to the appearance of new oscillations of the first and higher orders and to the violation of the polarization selection rules [19]. These effects have not yet been explained in the framework of the available theories of the ferroelectric phase transitions and formation of domain structures.

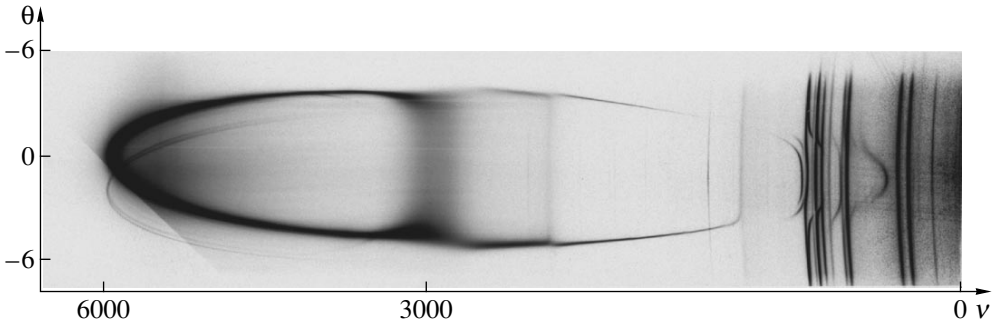
Concentration phase transition in the $KH_{2x}D_{2(1-x)}PO_4$ solid solution. When the proton concentration x decreases and the proton is replaced by the deuteron, a change is observed in the ordinary reflective index. The dependence is linear in the proton concentration range $x = 1–0.25$. However, as the deuteron concentration increases further, a kink characteristic of phase transitions appears in the dependence. The x scale is divided into two ranges with qualitatively different behaviors of the reflective index with varying deuteron concentration. It is worth noting that the position of this kink coincides with the x value at which inversion occurs in the valence oscillations of OH and OD bonds (their frequencies lie in the frequency range 1500–1800 cm^{-1} and their spectral widths reach 200 cm^{-1}) [20].

Anharmonicity of the oscillation states of the crystal lattice. The capabilities of SPDC spectroscopy are most brightly manifested in application to investigation of the anharmonicity of phonons; interaction of first-order oscillations with higher order oscillations, i.e., the Fermi resonance; and interference of susceptibilities of various orders, in particular, the Fano antiresonance. The coincidence of the frequencies of the first and second or higher orders is a rather rare phenomenon in Raman scattering spectra, whereas such a coincidence is a usual phenomenon in SPDC spectra due to the broad frequency band of phonon polaritons. The presence of higher order oscillations such as biphonons, sum and difference combination oscillations, and bands of uncoupled multiparticle states lead to a change in the optical parameters of crystals in the polariton region and thereby to a noticeable change in the behavior of the tuning curve of the SPDC spectrum.

A sufficiently complete theoretical description of these phenomena was given in a number of works by Agranovich and Lalov (see, e.g., [21]) and Strizhevskii *et al.* [22]. The Fermi resonance and Fano antiresonance are experimentally manifested in the spectra of almost all crystals in which scattering has been already observed (see, e.g., [14–16, 23, 24]). These effects are most brightly manifested in the spectra of crystals whose structure includes the hydrogen bond such as lithium formate, hydroiodic acid, potassium pentaborate, KDP, etc. The figure shows an example of the characteristic SPDC spectrum.

Investigation of effects of the electron–phonon interaction. As is shown, owing to the electron–phonon interaction, the localization of free electrons, and the formation of polaron states that lead to a change in the crystal absorption spectrum in the transparency range, various photoinduced effects, formation of holographic structures, etc., can occur in dielectric crystals. Additional changes in the phonon subsystem level are theoretically studied insufficiently and, as far as we known, have not yet been observed experimentally. In a recent investigation of SPDC spectra in chemically reduced crystals of doped lithium niobate Mg:$LiNbO_3$, it was found that the formation of polarons led to a change in the dispersion of the dielectric constant of the crystals near the phonon absorption edge [25]. The changes increase both with the polaron concentration and when approaching the phonon resonance region. The nature of these phenomena is not yet understood.

The above discussion concerns SPDC in spatially homogeneous samples in the form of a plane layer. The



Frequency–angular spectrum of spontaneous parametric down-conversion in the hydroiodic acid crystal. The horizontal axis shows the idler wavenumber ν in inverse centimeters and the vertical axis shows the scattering angle in degrees.

situation is much more complicated if a sample has either a regular or a stochastic spatial structure. Reflection from the internal boundaries between individual regions significantly complicates the spectrum. Various situations are possible: (i) only the linear parameters—the reflective index and absorption coefficient—undergo a significant jump at the internal boundaries, (ii) only the quadratic susceptibility or higher order susceptibilities undergo changes such as changes of the sign of even-order nonlinear susceptibilities in multidomain ferroelectrics, and (iii) both effects are manifested together. In order to adequately describe the form of the SPDC spectrum in the presence of spatial inhomogeneity, it is possible to use the generalized nonlinear Kirchhoff's law (GKL) formulated by Klyshko for the unified description of spontaneous and induced parametric processes in absorbing media [9, 26].

If the field at either the signal (ω_1) or idler (ω_2) frequency is present at the input of the medium along with the pumping, the processes of the induced conversion of the pumping photons accompanied by the formation of similar biphoton pairs occur along with the spontaneous process. If the external fields are sufficiently weak and the parametric transformation coefficient is much less than unity, the GKL describes the total field at the output of the three-wave parametric converter and is, in essence, the solution of the kinetic equations for the second moments of the field that take into account (in the consistent quantum consideration) both the thermal and quantum noises of the medium. According to the GKL, the mean number of photons in the output radiation modes of the signal range $N_{1'} \equiv \langle a_{1'}^{+} a_{1'} \rangle$ in the parametric down-conversion is given by the expression

$$\begin{aligned} N_{1'} &= \hat{U}_{1'1}(N_1 + \mathcal{N}_0 + I)\hat{U}_{1'1}^{+} \\ &- \hat{U}_{1'2}(\tilde{N}_2 - \mathcal{N}_0)\hat{U}_{1'2}^{+} - \mathcal{N}_0 - I. \end{aligned} \tag{4}$$

Here, $\hat{U}_{i'j}$ are the matrix elements of the scattering matrix $\hat{U} = \begin{pmatrix} \hat{U}_{1'1} & \hat{U}_{1'2} \\ \hat{U}_{2'1} & \hat{U}_{2'2} \end{pmatrix}$; the subscripts 1' and 2' run over all the modes of the output signal and idler radiations, respectively; the subscripts 1 and 2 run over all the modes of the input signal and idler radiations, respectively; I is the identity matrix; a_i^{+} and a_i are the operators of the production and annihilation of photons in the corresponding field modes, respectively; the matrices $(N_1)_{ij} \equiv \langle a_{1_i}^{+} a_{1_j} \rangle$ and $(N_2)_{ij} \equiv \langle a_{2_i}^{+} a_{2_j} \rangle$ describe the second moments of the input field; and the temperature factor $\mathcal{N}_0 \equiv 1/(\exp(\hbar\omega/kT) - 1)$ describing the contribution of the thermal fluctuations of the medium is much less than unity (i.e., than the contribution from zero-point electromagnetic vacuum fluctuations) even at the lowest idler frequencies and can be ignored.

The mean number of photons in the output radiation modes of the idler range, $N_{2'} \equiv \langle a_{2'}^{+} a_{2'} \rangle$, is determined by the expression

$$\begin{aligned} N_{2'} &= \hat{U}_{2'1}(\tilde{N}_1 + \mathcal{N}_0 + I)\hat{U}_{2'1}^{+} \\ &- \hat{U}_{2'2}(N_2 - \mathcal{N}_0)\hat{U}_{2'2}^{+} + \mathcal{N}_0. \end{aligned} \tag{5}$$

The difference between Eqs. (4) and (5) for the signal and idler photons is attributed to the assumption that absorption occurs only at the idler radiation frequencies. In this case, the correlation moment $K' \equiv \langle a_{1'} a_{2'} \rangle^{*}$ determining the statistical properties of the output biphoton field is given by the formula

$$K' = \hat{U}_{2'1}(\tilde{N}_1 + \mathcal{N}_0 + I)\tilde{\hat{U}}_{1'1} + \hat{U}_{2'2}(N_2 - \mathcal{N}_0)\tilde{\hat{U}}_{1'2}. \tag{6}$$

The formulation of the GKL in the form given by Eqs. (4)–(6) is called polariton Kirchhoff's law [9].

The GKL method allows the determination of the second correlation moments of the fields at the output of the nonlinear medium in terms of the second correlation moments specified at the input. Using this method, the general relations describing the dependence of the SPDC lineshape on the spatial distribution of the quadratic susceptibility in the nonlinear medium were obtained [27]. In these calculations, only a linear relation between the Heisenberg operators of the input and output fields is postulated. For a layered medium with the spatial modulation of the quadratic susceptibil-

ity $\chi \equiv \chi(\omega_1 = \omega_0 - \omega_2)$ along a single direction (Z), the two-dimensional scattering matrix U_{ij}, which describes the parametric conversion of the input signal and idler waves into the respective output waves in the approximation of linear amplification and given pumping, is calculated. Expressions obtained by means of the GKL allow the description of the frequency–angular distribution of the signal radiation of SPDC in a medium with arbitrary $\chi(z)$ dependence. The coordinate dependence $\chi(z)$ is represented by the Fourier spectrum of spatial harmonics

$$\chi(z) = \sum_{m=-\infty}^{\infty} \chi_m \exp(imqz) \tag{7}$$

with the amplitudes

$$\chi_m = \frac{1}{d}\int_{-d/2}^{d/2} \chi(z)\exp(-imqz)dz. \tag{8}$$

For a medium with regular periodic dependence $\chi(z)$, d is the period of the one-dimensional nonlinear superlattice, and $q \equiv 2\pi/d$ is the absolute value of the inverse-superlattice vector directed along the Z axis. If a sample with length l is characterized by irregular stochastic dependence $\chi(z)$, it can be considered as one period of the infinitely long regular lattice. In this case, $d = l$ and the number of periods $n \equiv l/d$ of the nonlinear lattice is formally equal to one. With allowance for the absorption of the medium at the idler-wave frequencies, the general expression for the SPDC signal intensity per unit angular and spectral intervals has the form

$$P_{\omega_1\Omega_1} = C_0 \sum_{m=-\infty}^{\infty} |\chi_m|^2 g(\Delta_m, y_2)$$

$$+ C_0 \sum_{m=-\infty}^{\infty} \sum_{\substack{m'=-\infty \\ m' \neq m}}^{\infty} (-1)^{n(m'-m)} \chi_m^* \chi_{m'} \tag{9}$$

$$\times \left[\frac{e^{-i\Delta - y_2} - 1}{(y_2 + i\Delta - i2\pi mn)(y_2 + i\Delta - i2\pi m'n)} + \frac{e^{-i\Delta - y_2} - 1}{(y_2 - i\Delta - i2\pi mn)(y_2 - i\Delta - i2\pi m'n)} \right].$$

Here, $\Delta_m \equiv \Delta - 2\pi mn$ is the detuning of quasi-phase matching for the mth order of nonlinear diffraction, where the detuning $\Delta_0 \equiv \Delta \equiv (\Delta\mathbf{k})_z l$ in the zeroth order coincides with the detuning of phase synchronism in the spatially homogeneous medium and is calculated, according to Eq. (2), in terms of the Z-axis projections of the signal, idler, and pumping wave vectors; $C_0 \equiv (\hbar\omega_1^4\omega_2/c^5 n_0 n_1 n_2) P_0 l^2/\cos\vartheta_2$ is the common factor proportional to the incident pumping power P_0; and $y_2 \equiv \alpha_2 l/2\cos\vartheta_2$, where α_2 is the Bouguer absorption coefficient at the idler-wave frequency and ϑ_2 is the angle between the normal to the nonlinear-layer plane and the wave vector of the idler wave. The frequency–angular shape of the scattering line is described by two contributions.

The first contribution is the additive sum of the contributions from the individual harmonics of quadratic susceptibility. Each term has the same distribution of the signal power with respect to the maximum as scattering in the homogeneous medium that is described by the function

$$g(\Delta, y_2) \equiv \frac{2}{(\Delta^2 + y_2^2)^2}[(\Delta^2 - y_2^2)(1 - e^{-y_2}\cos\Delta)$$

$$- 2y_2\Delta e^{-y_2}\sin\Delta + y_2(\Delta^2 + y_2^2)].$$

In this case, the position of new maximum lines—tuning curves for each term—is shifted with respect to the maxima of scattering in the homogeneous medium according to the quasi-phase matching condition $\Delta(\omega_1, \theta_1) = mql \equiv 2\pi mn$, when the inverse-lattice vector is additionally introduced under the wave synchronism condition. Each term is proportional to the amplitude modulus squared of the corresponding harmonic.

The second contribution to Eq. (9) contains the products of noncoinciding harmonics χ_m and is the result of their interference.

Analysis of the frequency–angular distribution of the signal intensity can be used as the basis for the measurement of the character of the periodic spatial variation in the quadratic nonlinearity in crystalline structures [28]. In principle, any three-wave parametric processes including induced up-conversion processes and generation of the second harmonic are appropriate for this aim [29]. However, when induced processes are used, the problem of taking into account nonuniform filling of the converter modes by the input radiation always arises. If SPDC is used, the absolutely regular filling of the input modes of the idler channel is ensured (in the linear-amplification regime). The SPDC method makes it possible to analyze inhomogeneous-structure islands, which are located in the crystal bulk, and to characterize the entire sample volume as a whole. This property constitutes its main advantage over the methods of chemical etching, atomic force or electron force microscopy, scanning electron microscopy, x-ray analysis, etc., that are applied to analyze only the surface of samples. It is possible to obtain nonlinear optical tomographic images of samples and to determine the spectral ranges most appropriate for use in real devices. An example of constructing the map of the distribution of the quality of a nonlinear optical superlattice along the input surface of a periodically poled sample based on the Y:$LiNbO_3$ crystal was given in [30].

Thus, SPDC spectroscopy is a simple, sensitive method for characterizing the most important micro- and macroscopic parameters of nonlinear media. In

the near future, interest in SPDC spectroscopy will evidently increase in view of the active development of the long-wavelength spectral range, particularly the terahertz range [31]. Unique possibilities of generating and detecting terahertz radiation in periodically poled nonlinear optical crystals were demonstrated in [32]. SPDC spectroscopy is very promising in this range of the lower polariton branch. Schemes for measuring dispersion characteristics and spatial structure, as well as for searching for and analyzing new highly efficient nonlinear structures, will be created using SPDC.

This work was supported by the Russian Foundation for Basic Research (project no. 05-02-16391) and the Council of the President of the Russian Federation for State Support of Young Scientists and Leading Scientific Schools (project no. NSh-166.2003.02 for Penin's scientific school).

REFERENCES

1. D. N. Klyshko, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 490 (1967) [JETP Lett. **6**, 23 (1967)].
2. S. A. Akhmanov, V. V. Fadeev, R. V. Khokhlov, and O. N. Chunaev, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 575 (1967) [JETP Lett. **6**, 85 (1967)].
3. S. E. Harris, M. K. Oshman, and R. L. Byer, Phys. Rev. Lett. **18**, 732 (1967).
4. D. Magde and H. Mahr, Phys. Rev. Lett. **18**, 905 (1967).
5. *The Physics of Quantum Information*, Ed. by D. Bouwmeester, A. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
6. C. H. Henry and J. J. Hopfield, Phys. Rev. Lett. **15**, 964 (1965).
7. S. P. S. Porto, B. Tell, and T. C. Damen, Phys. Rev. Lett. **16**, 450 (1966).
8. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **55**, 1006 (1968) [Sov. Phys. JETP **28**, 522 (1968)].
9. D. N. Klyshko, *Photons and Nonlinear Optics* (Nauka, Moscow, 1980) [in Russian].
10. M. V. Chekhova, G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Proc. SPIE **1863**, 192 (1993).
11. D. N. Klyshko, B. F. Polkovnikov, and A. N. Penin, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 11 (1970) [JETP Lett. **11**, 5 (1970)].
12. T. G. Giallorenzi and C. L. Tang, Phys. Rev. **184**, 353 (1969); C. H. Henry and C. G. B. Garrett, Phys. Rev. **171**, 1058 (1968).
13. O. H. Abroskina, G. Kh. Kitaeva, and A. N. Penin, Dokl. Akad. Nauk SSSR **280**, 584 (1985) [Sov. Phys. Dokl. **30**, 67 (1985)].
14. V. M. Ivanov, T. V. Laptinskaya, and A. N. Penin, Dokl. Akad. Nauk SSSR **260**, 321 (1981) [Sov. Phys. Dokl. **26**, 859 (1981)].
15. O. A. Aktsipetrov, G. Kh. Kitaeva, and A. N. Penin, Fiz. Tverd. Tela (Leningrad) **19**, 127 (1977) [Sov. Phys. Solid State **19**, 72 (1977)].
16. O. A. Aktsipetrov, G. Kh. Kitaeva, and A. N. Penin, Fiz. Tverd. Tela (Leningrad) **20**, 402 (1978) [Sov. Phys. Solid State **20**, 232 (1978)].
17. G. Kh. Kitaeva, K. A. Kuznetsov, A. N. Penin, and A. V. Shepelev, Phys. Rev. B **65**, 054304 (2002).
18. O. A. Aktsipetrov, G. Kh. Kitaeva, and A. N. Penin, Fiz. Tverd. Tela (Leningrad) **19**, 1001 (1977) [Sov. Phys. Solid State **19**, 582 (1977)].
19. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Fiz. Tverd. Tela (St. Petersburg) **34**, 3440 (1992) [Sov. Phys. Solid State **34**, 1841 (1992)].
20. I. V. Mityusheva, E. D. Mishina, and A. N. Penin, Fiz. Tverd. Tela (St. Petersburg) **22**, 2476 (1980) [Sov. Phys. Solid State **22**, 1443 (1980)].
21. V. M. Agranovich and I. I. Lalov, Zh. Éksp. Teor. Fiz. **61**, 656 (1971) [Sov. Phys. JETP **34**, 350 (1971)].
22. V. L. Strizhevskiĭ, Zh. Éksp. Teor. Fiz. **62**, 1446 (1972) [Sov. Phys. JETP **35**, 760 (1972)]; V. L. Strizhevskiĭ, H. Ponath, and Yu. N. Yashkir, Opt. Spektrosk. **31**, 388 (1971).
23. L. I. Kuznetsova, L. A. Kulevskiĭ, K. A. Prokhorov, and Yu. N. Polivanov, Kvantovaya Élektron. (Moscow) **2**, 2095 (1975).
24. Yu. N. Polivanov, Usp. Fiz. Nauk **126**, 185 (1978) [Sov. Phys. Usp. **21**, 805 (1978)].
25. G. Kh. Kitaeva, K. A. Kuznetsov, V. F. Morozova, *et al.*, Appl. Phys. B **78**, 759 (2004).
26. D. N. Klyshko, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1478 (1982).
27. G. Kh. Kitaeva and A. N. Penin, Zh. Éksp. Teor. Fiz. **125**, 307 (2004) [JETP **98**, 272 (2004)].
28. G. Kh. Kitaeva and A. N. Penin, Kvantovaya Élektron. (Moscow) **34**, 597 (2004).
29. G. Kh. Kitaeva, V. V. Tishkova, and A. N. Penin, J. Raman Spectrosc. **36**, 116 (2005).
30. G. Kh. Kitaeva, V. V. Tishkova, I. I. Naumova, *et al.*, Appl. Phys. B (2005) (in press).
31. J. Shan, A. Nahata, and T. F. Heinz, J. Nonlinear Opt. Phys. Mater. **11**, 31 (2002).
32. G. H. Ma, G. Kh. Kitaeva, I. I. Naumova, and S. H. Tang, physics/0503173.

Translated by R. Tyapaev

JETP Letters, Vol. 82, No. 6, 2005, pp. 356–365. Translated from Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 395–406.

Self-Consistent Turbulent Convection in a Magnetized Plasma

V. P. Pastukhov* and N. V. Chudin

Russian Research Centre Kurchatov Institute, pl. Akademika Kurchatova 1, Moscow, 123182 Russia

e-mail: past@nfi.kiae.ru

Received July 15, 2005

It has been shown that low-frequency vortex convection, which is self-consistently developed in a magnetized plasma, may lead to nondiffusive transport processes similar to those observed in various systems of the plasma magnetic confinement. To theoretically analyze such a convection, an approach is proposed on the basis of the direct computer simulation of the quasi-two-dimensional dynamics of a weakly dissipative plasma with the use of adiabatically reduced hydrodynamic-type equations. The derived equations ensure the description of both relatively fast nonlinear convective flows and slower resulting transport processes and allow the simulation of the plasma evolution at sufficiently long times comparable with the plasma lifetime. The simulation shows that the development of the convection leads to the formation of nonlinear large-scale stochastic vortex structures, which exhibit broad power-law frequency and wavenumber spectra, as well as the non-Gaussian statistics of fluctuations, and corresponds to the notion of structure turbulence. The resulting transport processes are nonlocal and nondiffusive and have a number of characteristic properties similar to those observed in real experiments. The self-consistency of the pressure and density profiles in the plasma, *L–H* transitions, impurity pinch, etc., are among these properties. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 52.30.–q, 52.35.Ra

1. INTRODUCTION

The problem of the anomalous transport of particles and energy is one of the key problems both in laboratory investigations on the magnetic confinement of a high-temperature plasma and in the study of a magnetized plasma in natural objects including the Earth's magnetosphere. In most theoretical works, anomalous transport processes are associated with plasma density and temperature fluctuations induced by various types of drift instabilities (see, e.g., reviews [1, 2]). Since the transverse scales of such fluctuations are small, the anomalous transport is traditionally discussed in terms of the diffusion approximation with local transport coefficients. However, numerous recent experiments indicate that anomalous transport processes may often have a different physical nature that cannot be reduced to the diffusion approximation. In particular, nondiffusive transport processes can be associated with the development of unsteady large-scale vortex convection of the plasma.

Concerning nondiffusive transport, it is worth noting that low-frequency fluctuations of the plasma and the resulting transport flows observed in many magnetic systems, which strongly differ in both the topology of the magnetic field and plasma parameters, exhibit quite common properties. In particular, low-frequency fluctuations of the electric field and plasma density in tokamaks [3, 4], stellarators [5, 6], tandem mirrors [7, 8], direct systems with a uniform magnetic field [9–11], etc., have spectra characteristic of strong turbulence in which stochastically excited nonlinear structures (e.g., vortices) dominate. In view of this circumstance, such a turbulence is often called structure turbulence [5, 6, 9]. In this case, quite large radial and poloidal correlation lengths of fluctuations [4, 6] indicate that such nonlinear structures are large-scale and, correspondingly, the transport processes associated with them are nonlocal and nondiffusive.

The statistical properties of fluctuations observed in various magnetic systems are also very similar despite the difference in certain specific details [5, 6, 9]. For example, the probability distribution functions for the increments of fluctuations that are obtained by processing of experimental results from various devices are non-Gaussian with "heavy tails," which indicates both an increased statistical contribution from large fluctuations and the nondiffusive character of the process.

Finally, the development of more perfect diagnostics made it possible to detect the presence of relatively long-lived macroscopic structures in a number of experiments. Blobs detected in the peripheral region of the plasma (in SOL) at the DIII-D tokamak [12] and then in many other systems are examples of such structures. Moving along the magnetic field at distances much larger than their sizes, blobs can carry the main part of the total fluxes of particles and energy in SOL. The diagnostics based on the x-ray tomography of a plasma column enabled one to reveal the generation of vortex structures whose sizes are comparable with the small plasma radius in the GAMMA 10 tandem mirror

* A member of the editorial board of the journal *JETP Letters* since 1988.

[8]. Thus, nondiffusive transport processes associated with the development of a certain type of structure turbulence are typical in the physics of magnetic plasma confinement and require serious experimental and theoretical investigations.

The difficulties of experimental investigations of structure turbulence are obvious. They will be undoubtedly overcome as diagnostic methods are developed and refined. At the same time, in investigations of such a complex and many-sided phenomenon as nondiffusive plasma transport, it is exclusively important to develop a theory for the adequate description and simulation of the most essential properties of structure turbulence itself and the resulting transport processes induced by it.

Analysis of the results of numerous experiments (see, e.g., [5–12]) shows that relatively large-scale structures are present and even dominate in the low-frequency turbulence in various magnetic systems, because such structures must be strongly extended along the magnetic field. As a result, turbulence appears to be two-dimensional (2D) or quasi-two-dimensional and such a turbulence is characterized by the presence of both direct and inverse cascades [13] in which the spectral redistribution of energy occurs in the direction of both small and large transverse scales. Under weak dissipation conditions, strong 2D turbulence must lead to the establishment of quite universal spectra, which decrease according to the power law in the direction of small scales and depend slightly on the spatial scale of feeding instability.

For theoretical analysis of 2D turbulence in the magnetized plasma, where relatively large-scale low-frequency structures are expected to dominate, an approach based on the direct simulation of the nonlinear dynamics of the plasma described by the moment hydrodynamic-type equations (equations of continuity, motion, energy transfer, etc.) seems to be very fruitful. Such equations form a finite set of integral consequences of the initial kinetic equations supplemented by certain closing conditions. It is expected that, despite the less detailed description compared to kinetics, this approach will provide a sufficiently complete self-consistent description of the evolution of the basic macroscopic parameters of the plasma (density, velocity, and temperatures of various plasma components). At the same time, the use of this approach significantly facilitates the qualitative understanding and numerical simulation of the dynamic processes under investigation.

It is essential that, as applied to the simulation of low-frequency turbulence (in particular, with frequencies on the order of drift ones), this approach allows further simplification of the dynamic problem. Indeed, the aforementioned moment equations describe too wide a variety of collective degrees of freedom. They usually include fast (i.e., high-frequency) stable collective degrees of freedom such as, e.g., magnetosonic waves. In view of strong difference in the characteristic frequencies, such fast degrees of freedom are not excited by the low-frequency turbulence under investigation and may be excluded from consideration by a certain reduction of the initial system of moment equations.

The reduced magnetohydrodynamic (MHD) equations were first proposed for describing the dynamics of the tokamak plasma by Kadomtsev and Pogutse [14] and then this idea was developed in [15–17] and further in many works that are primarily associated with H. Strauss. In all these works, the small parameter B_p/B_T (the ratio of the poloidal and toroidal magnetic fields) is used to exclude magnetoacoustic waves from the equations of the nonlinear low-frequency dynamics of the plasma in tokamaks and stellarators. More recently, a variational method of the adiabatic separation of fast and slow motions in continuous Lagrangian systems was developed in [18, 19]. This method allows the exclusion of fast stable collective degrees of freedom of the dynamic system by using various small parameters without the violation of kinematic and dynamic conservation laws inherent in the initial unreduced system of equations. The latter circumstance is exclusively important for the simulation of long-term dynamic evolution of the weakly dissipative plasma at times comparable with its lifetime.

Thus, the direct dynamic simulation of low-frequency turbulence and the resulting nondiffusive transport processes in various systems of plasma magnetic confinement using adiabatically reduced equations seems to be quite promising for theoretical investigations in plasma physics. In this work, the efficiency of this approach is demonstrated by a relatively simple but nontrivial example. In Section 2, the system of reduced equations used in the dynamic model under consideration is discussed. The results of the computer simulation of low-frequency turbulent convection and the resulting transport processes in the magnetized plasma are presented in Section 3. Section 4 briefly summarizes the basic results.

2. ADIABATICALLY REDUCED MHD MODEL OF THE NONLINEAR SELF-CONSISTENT DYNAMICS OF A WEAKLY DISSIPATIVE NONPARAXIAL PLASMA

In order to demonstrate the efficiency of the approach under development, we consider a relatively simple but quite nontrivial and conceptual example of the axisymmetric magnetic system with a “levitating dipole,” i.e., with a current-carrying ring placed into the plasma. This system schematically shown in Fig. 1 has the same topology of the magnetic surfaces as a spherical tokamak. In particular, the plasma equilibrium in it, as well as in the tokamak, is described by the Grad–Shafranov equation. However, owing to the absence of the toroidal magnetic field and, correspondingly, magnetic shear, the physically meaningful and quite non-

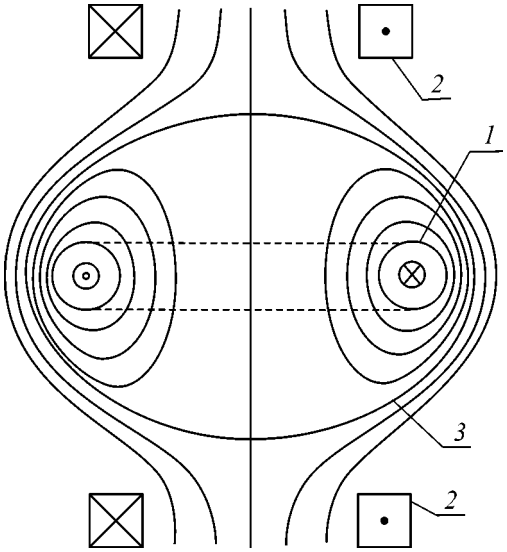


Fig. 1. Configuration of field lines in the trap with a levitated internal ring: (*1*) inner ring, (*2*) external-field coil, and (*3*) separatrix.

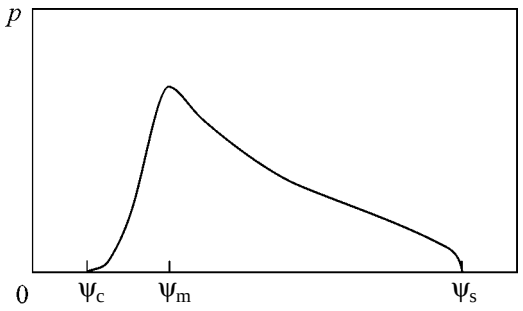


Fig. 2. Pressure profile in the magnetic trap with a levitated internal ring: ψ_c is the internal-ring surface, ψ_m is the maximum-pressure surface, and ψ_s is the outer separatrix surface.

trivial results concerning the nonlinear dynamics of the plasma in this system can be obtained even in the framework of the simplest one-fluid MHD model with the isotropic pressure of the plasma.

The magnetic system under consideration belongs to a quite wide class of strongly nonparaxial magnetic systems, where plasma confinement is based on an alternative approach to the problem of MHD stability [20]. The magnetic field in such systems decreases in the direction of the outer magnetic separatrix surface. As a result, only pressure profiles decreasing quite slowly and smoothly to the periphery are MHD stable. These systems exhibit a pronounced tendency to self-organization in the form of the maintenance of a pressure profile marginally stable against the flute MHD mode. In the framework of the simplest isotropic MHD model, the marginally stable pressure profile satisfies the condition $S \equiv pU^\gamma = \text{const}$, where $p(\psi)$ is the plasma pressure, $U(\psi) = \oint dl/B$ is the specific volume of a magnetic flux tube (i.e., the volume of the magnetic flux tube with unit magnetic flux), ψ is the function of the poloidal magnetic flux that is a generalized radial coordinate, γ is the adiabatic index, and S is the single-valued function of the plasma entropy in the magnetic flux tube with volume U. According to the linear theory of stability, the pressure profile $p(\psi)$ shown in Fig. 2 is expected to be formed in such a system. In this case, the plasma is stable against all MHD modes in the inner region $\psi_c < \psi < \psi_m$, where $S'(\psi) > 0$ and marginally stable with respect to the flute modes (S = const) in the outer region $\psi_m < \psi < \psi_s$. The S = const profile in the outer region is also stable under all incompressible Alfvén perturbations if $\beta \equiv 8\pi p/B^2$ does not exceed a certain critical value $\beta_{cr} \sim 1$.

In [20, 21], to simulate the low-frequency nonlinear convection of the plasma, an additionally simplified model of the magnetic configuration in the form of a cylindrical quasi-equilibrium plasma column with a rigid current-carrying rod with radius r_c on its axis was used. The self-consistent evolution of the plasma is expected to develop according to the following scenario. The heating of the plasma and initial local collisional thermal conductivity distort the initial pressure profile making it weakly unstable under the flute modes ($S'(\psi) < 0$). Instability excites and maintains nonlinear MHD convection, which in turn tends to recover the marginally stable pressure profile, leading to the substantially nonlocal increased transport of particles and energy. To correctly describe and simulate the long-term evolution of the plasma and to overcome the difficulties associated with the significant difference between the ideal and dissipative time scales, the method of adiabatic separation of fast and slow motions [18, 19] (ASM method) was used in [20]. By means of the ASM method under the assumption $\beta < \beta_{cr}$, fast stable magnetosonic incompressible Alfvén, and longitudinal acoustic modes of plasma oscillations were excluded from the consideration and the reduced equations were obtained for describing slower (adiabatic) flute 2D convection of the plasma, as well as the resulting transport processes with allowance for small dissipation caused by collisional thermal conductivity, viscosity, and diffusion.

In [22], the generalized reduced MHD equations were derived which allow the simulation of the low-frequency flute dynamics of the weakly dissipative plasma in arbitrary axisymmetric shearless magnetic systems with closed or open field lines. The possibility of excluding fast stable modes and the quasi-two-dimensional character of flute plasma motions holds for a more complex geometry of the magnetic field. Nevertheless, in this work, we discuss only results obtained with the simplified cylindrical model of the magnetic field, because the basic results of the numerical simula-

tion are obtained for this model. For corresponding generalizations, see [22].

As was shown in [20–22], fast stable MHD modes can be excluded under the assumption that the dynamic plasma states under consideration deviate slightly from the marginally stable state in flute modes. For the weakly nonideal MHD model, the small parameter ϵ, which allows the application of the adiabatic separation of motions, has the form

$$\epsilon \sim (\chi/ac_s)^{1/3} \ll 1, \quad (1)$$

where χ is the initial local collisional thermal diffusivity, a is the size on the order of the small plasma radius, and c_s is the speed of sound. After the exclusion of fast degrees of freedom, the reduced dynamics of the system in the cylindrical geometry of the field is completely described by the adiabatic velocity field

$$\mathbf{v}_a = [\mathbf{B} \times \nabla\Phi]/B^2 \sim \epsilon c_s, \quad (2)$$

where an arbitrary function $\Phi(t, r, z)$ serves as the 2D electric potential. The expected scale of the velocity $v_a \sim \epsilon c_s$ is estimated from the expected available potential energy and agrees with the condition of nonlinear saturation of the convection. According to [22], in the general geometry, the velocity field is three-dimensional but is as earlier determined by only the 2D potential $\Phi(t, \psi, \varphi)$ in the flux coordinates ψ and φ. Since the z coordinate in the cylindrical geometry simulates the toroidal angle φ of real systems, the effective angle $\varphi = z/R$ can be introduced in this geometry and all functions can be supposed to be periodic in this angle.

Reduced equations are written in terms of more appropriate variables: the specific volume of the magnetic flux tube $U = 2\pi r/B$, the entropy function $S = pU^\gamma$, and the renormalized density $\lambda = \rho U$ (or the number of particles in the magnetic flux tube with volume U), which are introduced instead of the magnetic field B, pressure p, and density ρ, respectively. These variables allow the explicit inclusion of the invariant properties of the initial MHD equations. By analogy with the case of more general field geometry, we introduce the flux coordinate ψ instead of r and normalize it such that $\mathbf{B} = [\nabla\psi \times \nabla\varphi]$. Under the assumption of the closeness to the marginally stable state, the entropy function is represented in the form $S(t, \psi, \varphi) = S_0(t, \psi) + \tilde{S}(t, \psi, \varphi)$, where $\tilde{S} \sim \epsilon^2 S_0$ describes fluctuations of the entropy function, and the quasi-equilibrium component $S_0(t, \psi)$ averaged with respect to φ is assumed to be such that its derivative with respect to ψ cannot take negative large-magnitude values; i.e., $\partial_\psi S_0(t, \psi) > -\epsilon^2 S_0/\psi_s$. The function $\lambda(t, \psi, \varphi)$ allows a similar representation. The reduced equations include the subsystem of "slow" equations ($\partial_t \sim \epsilon^3 c_s/a$), which describe the 1D (radial) transport of the plasma in the presence of convective motions and fluctuations in entropy and density, and the subsystem of "fast" equations ($\partial_t \sim \epsilon c_s/a$), which describe the 2D dynamics of the plasma and fluctuations of entropy $\tilde{S}$ and density $\tilde{\lambda}$.

The adiabatic velocity field (2) does not disturb the magnetic field and, therefore, fast fluctuations of **B** and, correspondingly, U are absent. In the simplified cylindrical model, the first slow equation determining $U(t, \psi)$ in terms of the known function $S_0(t, \psi)$ corresponds to the balance of the main radial forces (quasiequilibrium)

$$U^{-1}\partial_\psi\left(\frac{\gamma S}{\gamma - 1}U^{-(\gamma-1)} + \frac{r^2}{U}\right) - \frac{U^{-\gamma}}{\gamma - 1}\partial_\psi S = 0. \quad (3)$$

According to [22], in the case of the general axisymmetric shearless magnetic field, the quasiequilibrium condition corresponds to the Grad–Shafranov equation

$$\Delta^*\psi + 4\pi r^2 \partial_\psi (S_0(t, \psi)/U^\gamma) = 0,$$
$$\Delta^*\psi \equiv r^2\left(\nabla\frac{\nabla\psi}{r^2}\right), \quad (4)$$

which enables one to determine the function $\psi(t, r, z)$ and, then, **B** and $U(t, \psi)$ are reconstructed using this function. In turn, the function $S_0(t, \psi)$ is determined from the heat transfer equation averaged with respect to φ. After a change to new variables, this equation assumes the form

$$\partial_t|_\psi \hat{S}_0 - \partial_\psi(\overline{\tilde{S}\partial_\varphi\Phi})$$
$$= 2(\pi R)^2(\gamma - 1)U^{\gamma-1}\partial_\psi\left(\frac{r^2\lambda_0\chi}{U^2}\partial_\psi\left(\frac{S_0}{\lambda_0 U^{\gamma-1}}\right)\right) \quad (5)$$
$$+ (\gamma - 1)U^\gamma Q_E,$$

where the time derivative is calculated at a given ψ value, the bar stands for averaging with respect to φ, and the function $Q_E(t, \psi)$ describes the energy source. All the terms in Eq. (5) have an order of ϵ^3. The second term on the left-hand side of Eq. (5) describes the nondiffusive heat transfer caused by 2D convection. The equation for the function $\lambda_0(t, \psi)$ is similar to the equation for $S_0(t, \psi)$.

The reduced dynamic equation assumes the form

$$\partial_t|_\psi \hat{w} + [\Phi, \hat{w}] - \left[\lambda, \frac{v_a^2}{2}\right] + \frac{1}{U^\gamma}\partial_\psi U \partial_\varphi \tilde{S}$$
$$= U\nabla\left(C_A^{-2}\nabla\left(\frac{\eta C_A^2}{U}\hat{w}\right)\right), \quad (6)$$
$$[\Phi, f] \equiv \partial_\psi\Phi\partial_\varphi f - \partial_\varphi\Phi\partial_\psi f,$$

where $\hat{w}$ is associated with the vorticity of the convective flow described by velocity field (2) and is given by the expression

$$\hat{w} = U\nabla(\nabla\Phi/C_A^2), \quad (7)$$

η is the local kinematic viscosity, and C_A is the Alfvén velocity. As was emphasized in [20, 22], the quantity $\hat{w}$, which can be called the specific dynamic vorticity of the magnetic flux tube, has the meaning of the generalized momentum canonically conjugate to the generalized adiabatic coordinate of the flute convective motion. We also note that velocity field (2), as well as Eqs. (6) and (7), describes both pure vortex motion of the plasma and flows homogeneous in φ but inhomogeneous in the radius, which are often called shear or zonal flows. It is well known that such flows can exist even in the stationary equilibrium states of the plasma. In our consideration, shear flows constitute an essential part of the general nonlinear dynamic process and can self-consistently change under the action of vortex convection.

The equation for entropy fluctuations is obtained from the general transport equation for the plasma thermal energy by subtracting Eq. (5) from it:

$$\partial_t|_\psi \tilde{S} + [\Phi, \tilde{S}] + \partial_\psi(\overline{\tilde{S}\partial_\varphi\Phi}) - \partial_\varphi\Phi\partial_\psi S_0 = \frac{\gamma-1}{2\gamma}U^\gamma\nabla\left(\frac{\lambda_0\chi c_s}{U}\nabla\left(c_s\frac{\tilde{S}}{S_0}\right)\right). \quad (8)$$

The left-hand sides of Eqs. (6) and (8) describe the ideal plasma dynamics including the generation of linear instability and the nonlinear coupling and saturation of the modes. According to the expected order of magnitude of the quantities, all the terms on the left-hand sides have the same order of magnitude. The dissipative terms on the right-hand sides of Eqs. (6) and (8) are small as $\epsilon^2a^2|\nabla|^2$ compared to the terms on the left-hand sides and describe the dissipation of small-scale fluctuations. The equation for fluctuations $\tilde{\lambda}$ is similar to the equation for $\tilde{S}$. The generalization of the above equations to the case of an arbitrary axisymmetric shearless magnetic field was given in [22].

3. COMPUTER SIMULATION RESULTS

The promising application of adiabatically reduced equations to the simulation of the low-frequency dynamics of the magnetized plasma is directly confirmed by the numerical calculation results. The self-consistent convection of the plasma and convective enhancement of the transport processes are simulated using these equations as an evolution problem with the given initial and boundary conditions. The numerical code provides the realization of various regimes of convection and plasma transfer processes by varying the conditions of the heating, power takeoff at the separatrix, penetration of impurities, maintenance of radially inhomogeneous (shear) plasma flows, and other controlling parameters. The dimensionless time is introduced taking into account the characteristic size (small plasma radius) and the expected velocity scale for nonlinear convection (2) and is specified by the relation $\epsilon c_s t/a \longrightarrow t$, where the small parameter ϵ is given by Eq. (1).

In the model magnetic configuration under consideration, there is no strong toroidal magnetic field. For this reason, the results can be used with care to quantitatively compare with tokamak data. Nevertheless, the characteristic times of a number of simulated processes are quite close to the times of the processes observed in real experiments, in particular, at tokamaks. For the convenient comparison of the calculated times with experimental data, we note that the unit of dimensionless time is about 60 μs for a device of the T-10 scale (i.e., for a device with the same characteristic sizes and poloidal magnetic field, density, and temperature).

The total dimensionless calculation time t_{tot} usually exceeds the energy lifetime of the plasma and ranges from 50 to 500 for various calculation variants. In the absence of the initial high-power shear flows, the initial linear and quasilinear stages of MHD instability are completed at $t = 5$. Then, convective flows and entropy fluctuations are transformed to the stage of strongly nonlinear rearrangement. This stage is very irregular and leads to the formation of wide frequency and wavenumber spectra. At $t \approx 20$, the plasma enters the quasistationary stage at which the S_0 profile, as well as the spectral widths, remains almost unchanged. At the same time, the convective flows acquire a certain characteristic large-scale vortex structure consisting of spontaneously appearing and disappearing vortex pairs each of which exists during several dimensionless time units and has a quite high level of stochasticity of local velocities. Such behavior likely indicates that the dynamic system passes to the strange-attractor regime [23]. Indeed, in this case, there are all the necessary conditions for realizing such a regime; namely, we have an unclosed strongly nonlinear dynamic system with the number of freedom more than two with weak dissipation and the energy flux from the heating region to the periphery. At the same time, more detailed analysis (including statistical analysis) shows that the properties of the fluctuations in the velocity, temperature, and density of the plasma in this regime are similar to those observed in many experiments and are characterized as structure turbulence [5, 6, 9]. Below, we discuss these properties in more detail.

In Fig. 3, which characterizes the quasistationary stage at $t = 41.0$, the typical two-dimensional structures of convective flows and entropy fluctuations are shown in the (x, φ) plane, where $x = r/r_m$ is the dimensionless small radius and r_m corresponds to a point near which the heating source Q_E in Eq. (5) is localized. The region boundaries $x = 0.5$ and 2.0 correspond to the surface of the current-carrying rod and separatrix, respectively. Figure 3a shows $\Phi(t, x, \varphi) = \text{const}$ contours or, equivalently, streamlines. The light and dark regions correspond to positive and negative potential values, respectively. In the general structure of flows, one or two long-lived pairs of coupled large-scale vortices with strictly

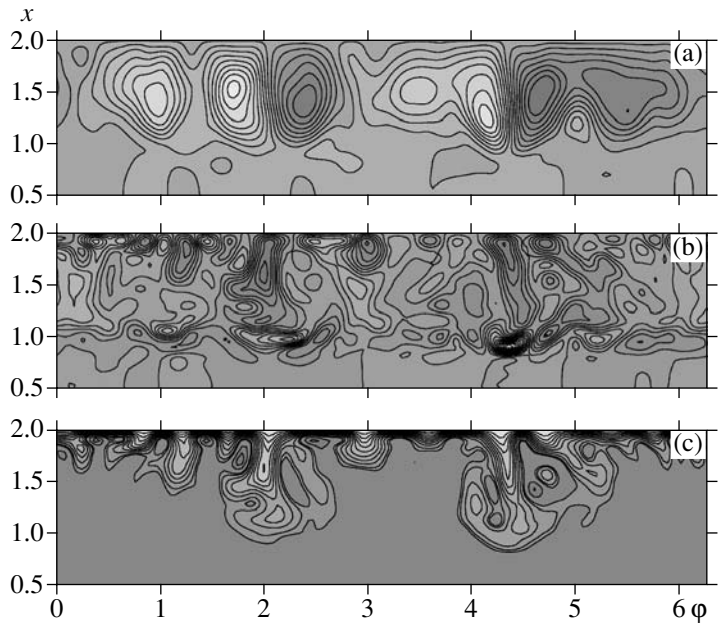


Fig. 3. Typical 2D structures of vortex convection: (a) $\Phi(t, x, \varphi)$, (b) $\tilde{S}(t, x, \varphi)$, and (c) $\lambda_*(t, x, \varphi)$.

determined polarity dominate. Vortices are almost completely localized in the region $1 < x < 2$, where $S_0' \approx 0$, and evolve very complexly and irregularly, changing their intensity and shape, as well as drifting along φ. The presence of quite fast plasma jets, which are localized in φ (φ ≈ 2.1 and 4.4 in Fig. 3a) and directed from the separatrix to the plasma center, is a feature of these vortex pairs. In their properties, these jets are very similar to the objects that are called streamers in many experiments. Figure 3b shows the entropy fluctuation contours $\tilde{S}(t, x, \varphi) = \text{const}$. Figure 3c illustrates an early stage of the fast penetration of the "passive impurity" from the plasma periphery to the central area. The impurity source was turned on near the separatrix ($x = x_s = 2$) at $t = 40.0$. It is seen that the jets give rise to the rapid (in time $\Delta t \sim 1$) injection of the impurity to the plasma core. A similar effect of the fast penetration of the impurity to the central area of a tokamak is observed for many devices. In particular, anomalously fast penetration of a tritium impurity was observed at the ASDEX device [24].

Figure 4 shows the radial profiles of the quasi-stationary parameters at the developed-turbulence stage ($t = 100$). The profile $S_0(x, t)$ (the solid line in Fig. 4a) exhibits a plateau in the intense-convection region $1 < x < 2$, which ensures the expected closeness to the marginally stable state. The maintenance of such a plateau (the deviation from $S_0 = \text{const}$ is less than 3%) is observed in all the time intervals including transient regimes. The latter property means a certain self-consistency of the pressure and temperature profiles in the convective region. A similar self-consistency of the profiles was observed in many experiments (see, e.g., [4]). At a fixed energy source, the stationary level S_0 in the plateau region and the corresponding plasma lifetime depend only on the coefficient ν in the boundary condition for the heat flux, which characterizes the energy loss intensity at the periphery [20, 21]. Following the traditional interpretation of experimental data on the anomalous transport, one can introduce the effective (anomalous) local thermal diffusivity $\chi_{\text{eff}} = -q/\rho T'$, where $q(x, t)$ and $T(x, t)$ are the heat-flux density and temperature, respectively, as well as the "anomalous factor," which defines as the ratio of the effective thermal diffusivity to the classical thermal diffusivity: $F_a = \chi_{\text{eff}}/\chi_{\text{cl}}$. The anomalous factor is shown by the dotted line in Fig. 4a. In the range $0.5 < x < 0.9$, $F_a \approx 1$, which indicates that the heat transfer in this range is classical. The sharp spikes near the $x = 1.1$ magnetic surface correspond to the condition of vanishing $q(x, t)$ and $T'(x, t)$ on different magnetic surfaces, because the heat-transfer process is nondiffusive. In the region of intense convection for $x > 1.2$, the factor F_a first increases up to its maximum (depending on ν) and then decreases to unity near the plasma boundary. A similar effect of the sharp decrease in F_a near the plasma boundary, which was called the "external transport barrier," was first observed at the ASDEX device [25] and then in many other experiments at tokamaks.

Figure 4b illustrates the radial density distribution of the "passive impurity." The dotted line is the φ-averaged impurity-particle number $\lambda_*(x, t)$ in the magnetic flux tube, which [similar to $S_0(x, t)$] has a plateau in the intense-convection region. It is essential that the corre-

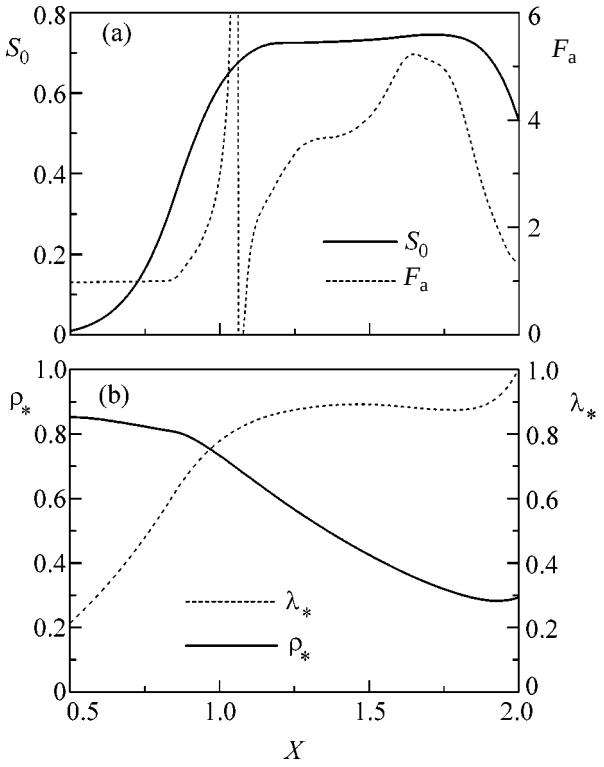


Fig. 4. Radial profiles of the (a) entropy and anomaly factor and the (b) averaged impurity density.

sponding impurity density profile $\rho_*(x, t)$ increases in the direction of the plasma center rather than in the opposite direction as is predicted by the diffusive-transport model. Such an effect in experiments is usually called "impurity pinch." Since the ion flux of the main plasma component can be organized similarly, the above mechanism can be considered as a promising method for maintaining the necessary plasma density in the central region.

Figure 5 shows the time evolution of the anomalous factor F_a at three radii in the calculation variant, where the energy absorption coefficient ν at the plasma edge is halved at time $t = 20$. This figure shows that the anomalous factor F_a also decreases noticeably over the entire convection region in time on the order of three or four dimensionless time units, whereas the relative temperature gradients in the main plasma bulk remained unchanged. Such a behavior of the effective transport coefficients is similar to that in the *L–H* transitions in tokamaks. The time of changing the anomalous factor F_a for the T-10-scale device is estimated as about 200 μs, which quite well agrees with the characteristic time of the *L–H* transitions observed in the experiment [4].

The presence of shear or zonal flows in the plasma can noticeably affect the development of large-scale vortex convection. The simulation results show that the shear flows with sufficiently large vorticity can temporarily suppress the flute mode and nonlinear broadening of the spatial perturbation spectrum at the initial evolution stage, leading to considerable delay in the plasma transition to the quasi-stationary turbulent state. However, the further nonlinear evolution of convective flows is similar to the evolution in the absence of initial zonal flows and leads to the formation of intense large-scale stochastic vortex structures (stochastic convective cells) with the corresponding increase in the resulting fluxes of particles and heat.

Figure 6 shows the evolution of heat fluxes to the outer separatrix in three different regimes. Line *1* corresponds to the above-discussed regime with relatively weak shear flows, when the integral dimensionless vorticity W_0 flows [which, according to Eq. (6), is an integral of motion of the system] are equal to zero. Line *2* corresponding to the large vorticity of the shear flows ($W_0 = 10$) exhibits a considerable delay in the turbulence development and formation of the resulting anomalous heat flux. In this regime, the transition to developed turbulence occurs after the rearrangement and smoothing of the initial profile of the dynamic vor-

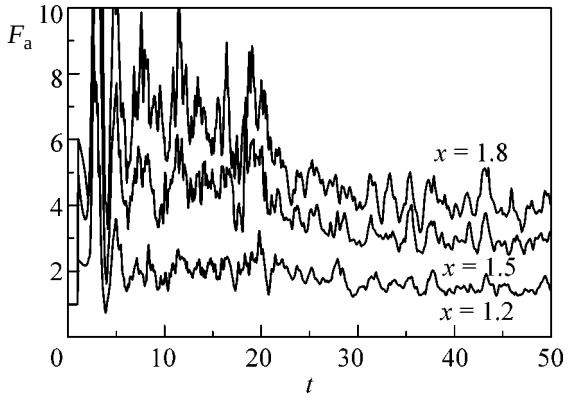


Fig. 5. Evolution of the anomaly factor at three radii.

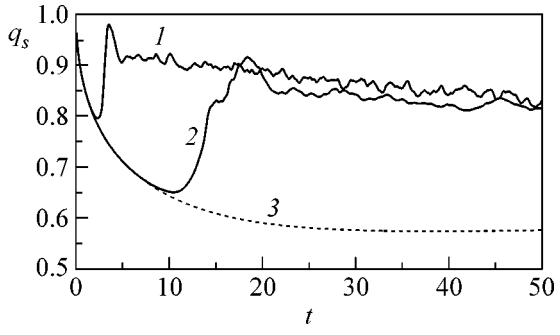


Fig. 6. Evolution of heat fluxes to the outer separatrix in the regimes of (*1*) weak shear flows, (*2*) strong shear flows, and (*3*) without fluctuations.

ticity of the shear flows. The delay time depends on both the integral dynamic vorticity and its initial profile. Line *3* is shown for comparison and corresponds to the hypothetical regime, where any fluctuations are absent (i.e., the regime with the purely collisional diffusive heat flux).

Time-averaged spatial spectra of fluctuations (more strictly, spectra in toroidal wavenumbers n) at the developed-turbulence stage are shown in Fig. 7. Figure 7a shows the spectra (in arbitrary units) of fluctuations in the electric potential for the regimes with strong (W_0 = 10) and weak (W_0 = 0) shear flows. The shape of the spectra is independent of the structure of the initial perturbations. The spectra differ most strongly for $n \le 3$, which is associated with a certain difference of the spatial form of the dominating large-scale vortex structures in the indicated regimes. For $n \ge 4$, the spectra in both regimes exhibit a characteristic power decrease with approximately identical slops. The increase in the slope of the spectra for $n > 10$ (as is expected) is associated with the dissipative suppression of small-scale fluctuations. As the dissipation decreases, this effect is manifested for large n values. The corresponding spectra of fluctuations in the entropy function are shown in Fig. 7b.

Figure 8 shows an example of the characteristic time realization of fluctuations in the (a) radial velocity and (b) entropy function at a fixed point in the center of the convective region. These fluctuations, which evidently are of stochastic character, correspond to developed turbulence in the regime with the high vorticity of the shear flow. In fluctuations of both the radial velocity and entropy function, high spikes (so-called bursts) are observed, whose value is much higher than the statistically average level of the fluctuations. Analysis of the time evolution of flows enables one to associate the appearance of bursts in the radial velocity with the drift of the "streamer" through the observation point.

The frequency spectra of the appearing turbulent convection, as well as its statistical properties, are very similar to those observed in many experiments (including experiments at tokamaks and stellarators) and, as was mentioned in the Introduction, are associated with the development of structure turbulence. For the regime with the high vorticity of the shear flow, Fig. 9a shows the frequency spectrum of fluctuations in the potential, where individual narrow peaks exist along with the broad stochastically irregular spectral component (broad band). We emphasize that, in contrast to the traditional interpretation of the experimental results, these peaks cannot be attributed to drift instabilities, because such instabilities are absent in the dynamic model under consideration. Moreover, owing to the strong nonlinearity of the vortex convective flows, the time realizations of the fluctuations contain almost no information on the frequency characteristics of the feeding linear instabilities. At the same time, analysis of the time evolution of the flows provides another natural

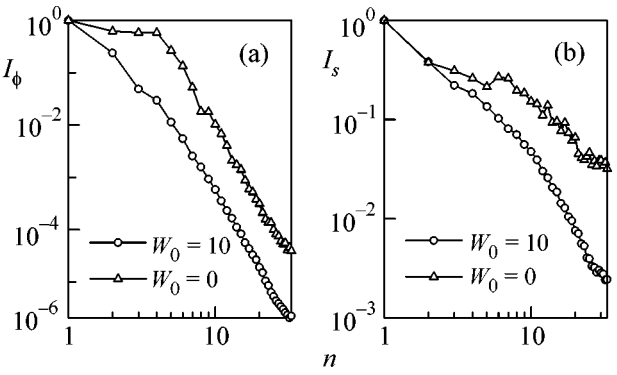


Fig. 7. Toroidal-wavenumber spectra of fluctuations in (a) potential and (b) entropy.

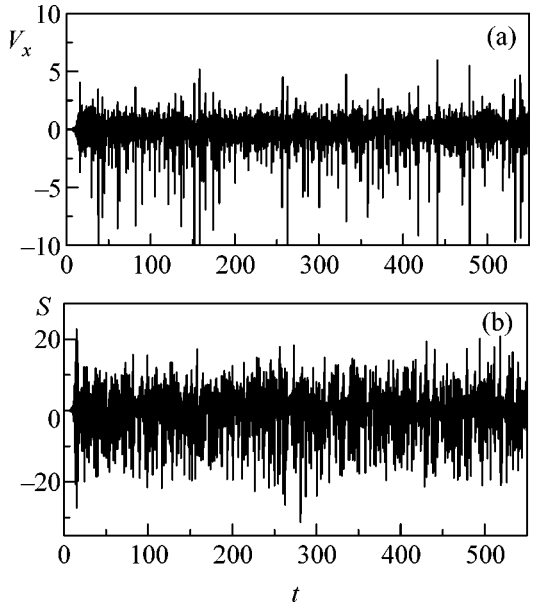


Fig. 8. Fluctuations of the (a) radial velocity and (b) entropy function in the convective-vortex region.

interpretation of the peaks in the potential fluctuation spectrum. In particular, the main peak in Fig. 9a with the dimensionless frequency $\omega = 8.2$ corresponds to the drift of the dominant large-scale vortex structure through the observation point and the peak next in height with the dimensionless frequency $\omega = 1.9$ corresponds to the above bursts. This circumstance gives rise to certain doubts in the correctness of the traditional interpretation of the experimental results according to which the narrow peaks in the spectrum are associated with drift instabilities. In Fig. 9b, where the same spectrum is shown in the log–log scale, it is seen that the basic changes in the dominate large-scale vortex structures are characterized by the frequency range $1 < \omega < 10$. At higher frequencies, the spectrum decreases

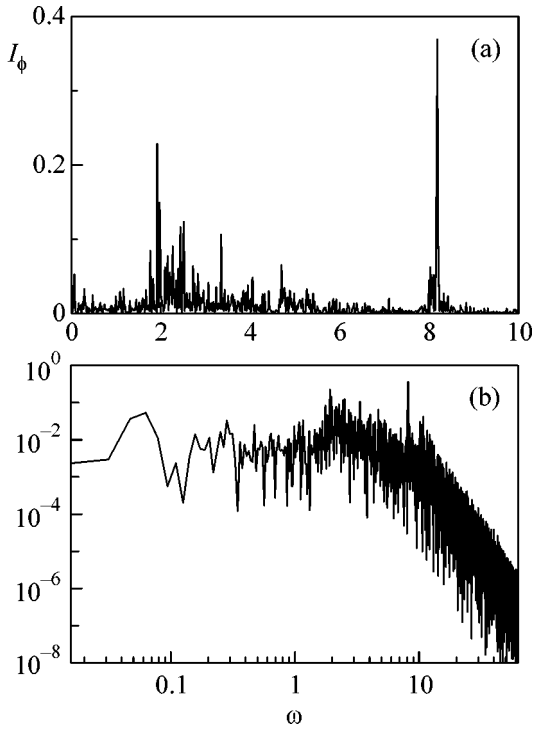


Fig. 9. Frequency spectrum of fluctuations in the potential in the (a) linear and (b) log–log scales.

according to a power law on average, which is characteristic of strong turbulence.

The probability density distribution functions for the increments of fluctuations in the potential, entropy, and radial velocity that are calculated with various time lags in this model are similar to those obtained by processing the experimental data. As in most experiments (see [5, 6, 9]), the probability distribution functions for a short time lag strongly differ from the Gaussian distribution, have heavy tails, and are transformed to the Gaussian distribution for a sufficiently large lag. In these calculations, such a transition occurs at a lag of two or three dimensionless time units, which approximately corresponds to the time of the rotation of the dominant vortex structure. After the evaluation for the parameters of the T-10 tokamak, it is equal to 120–180 μs, which agrees well with the experimental results.

4. CONCLUSIONS

The above results confirm the efficiency of the approach based on the self-consistent dynamic simulation of low-frequency turbulence and the resulting transport processes in the magnetized plasma. By an example of a relatively simple magnetic configuration, it has been shown that the substantial nonlinear low-frequency dynamics of the plasma can be quite adequately simulated by means of adiabatically reduced hydrodynamic-type equations. Equations obtained by the method of the adiabatic separation of fast and slow motions that was developed in [18–20, 22] hold the invariant properties of the initial equations and thereby can be used to simulate the plasma evolution in sufficiently long time intervals exceeding the plasma lifetime.

For the magnetic configuration under consideration, as well as for other magnetic systems with closed or open field lines, very meaningful physical results can be obtained even in the simplest one-fluid MHD model with isotropic plasma pressure. In this case, the method of the adiabatic separation of motions enables one to exclude three classes of relatively high-frequency stable oscillations: magnetoacoustic, Alfvén, and longitudinal acoustic modes. The resulting reduced equations describe slower nonlinear 2D convection, which is self-consistently developed near the threshold of the ideal flute MHD instability.

In the above-considered example, convection is generated and maintained due to the competition between two processes: plasma heating and local dissipative processes that distort the initial marginally stable profile of the plasma pressure, making it weakly unstable and leading to the growth of flute perturbations, which then recover the marginally stable pressure profile and initiate the substantially nonlocal anomalous transport of particles and energy. Strictly speaking, other mechanisms of the generation and maintenance of low-frequency convection in the magnetized plasma are possible. However, in this case, it might also be expected that nonlinearity leads to the formation of large-scale vortex structures. The adiabatically reduced equations allow the self-consistent description of both relatively fast nonlinear convective flows and slower transport processes in the weakly dissipative plasma. As a result, the description of both dynamic and transport processes appears to be completely self-consistent and does not require the inclusion of any additional assumptions.

Computer simulation shows that the convection at the developed stage is substantially nonlinear and stochastic and has broad frequency and wavenumber spectra. A finite number of large-scale long-lived vortices (stochastic convective cells) dominate in its 2D structure. Convection maintains a certain quasistationary turbulent-relaxed plasma state, which is characterized by the existence of the marginally stable plateau in the radial profiles of the renormalized densities of the plasma components. The tendency to the maintenance of such plateaus can be treated as the self-consistency of profiles in the convection region. The formation of a similar plateau in the impurity distribution can be responsible for an effect called impurity pinch.

Convection leads to increased heat transfer whose magnitude is determined primarily by energy losses at the plasma periphery rather than by the gradients of the plasma parameters in the main confinement region. This effect is similar to *L–H* transitions in tokamaks. Similar to experiment, the heat flux is quite rapidly syn-

chronized with the change in the energy losses at the plasma periphery.

The numerically simulated frequency spectra of fluctuations in the electric potential, pressure, and density of the plasma contain both a broad stochastically irregular spectral component (broad band) and individual narrow peaks, which, in contrast to the traditional interpretation of the experimental spectra, are attributed to the drift of dominant vortex structures rather than to drift instabilities. This circumstance could lead to the revision of the traditional interpretation of the experimental spectra.

The statistical properties of the calculated fluctuations are also very similar to those observed in various magnetic systems. In particular, the probability distribution functions of increments of fluctuations that are obtained by processing the calculation results are non-Gaussian with heavy tails. On the whole, it might be expected that the dynamic simulation of low-frequency turbulence and the resulting transport processes will enable one to better understand the transport processes in various magnetic systems, as well as make a noticeable contribution to the development of the general theory of nondiffusive transport processes in magnetized plasma.

This work was supported in part by the State Contract (project no. RI-112/001/144) and by the Department of Atomic Science and Engineering, the State Agency for Atomic Energy of the Russian Federation.

REFERENCES

1. B. B. Kadomtsev and O. P. Pogutse, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1967), No. 5, p. 209.
2. W. Horton, in *Basic Plasma Physics*, Ed. by A. A. Galeev and R. N. Sudan (Énergoatomizdat, Moscow, 1984; North-Holland, Amsterdam, 1984), Vol. 2, p. 362.
3. J. W. Connor, P. Buratti, J. D. Cordey, *et al.*, Plasma Phys. Controlled Fusion **41**, 693 (1999).
4. M. V. Ossipenko and T-10 Team, in *Proceedings of 19th IAEA Fusion Energy Conference, Lyon, 2002* (IAEA, Vienna, 2003), Rep. OV/5-2.
5. G. M. Batanov, L. V. Kolik, A. E. Petrov, *et al.*, Fiz. Plazmy **29**, 395 (2003) [Plasma Phys. Rep. **29**, 363 (2003)].
6. G. M. Batanov, V. E. Bening, V. Yu. Korolev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 974 (2003) [JETP Lett. **78**, 502 (2003)].
7. T. Cho, M. Yoshida, H. Higaki, *et al.*, J. Plasma Fusion Res. **80**, 81 (2004).
8. T. Cho, M. Yoshida, J. Kohagura, *et al.*, Phys. Rev. Lett. **94**, 085002 (2005).
9. V. V. Abrakov, A. E. Petrov, K. A. Sarksyan, and N. N. Skvortsova, Fiz. Plazmy **20**, 1069 (1994) [Plasma Phys. Rep. **20**, 959 (1994)].
10. G. Y. Antar, S. I. Krasheninnikov, P. Devynck, *et al.*, Phys. Rev. Lett. **87**, 065001 (2001).
11. N. Ohno, D. Nishijima, S. Takamura, *et al.*, Nucl. Fusion **41**, 1055 (2001).
12. J. A. Boedo, D. L. Rudakov, R. Moyer, *et al.*, Phys. Plasmas **8**, 4826 (2001).
13. S. D. Danilov and D. Gurarie, Usp. Fiz. Nauk **170**, 921 (2000) [Phys. Usp. **43**, 863 (2000)].
14. B. B. Kadomtsev and O. P. Pogutse, Zh. Éksp. Teor. Fiz. **65**, 575 (1973) [Sov. Phys. JETP **38**, 283 (1973)].
15. R. White, D. Monticello, M. N. Rosenbluth, *et al.*, in *Proceedings of 5th International Conference on Plasma Physics and Controlled Nuclear Fusion Research, Tokyo, 1974* (IAEA, Vienna, 1975), Vol. 1, p. 495.
16. H. R. Strauss, Phys. Fluids **19**, 134 (1976).
17. H. R. Strauss, Phys. Fluids **20**, 1354 (1977).
18. V. P. Pastukhov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 892 (1998) [JETP Lett. **67**, 940 (1998)].
19. V. P. Pastukhov, Fiz. Plazmy **26**, 566 (2000) [Plasma Phys. Rep. **26**, 529 (2000)].
20. V. P. Pastukhov and N. V. Chudin, Fiz. Plazmy **27**, 963 (2001) [Plasma Phys. Rep. **27**, 907 (2001)].
21. V. P. Pastukhov and N. V. Chudin, in *Proceedings of 19th IAEA Fusion Energy Conference, Lyon, 2002* (IAEA, Vienna, 2003), Rep. TH/2-5.
22. V. P. Pastukhov, Fiz. Plazmy **31**, 628 (2005) [Plasma Phys. Rep. **31**, 577 (2005)].
23. M. I. Rabinovich, Usp. Fiz. Nauk **125**, 123 (1978) [Sov. Phys. Usp. **21**, 443 (1978)].
24. K.-D. Zastrow, J. M. Adams, Yu. Baranov, *et al.*, Plasma Phys. Controlled Fusion **49**, B255 (2004).
25. F. Wagner, G. Becker, K. Behringer, *et al.*, Phys. Rev. Lett. **49**, 1408 (1982).

Translated by R. Tyapaev

JETP Letters, Vol. 82, No. 6, 2005, pp. 366–370. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 407–411.

Vibrational Enhancement of the Effective Donor–Acceptor Coupling¶

M. Lazrek[a, b], D. J. Bicout[a, c], S. Jaziri[b], and E. Kats[a, d, *]

[a] *Institut Laue-Langevin, BP 156, Grenoble, France*

[b] *Laboratoire de Physique de la Matière Condensée, Faculté des Sciences de Bizerte, 7021 Jarzouna Bizerte, Tunisia*

[c] *Biomathematics and Epidemiology, ENVL–TIMC, 69280 Marcy l'Etoile, France*

[d] *Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117940 Russia*

e-mail: kats@ill.fr

Received May 31, 2005

This letter deals with a simple three-site model for charge transfer phenomena in a one-dimensional donor (D)–bridge (B)–acceptor (A) system coupled with vibrational dynamics of the B site. It is found that, in a certain range of parameters, the vibrational coupling leads to an enhancement of the effective donor–acceptor electronic coupling as a result of the formation of the polaron on the B site. This enhancement of the charge transfer efficiency is maximum at the resonance, where the effective energy of the fluctuating B site coincides with the donor (acceptor) energy. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 34.70.+e, 73.40.Gk, 82.39.Jn

Molecular electronics is progressing so rapidly that it is now possible to perform measurements and assembly at the level of individual or a few molecules [1, 2]. Charge transport is known to occur in a wide range of linear chain molecules including DNA double strand molecules. For DNA, it is believed that the charge transport phenomenon is involved in the protection of the DNA encoded information against oxidative damage [3]. As the DNA molecule is essentially a dynamic structure on the time scale of charge transport, one would expect vibrational dynamics to play an important role in DNA electronics and, in general, for any property of biological molecules, because the biological functions of life are associated with molecular motion not a static or dead structure (i.e., the equilibrium positions of all the atoms).

In this letter, we are interested in a one-dimensional DNA wire or bridge (B) connecting a donor (D) and an acceptor (A) site. Usually, the bridge consists of N sites with one state per site (see the abundant literature devoted to this issue in [4–23]). The theoretical analysis of this problem requires solving a system of $N + 2$ nonlinear coupled equations. Unfortunately, such a problem cannot be solved analytically for $N \gg 1$, and we have to recourse to numerical solution. However, many insights and essential features of the dynamics can already be gained by studying three simple sites: D–B–A. Generally speaking, the interaction between the donor and acceptor involves all states of the bridging subsystem. This bridge mediated interaction can be characterized, under certain conditions, by a single energy-dependent parameter, effective coupling, which plays the key role in the charge transfer. For small systems, the phase coherence of the charges is maintained over the entire system and the quantum effects are crucial in determining the system properties. In contrast, for long $N \gg 1$ bridges, the fast relaxation processes result in a strong dephasing between the charges in the system. Therefore, this leads to a rapid falloff of the off-diagonal elements of the density matrix such that the diagonal elements can be described by a set of kinetic equations [17, 24].

Our concern in this paper is to investigate a Hamiltonian model describing the D–A coupling in the presence of dynamic structural fluctuations. Such local fluctuations, including local vibrations, twisting motions, radial deformations, and hydrogen-bond stretching or opening, are known to strongly influence charge transfer in DNA molecules [25–30]. For simplicity, we consider a three-site D–B–A system where the electronic degree of freedom is coupled to an effective local vibrational degree of freedom.

Let us assume that, initially, the charge is entirely localized on the donor site with energy ϵ. Then, owing to the nonzero overlapping integrals of the electronic wave functions between the two neighboring sites, the tunneling of the charge takes place from the donor to the acceptor site with the same energy. Denoting by $\{|d\rangle, |b\rangle, |a\rangle\}$ the localized states on the donor, bridge, and acceptor, respectively, the Hamiltonian of the

¶ The text was submitted by the authors in English.

* A member of the editorial board of the journal *JETP Letters* from 1991 to 1997.

bridge-mediated charge transfer between the donor and acceptor acquires the form

$$H = H_e + \frac{p^2}{2m} + \frac{m\omega^2 r^2}{2} + kr|b\rangle\langle b|, \quad (1)$$

in which the bare electronic part reads as

$$H_e = \varepsilon[|d\rangle\langle d| + |a\rangle\langle a|] + \varepsilon_b|b\rangle\langle b| + V_{db}[|d\rangle\langle b| + |b\rangle\langle d|] + V_{ba}[|b\rangle\langle a| + |a\rangle\langle b|], \quad (2)$$

where ϵ is the one-site energy of the donor and acceptor, ϵ_b is the one-site energy of the bridge, m ($m \simeq 300$ amu) is the mass of the bridge base pair, r is its radial displacement in the localized vibrational mode with the frequency ω, the momentum p is conjugated to r, and k is the electron-localized vibration mode coupling constant. The localized bridge mode can be treated classically, since the corresponding vibrational displacement amplitude is larger than the zero-point quantum fluctuations for the characteristic DNA parameters [25–28] (see also the following).

The frequency of the typical vibrations in DNA is $\omega \simeq 10^{11}$–10^{12} s^{-1}, and the scale of the electronic overlap integrals between the base pairs in DNA is $V = \sqrt{2(V_{db}^2 + V_{ba}^2)} \simeq 0.2$ eV, which leads to an electronic characteristic frequency of $V/\hbar \simeq 3 \times 10^{14}$ s^{-1}. As a consequence of the small (adiabatic) parameter, $V\omega/\hbar \ll 1$, the slow vibrational and fast electronic motions can be decoupled. Therefore, to solve the problem of the bridge-mediated charge transfer between the donor and the acceptor, we employ the adiabatic procedure to eliminate the slow vibrational motions and derive an effective Hamiltonian for the fast electronic motions. To proceed, we take the wave function of the charge in the form $|\Psi(t)\rangle = c_d(t)|d\rangle + c_b(t)|b\rangle + c_a(t)|a\rangle$, where $c_n(t)$ are the time-dependent amplitudes of the probability of the charge being at the nth site. From the Hamiltonian (1), we arrive at the following equations of motion for the quantum amplitude $c_n(t)$,

$$i\hbar\frac{d}{dt}\begin{pmatrix} c_d \\ c_b \\ c_a \end{pmatrix} = \begin{pmatrix} \varepsilon & V_{db} & 0 \\ V_{db} & \varepsilon_b + kr & V_{ba} \\ 0 & V_{ba} & \varepsilon \end{pmatrix}\begin{pmatrix} c_d \\ c_b \\ c_a \end{pmatrix}, \quad (3)$$

and for the classical dynamic mode,

$$m\frac{d^2 r}{dt^2} = -m\omega^2 r - k|c_b|^2. \quad (4)$$

Next, we seek stationary solutions of the form $c_n(t) = c_n e^{-iEt/\hbar}$ oscillating with the frequency $E/\hbar$. To work with dimensionless quantities, we use, from now on, the dimensionless variables u, σ, and κ defined in the table. Using $c_n(t)$ in the equations of motions, we find

Definition of dimensionless variables

Definition	Variable
Energy scale	$V = \sqrt{2(V_{db}^2 + V_{ba}^2)}$
Length scale	$\xi = \sqrt{V/m\omega^2}$
Coupling asymmetry	$\eta = V_{db}/V_{ba}$
Dimensionless polaron energy	$u = (E - \varepsilon)/V$
Dimensionless energy barrier	$\sigma = (\varepsilon_b - \varepsilon)/V$
Reduced electron-vibration coupling	$\kappa = k\xi/V$

that the stationary bridge displacement is $r = -\kappa c_b^2$, and the stationary probability distribution is

$$c_d^2 = \eta^2 c_a^2, \quad c_b^2 = 2u^2/(1 + 2u^2), \quad c_a^2 = 1/[(1 + \eta^2)(1 + 2u^2)], \quad (5)$$

where the root u satisfies the characteristic equation

$$4u^4 + 4(\kappa^2 - \sigma)u^3 - 2\sigma u - 1 = 0. \quad (6)$$

Solution of this equation provides the ground u_g plus one or three (depending upon the values of κ and σ) excited energies of the "polaron," i.e., the state created by the charge coupling with the DNA structural deformation. These stationary polaron solutions allow us to eliminate from (1) the structural deformation r and to obtain the effective Hamiltonian for the charge transfer as

$$H_{\text{eff}} = \frac{1}{V}(H_e - \varepsilon\hat{1}) = \Delta(u)|b\rangle\langle b| + \frac{\eta}{\sqrt{2(1+\eta^2)}}[|d\rangle\langle b| + |b\rangle\langle d|] + \frac{1}{\sqrt{2(1+\eta^2)}}[|b\rangle\langle a| + |a\rangle\langle b|], \quad (7)$$

where $\Delta(u) = \sigma - \kappa^2 c_b^2(u)$ is the renormalized effective energy of the bridge due to electron-vibration coupling and $\hat{1}$ is the unit matrix. To calculate the effective charge coupling between the donor and the acceptor, we first have to solve the time-dependent problem to determine the probability of charge transfer defined as $P_{d\to a}(t) = |\langle a|\Psi(t)\rangle|^2$, where $|\Psi(t)\rangle$ is the solution of the Schrödinger equation

$$i\hbar\frac{d|\Psi(t)\rangle}{dt} = H_{\text{eff}}|\Psi(t)\rangle,$$

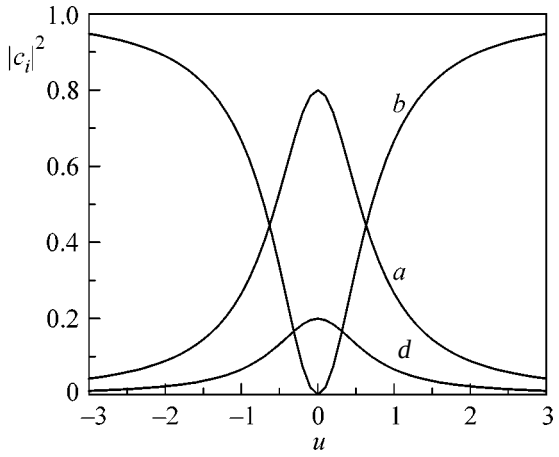


Fig. 1. Stationary charge density in Eq. (5) versus the energy for the coupling asymmetry $\eta = 0.5$. The quoted letters *d*, *b*, and *a* stand for the donor, bridge, and acceptor, respectively.

with the initial condition $|\Psi(0)\rangle = |d\rangle$. It is easy to show that, when $t \longrightarrow \infty$ we have $P_{d\to a}(t) \simeq k_{da}t$, where the charge transfer rate k_{da} is given by the Fermi golden rule

$$k_{da} = \frac{2\pi}{\hbar}|H_{da}|^2\begin{cases}\delta(E_2(u)-E_1(u)); & \sigma \geq \kappa^2 c_b^2\\ \delta(E_3(u)-E_1(u)); & \sigma \leq \kappa^2 c_b^2\end{cases}, \quad (8)$$

with the (dimensionless) effective donor–acceptor coupling [18]

$$|H_{da}|^2 = \frac{\eta^2}{4(1+\eta^2)^2} \times\begin{cases}1/[(E_3(u)-E_1(u))(E_3(u)-E_2(u))]; & \sigma \geq \kappa^2 c_b^2\\ 1/[(E_1(u)-E_2(u))(E_3(u)-E_2(u))]; & \sigma \leq \kappa^2 c_b^2\end{cases}; \quad (9)$$

where $E_i(u)$ are the eigen energies of H_{eff} given by $E_1(u) = 0$ and $2E_{2,3}(u) = \Delta \mp \sqrt{\Delta^2+2}$. Finally, we end up with the effective D–A coupling given by

$$|H_{da}(u,\eta,\sigma,\kappa)|^2 = \frac{\eta^2}{2(1+\eta^2)^2\sqrt{\Delta^2+2}} \times\frac{1}{[|\Delta|+\sqrt{\Delta^2+2}]}, \quad (10)$$

and the ratio ρ, thus, allowing us to measure the effect of vibrations on the D–A coupling, which reads

$$\rho(u,\sigma,\kappa) = \frac{|H_{da}(u,\eta,\sigma,\kappa)|^2}{|H_{da}(u,\eta,\sigma,0)|^2} = \left[\frac{|\sigma|+\sqrt{\sigma^2+2}}{|\Delta|+\sqrt{\Delta^2+2}}\right]\left[\frac{\sigma^2+2}{\Delta^2+2}\right]^{1/2}. \quad (11)$$

Expressions (10) and (11) are the main results of this paper. They provide close formulas for evaluating how the dynamical disorder affects the effective donor–acceptor coupling in various situations. As direct applications of our main findings, we consider the following illustrative examples.

Charge density versus polaron energy. It results from Eq. (5) that the charge densities on the donor, bridge, and acceptor are even functions of the polaron energy *u*. For all *u*, the ratio of the density of the donor to that of the acceptor is equal to the square of the asymmetry energy η (see the table). For $u = 0$, there is no charge on the bridge site, and the charge density is distributed between the donor and acceptor sites in proportion to η. In contrast, for the limits of very high (or low) polaron energy, when $|u|$ gets larger, the charge density decreases considerably on the donor and acceptor sites, while it gets higher on the bridge site, hence, leading to small charge transfer efficiency. These features are illustrated in Fig. 1 for the energy asymmetry parameter $\eta = 0.5$. Likewise, at the resonance u_m, where the renormalized effective energy of the bridge is equal to zero (the donor/acceptor energy),

$$\Delta(u_m) = 0 \Rightarrow u_m = \pm\frac{\sqrt{\sigma}}{\sqrt{2(\kappa^2-\sigma)}}, \quad (12)$$

and the distribution of the charge density reduces to

$$c_d^2(u_m) = \eta^2 c_a^2(u_m), \quad c_b^2(u_m) = \sigma/\kappa^2, \quad c_a^2(u_m) = [1-c_b^2(u_m)]/(1+\eta^2). \quad (13)$$

At this resonance point, the charge density on the bridge site decreases either upon approaching the bare resonance for $\sigma \longrightarrow 0$ or by increasing the electron-vibration coupling parameter κ above $\sqrt{\sigma}$.

Effective D–A coupling versus *u*. The effective D–A coupling scales as $|H_{da}| \sim 1/\Delta$ for $\Delta \gg 1$. However, at the resonance $\Delta = 0$ defined in Eq. (12), the $|H_{da}(u, \eta, \sigma, \kappa)|^2$ and, thus, the ratio $\rho(u, \sigma, \kappa)$ attain their maxima given by

$$|H_{da}(u_m,\sigma,\kappa)|^2 = \frac{\eta^2}{4(1+\eta^2)^2}, \quad (14)$$

and

$$\rho(u_m,\sigma,\kappa) = \frac{(\sigma+\sqrt{\sigma^2+2})\sqrt{\sigma^2+2}}{2}. \quad (15)$$

Simple inspection of this equation shows two characteristic features. First, at the resonance, the effective D–A coupling is enhanced by the coupling to the structural dynamics, and, second, as is illustrated in Fig. 2, the enhancement factor ρ of the effective D–A coupling due to the electron-vibration interactions increases with

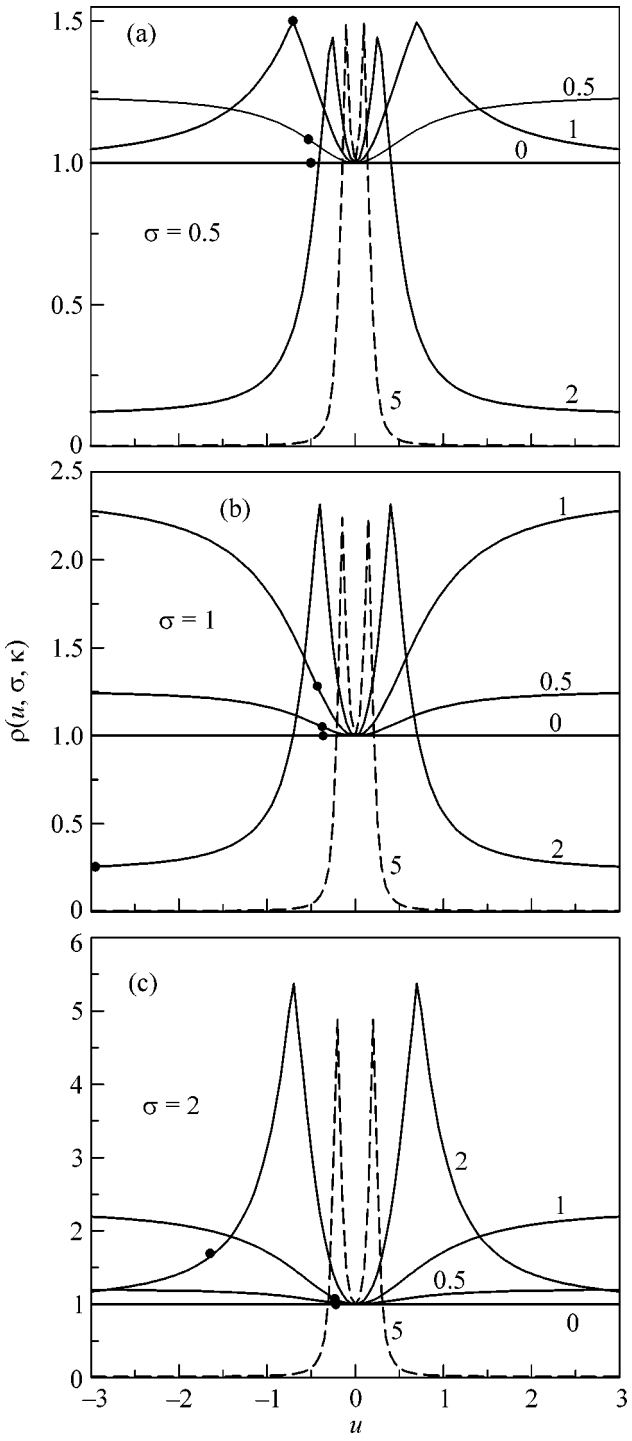


Fig. 2. Effective coupling ratio in Eq. (11) as a function of the energy. Filled circles correspond to $\rho(u_g, \sigma, \kappa)$ at the ground-state energy u_g. The quoted numbers correspond to the electron-vibration coupling values, i.e., $\kappa = 0, 0.5, 1, 2, 5$.

the energy barrier σ. To rationalize these observations in terms of the polaron energy and electron-vibration coupling, we have depicted in Fig. 2 the ratio $\rho(u, \sigma, \kappa)$ as a function of u for increasing values of σ and κ. Two different regimes can be distinguished:

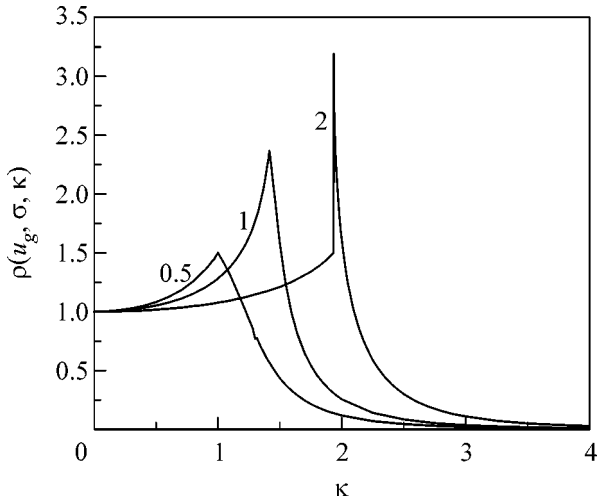


Fig. 3. Effective coupling ratio in Eq. (11) at the ground-state energy u_g as a function of the electron-vibration coupling. The quoted numbers correspond to the energy barrier values, i.e., $\sigma = 0.5, 1, 2$.

$\kappa < \sqrt{\sigma}$—below the resonance value, the effective D–A coupling is a monotonic increasing function of the polaron energy $|u|$; and

$\kappa \geq \sqrt{\sigma}$—the effective D–A coupling increases for $|u| < |u_m|$, attains its maximum at the resonance $|u| = |u_m|$, and decreases for $|u| > |u_m|$. As a consequence of $|H_{da}(u, \sigma, 0)| \sim 1/\sigma$, both the maximum of ρ at the resonance and its limit at high $|u|$ increase with σ.

Effective D–A coupling versus κ. As we have discussed above and illustrated in Fig. 2, the electron vibration may lead to an increase or decrease of the effective D–A coupling depending on the value of $|u|$ and the regime of σ. Similarly, Fig. 3 displays the enhancement factor ρ at the polaron ground state as a function of the electron-vibration coupling κ. It is clear that there is a certain threshold value $\kappa_c(u_g)$ below which the electron-vibration coupling leads to enhancement of the effective D–A coupling and above which the effective D–A coupling is drastically reduced, hence, affecting the charge transfer efficiency.

As above, two different regimes can be distinguished:

(1) $\sigma \leq 1$—equation (6) has two distinct roots corresponding to the ground and excited states, respectively. The maximum enhancement $\rho(u_m, \sigma, \kappa_c)$ given by Eq. (15) is attained at the resonance $\Delta = 0$, where $u_g = -\sqrt{2}/2$ (and the excited state $-u_g$) and $\kappa_c(u_g) = \sqrt{2\sigma}$, obtained from the combination of Eqs. (6) and (12). In this regime, the ground state coincides with the resonant energy $u_g = u_m$.

(2) $\sigma > 1$—there is an interval $\kappa_c(u_g) \leq \kappa \leq \kappa_{max}$ within which Eq. (6) admits four distinct roots (the lowest one corresponding to the ground state) and out of

which it has two distinct roots. In this case, the ground state is no longer resonant, $u_g \neq u_m$, but two excited states coincide with the resonant energies given by $u_m = \pm\sqrt{2}/2$. As a result, the maximum enhancement $\rho(u_g, \sigma, \kappa) < \rho(u_m, \sigma, \kappa_c)$ as $\Delta(u_g) \neq 0$. For instance, for $\sigma = 2$, the interval of four distinct roots is $1.9336 \leq \kappa \leq 2.175$ with $\kappa_c = 1.9336$ and the maximum enhancement $\rho(u_g, \sigma, \kappa_c) = 3.198$ is obtained for $u_g = -1.097$. At the resonance $\Delta = 0$ for $\sigma = 2$, we have $\kappa = 2$, $u_g = -1.707$, $\Delta(u_g) = -\sqrt{2}$, and $\rho(u_g, \sigma, \kappa) = 1.596$.

In summary, we have shown that the electronic coupling with the vibration dynamics of the bridge results in the formation of a polaron that may, under certain conditions, lead to an enhancement of the charge transfer efficiency. Figures 2 and 3 show that the enhancement factor ρ is greater than one for a wide range of the energy barrier σ and the electron-vibration coupling κ. These findings are very suggestive for the issue of charge transport assisted by structural dynamics along the DNA chain. To study the basic mechanism of vibration enhancement of charge transport, we have focused in this work on the simple three-site model with a single harmonic structural dynamic mode (reaction coordinate). Meanwhile, the method employed in this work is not limited to this model and the extension of the theory to several sites and anharmonic reaction coordinates (see, e.g., [27, 28]) and several resonance states can be handled within the framework developed in [18]. Nevertheless, further theoretical studies need to be conduct along the lines outlined above in order to gain a better understanding of the charge transport properties in biological systems and technological applications of significant importance.

The work of E.K. was supported in part by INTAS, grant no. 01-0105.

REFERENCES

1. C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struct. Biol. **10**, 279 (2000).
2. D. Porath, A. Bezryadin, S. de Vries, and C. Dekker, Nature **403**, 635 (2000).
3. A. Heller, Faraday Discuss. **116**, 1 (2000).
4. M. R. Arkin, E. D. A. Stemp, R. E. Holmin, *et al.*, Science **273**, 475 (1996).
5. J. W. Evenson and M. Karplus, Science **262**, 1247 (1993).
6. P. J. Dandliker, R. E. Holmin, and J. K. Barton, Science **275**, 1465 (1997).
7. D. B. Hall and J. K. Barton, J. Am. Chem. Soc. **119**, 5045 (1997).
8. M. R. Arkin, E. D. A. Stemp, S. C. Pulver, and J. K. Barton, Chem. Biol. **4**, 369 (1997).
9. K. Fukui and K. Tanaka, Angew. Chem. Int. Ed. Engl. **37**, 158 (1998).
10. H. W. Fink and C. Schonenberg, Nature **398**, 407 (1999).
11. C. Wan, T. Fiebeg, O. Schiemann, *et al.*, Proc. Natl. Acad. Sci. USA **97**, 14 052 (2000).
12. B. Giese, J. Amaudrut, A.-K. Kohler, *et al.*, Nature **412**, 318 (2001).
13. H. D. Sikes, J. F. Smalley, S. P. Dudek, *et al.*, Science **291**, 1519 (2001).
14. Yu. A. Berlin, A. L. Burin, and M. A. Ratner, Chem. Phys. **275**, 61 (2002).
15. A. Nitzan and M. A. Ratner, Science **300**, 384 (2003).
16. C. R. Treadway, M. G. Hill, and J. K. Barton, Chem. Phys. **281**, 409 (2002).
17. J. Jortner, M. Bixon, T. Langenbacher, and M. E. Michel-Beyerle, Proc. Natl. Acad. Sci. USA **95**, 12 759 (1999).
18. D. J. Bicout, F. Varchon, and E. Kats, Europhys. Lett. **70**, 457 (2005).
19. M. Bixon and J. Jortner, J. Phys. Chem. B **104**, 3906 (2000).
20. E. G. Petrov, Ye. V. Shevchenko, V. I. Teslenko, and V. May, J. Chem. Phys. **115**, 7107 (2001).
21. E. G. Petrov and V. May, J. Phys. Chem. A **105**, 10176 (2001).
22. M. Bixon and J. Jortner, J. Phys. Chem. A **105**, 10322 (2001).
23. E. I. Kats and V. V. Lebedev, JETP Lett. **75**, 37 (2002).
24. D. J. Bicout and E. Kats, Phys. Lett. A **300**, 479 (2002).
25. R. Bruinsma, G. Grüner, M. R. D'Orsogna, and J. Rudnik, Phys. Rev. Lett. **85**, 4393 (2000).
26. M. R. D'Osorgna and J. Rudnik, Phys. Rev. E **66**, 041804 (2002).
27. D. Hennig, J. Chem. Phys. **112**, 10017 (2000).
28. V. D. Lakhno, J. Biol. Phys. **26**, 133 (2000).
29. D. Hennig, J. F. R. Archilla, and J. Agarwal, Physica D (Amsterdam) **180**, 256 (2003).
30. A. Omerzu, M. Licer, T. Mertelj, *et al.*, Phys. Rev. Lett. **93**, 218101 (2004).

JETP Letters, Vol. 82, No. 6, 2005, pp. 371–376. From Pis'ma v Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 82, No. 6, 2005, pp. 412–417.

Quenched Disorder Effects in Electron Transport in Si Inversion Layers in the Dilute Regime¶

V. M. Pudalov[a, *], M. E. Gershenson[b], N. N. Klimov[b, c], and H. Kojima[b]

[a] *Lebedev Physics Institute, Russian Academy of Sciences, Moscow, 119991 Russia*
e-mail: pudalov@Mail1.lebedev.ru
[b] *Department of Physics and Astronomy, Rutgers University, New Jersey, 08854 USA*
[c] *Lebedev Physics Research Center, Moscow, 119991 Russia*

Received July 21, 2005

In order to reveal the effects of disorder in the vicinity of the apparent metal–insulator transition in 2D, we studied electron transport in the same Si device after cooling it down to 4 K at different fixed values of the gate voltage V^{cool}. Different V^{cool} did not significantly modify either the momentum relaxation rate or the strength of electron–electron interactions. However, temperature dependences of the resistance and the magnetoresistance in parallel magnetic fields in the vicinity of the 2D metal–insulator transition carry a strong imprint of the quenched disorder determined by V^{cool}. This demonstrates that the observed transition between the metallic and insulating regimes, besides the universal effects of electron–electron interaction, depends on the sample-specific localized states (disorder). We report on evidence for a weak exchange of electrons between the reservoirs of extended and resonant localized states that occur at low densities. The strong cool-down dependent variations of $\rho(T)$, we believe, are evidence for a developing spatially inhomogeneous state in the critical regime. © *2005 Pleiades Publishing, Inc.*

PACS numbers: 71.27.+a, 71.30.+h, 72.20.Ee, 73.40.Qv

After about a decade of intensive research, the apparent metal–insulator transition (MIT) in two-dimensional (2D) systems remains a rapidly evolving field [1]. One of the central problems is to understand the individual roles of the two major driving forces: disorder and electron–electron (*e–e*) interactions. A great body of experimental data demonstrates that, at sufficiently large carrier densities, the low-temperature behavior of disordered systems is governed by the universal quantum interaction corrections to the conductivity [2–4]. These purely interaction effects between mobile 2D electrons have been intensively studied both theoretically [2–8] and experimentally [9–13]; the role of disorder in these studies is solely limited to scattering of mobile electrons.

In contrast, the interplay of disorder and interactions, particularly, interactions between localized and mobile electrons, is considered much more rarely [14–18]. There are clear observations that, near the apparent 2D MIT, the behavior of dilute systems is very rich and does not necessarily follow the same pattern [18–20]. One might expect that the interplay of disorder and interactions should become more and more important as the electron density decreases and approaches the critical density of the 2D MIT.

Usually, the presence of the localized states themselves in 2D transport is masked by mobile electrons. In order to reveal their contribution in the vicinity of the 2D MIT, we have studied electron transport in the same Si-MOS structure that was slowly cooled down from room temperature to $T = 4$ K at different fixed values of the gate voltage $V_g = V^{cool}$. Changing the cooling conditions primarily affects the thickness of the potential well [21]. We believe that this allowed us to vary the fine details of disorder—the structure of the resonant (localized) states—without affecting the type of disorder (short-ranged), the scattering rate, and the strength of electron–electron interactions in the system of mobile electrons. We focused on two key features of the 2D MIT: the strong dependences of the resistivity on temperature and the parallel magnetic field, and studied them in the density range $n = (0.7–3) \times 10^{11}$ cm^{-2} for a system with a snapshot disorder pattern.

We have observed that, at relatively high densities (resistivity $\rho \le 0.1h/e^2$, the dependences $\rho(T)$ and $\rho(B_\parallel)$ in weak parallel magnetic fields $B_\parallel$ are very similar for different cool downs. This "universal" behavior of transport at high densities agrees with our observation [11] of the sample-independent $\rho(T\tau, n)$ for samples with different mobility ($\mu \propto \tau$). In contrast, at low densities ($\rho \sim (0.1 – 1)h/e^2$), or in moderate and strong parallel fields $g\mu_B B_\parallel \sim E_F \gg k_B T$, the cooling conditions dramatically affect transport even though the main parameters of disorder and of the electron interactions

¶ The text was submitted by the authors in English.

*A member of the editorial board of the journal *JETP Letters* since 2000.

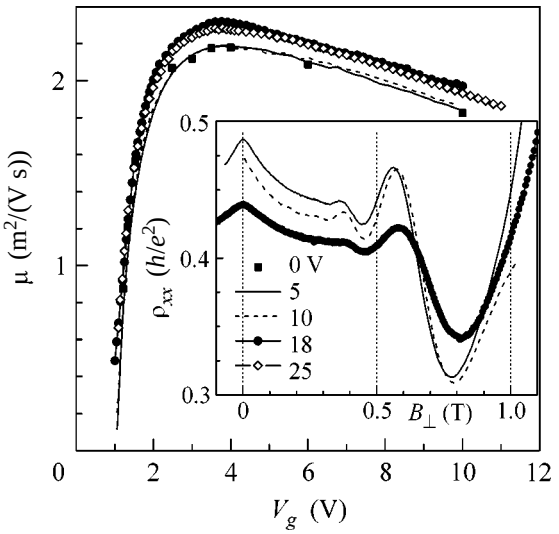


Fig. 1. Mobility versus gate voltage for different cool downs. The $V^{\rm cool}$ values for both the main panel and the inset are shown in the figure. Examples of the SdH oscillations, shown in the inset for the same $V_g = 1.15$ V, $T = 0.1$ K, $B_\parallel = 0.03$ T, demonstrate that the quantum time τ_q is not very sensitive to the cooling conditions. The carrier densities vary within the interval $(1.07–1.09) \times 10^{11}$ cm^{-2}.

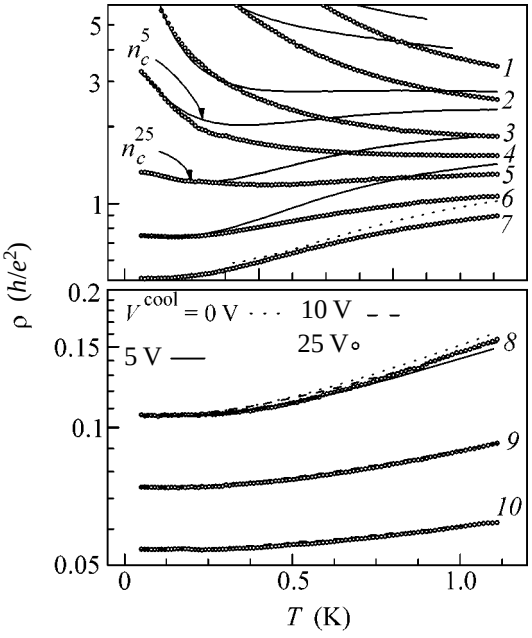


Fig. 2. Temperature dependences of the resistivity for four different cool downs. For curves *1* to *10*, the density values are 0.827, 0.882, 0.942, 0.972, 1.00, 1.038, 1.07, 1.18, 1.31, and 1.53 in unites of 10^{11} cm^{-2}. n_c^5 and n_c^{25} mark two critical dependences for cool downs at $V_g = 5$ and 25 V, respectively.

remain unchanged. This observation provides direct experimental evidence that, near the 2D MIT, electron transport at finite temperatures in dilute systems becomes sample-specific and dependent on more subtle details of disorder.

We have also observed that the frequency of the weak-field Shubnikov–de Haas (SdH) oscillations varies with temperature and in-plane field; these variations (of the order of a few %) grow as the density approaches the 2D MIT critical density. The SdH frequency is directly related with the density of mobile electrons, whatever strong interactions there are. Therefore, the observed variations of the mobile charge at a fixed total charge in the MOS capacitor gives evidence for the redistribution of electrons between the bands of mobile and localized states. The weak T dependence of this electron exchange, if it is attributed to activation processes, indicates the presence of very low energy barriers (~1 K) between the mobile and localized electron states. We relate the finite temperature cool-down effects to the hybridization of the mobile and spatially separated resonant (localized) states present at the Fermi energy at low densities.

The resistivity measurements were performed on a high mobility Si-MOSFET sample [22] at bath temperatures of 0.05–1.2 K. The crossed magnetic field system allowed us to accurately align the magnetic field parallel to the plane of the 2D electron system [9]. Five different cool downs were performed with $V^{\rm cool} = 0$, 5, 10, 18, and 25 V. The carrier density, found from the period of the SdH oscillations, varies linearly with V_g: $n \approx C \times (V_g - V_{\rm th})$, where C (1.10×10^{11}/(V cm^2) for the studied sample) is determined by the oxide thickness. The threshold voltage $V_{\rm th}$ varied little (within 0.15 V) for different cool downs and remained fixed as long as the sample was maintained at low temperatures (up to a few months). Within the same cool down, both C and $V_{\rm th}$ remained constant (within a few %) in the overall studied range of densities.

Figure 1 shows the mobility μ versus V_g for five different cool downs with $V^{\rm cool} = 0$, 5, 10, 18, and 25 V. The peak mobility for different $V^{\rm cool}$ varies by less than ~7%; this demonstrates that the momentum relaxation time τ is not strongly affected by the cooling conditions. Comparing the μ(n) curves with the conventional transport theory [23, 24], we conclude that the density of the charged impurities varies by less than 10^{10} cm^{-2} for different cool downs. We also observed that the amplitudes of the SdH oscillations are similar for different cool downs, as shown in the inset to Fig. 1. These two observations are consistent with each other, since the quantum lifetime τ_q is nearly equal to τ for Si-MOSFETs.

Figure 2 shows that the ρ(T) dependences are cool down independent far from the transition (at $\rho \ll h/e^2$) (see, e.g., curves *9*, *10*). However, in the vicinity of the transition ($\rho \sim h/e^2$), a dramatically different behavior is

observed as temperature increases. The irreproducibility of $\rho(T)$ for different cool downs is clearly seen for the curves in Fig. 2, which correspond to nearly the same ρ at the lowest T: these curves, being different at higher temperatures, converge with decreasing T. We have verified that the renormalization of the electron spin susceptibility and effective mass (and, thus, the two Fermi-liquid coupling constants) do not change for different cool downs to within 5%. Thus, electron–electron interaction effects [11] also cannot account for the changes in $\rho(T)$.

The sample-specific variations vanish at sufficiently low temperatures: this suggests that the underlying mechanism is related to the finite-temperature effects in a system that retains a quenched disorder. These results also suggest that, in addition to universal effects, a finite-temperature and sample-specific mechanism, which strongly affect the resistivity, come into play.

If the behavior shown in Fig. 2 is characterized by the critical density n_c, which corresponds to the transition, the latter would have been cool down dependent. The labels in Fig. 2 mark two $\rho(T)$ dependences, which corresponded to $n = n_c$ for two different cool downs; they were estimated from linear extrapolation to zero of the activation energy $\Delta(n)$ measured in the insulating regime $\rho(T) \propto \exp(\Delta/T)$ [20, 25] away from the critical regime. It is clear that the critical dependences $\rho(T, n = n_c)$ are nonmonotonic (see also [25, 26]). The nonmonotonicity is not caused by electron overheating; we applied sufficiently low source-drain current in order to reduce the excess electron temperature δT_e to a few mK. For example, for the curve n_c^{25} (the source drain resistances are 150 kOhm each; the channel resistance is ~30 kOhm/□), the chosen source-drain excitation 10 μV corresponds to the dissipation $\sim 10^{-15}$ W, which might cause electron overheating $\lesssim 1$ mK [26].

Since the cool down dependent changes in $\rho(T)$ vanish with decreasing temperature, we have attempted to analyze the variations $\delta\rho(T) = \rho(V_1^{\text{cool}}, T) - \rho(V_2^{\text{cool}}, T)$ in terms of the exponential $\exp(-\Delta/T)$ dependence as demonstrated in Fig. 3. The corresponding activation energy Δ is very low: it varies within the interval ~(0.7–1) K. This proves that the low-lying band of localized states (located close to the bottom of the conduction band, ≈8 K below the Fermi level) is irrelevant. The smallness of Δ, therefore, points to the involvement of the localized states that are located close to the Fermi level. Similar resonant localized states are known in narrow band-gap semiconductors and must be spatially separated from the mobile states.

We now turn to the magnetoresistance (MR) in parallel fields; the data are shown in Figs. 4 and 5. This MR is usually associated with the spin effects [1, 19]. In the theoretical models of the parallel-field MR based on electron–electron interactions, the MR is controlled by the effective g^* factor and the momentum relaxation time τ [6, 11, 4]. An important advantage of our method

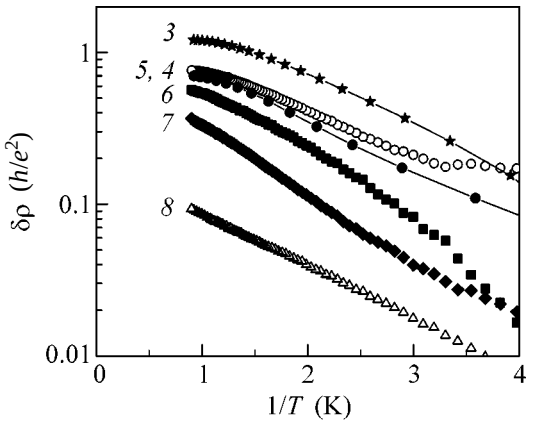


Fig. 3. Difference between resistivity values for two different cool downs (shown in Fig. 2) versus inverse temperature. The numbers label the curves for the densities the same as in Fig. 2.

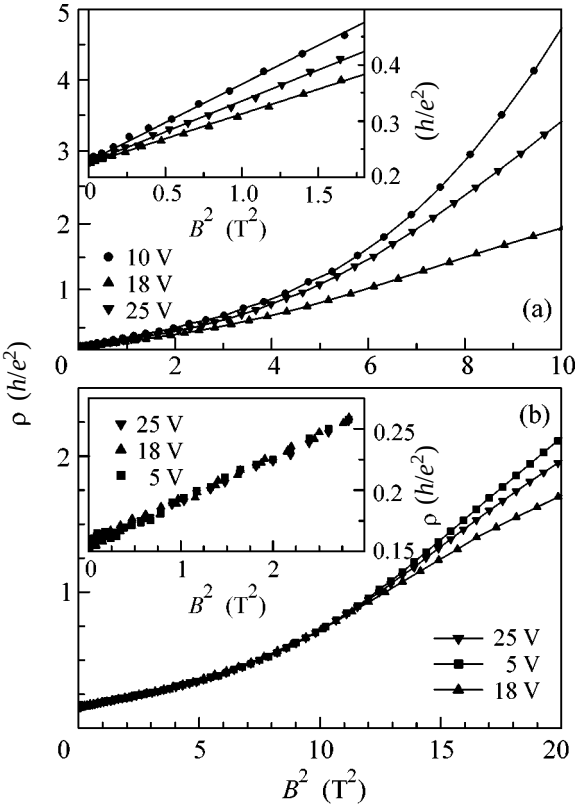


Fig. 4. Examples of the dependences $\rho(B_{\parallel}^2)$ at $T = 0.3$ K for the carrier density (a) 1.20×10^{11} cm^{-2} and (b) 1.34×10^{11} cm^{-2}. The insets blow up the low-field region of the quadratic behavior. The values of V^{cool} are indicated for each curve.

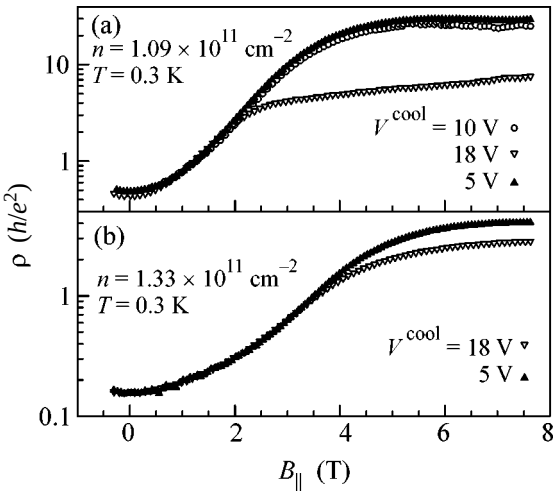


Fig. 5. Resistivity vs. $B_\parallel$ field for three cool downs at two electron densities.

is that, as we mentioned above, cooling of the same sample at different $V^{\rm cool}$ does not affect these parameters. Thus, one might expect to observe a sample-independent behavior if the MR is controlled solely by the universal interaction effects.

Firstly, we consider the range of fields much weaker than the field of complete spin polarization ($g^*\mu_B B_\parallel \ll E_F$). The insets to Figs. 4a and 4b show that the MR is proportional to $B_\parallel^2$ at $g^*\mu_B B_\parallel/k_B T \le 1$. We found that the slope $d\rho/dB^2$ is nearly cool down independent (i.e., universal) only for the densities $n > 1.3 \times 10^{11}$ cm^{-2} (which are 30% greater than the critical density n_c) or for the resistivities $\rho(0) < 0.16h/e^2$ (compare the insets to Figs. 4a and 4b); this is consistent with our earlier observations [11]. With approaching n_c, this universality vanishes: Fig. 4a shows that even when the zero-field resistivity is as small as $0.22h/e^2$, the slope varies by a factor of 1.3 for different $V^{\rm cool}$.

For the intermediate fields, $k_B T < g^*\mu_B B_\parallel < E_F$, the $\rho(B_\parallel)$ behavior is not universal over the whole density range $n = (1–3) \times 10^{11}$ cm^{-2} (Fig. 4). As n decreases and approaches n_c, the cool down dependent variations of $\rho(B_\parallel)$ increase progressively.

The influence of the cool down conditions on the magnetoresistance becomes even more dramatic in strong fields: $B_\parallel \gtrsim E_F/g^*\mu_B$. Despite the fact that the dependences $\mu(n) \propto \tau(n)$ for different cooldowns were very similar (Fig. 1), we observed very large variations in the high-field MR. Figures 5a and 5b show $\rho(B_\parallel)$ for different cool downs at two values of n. The cool down conditions cause factor-of-five changes in $\rho(B)$ in high fields and factor-of-two changes in the values of $B_\parallel = B_{\rm sat}$ at which the MR saturates at a given carrier density. The latter quantity was determined from the intercept of the tangents at the fields below and above the MR saturation [19].

The nonuniversal, sample-dependent behavior of $\rho(B_\parallel)$ agrees with earlier observations made on different samples [19]. We emphasize that the curves for different $V^{\rm cool}$ (in both Figs. 5a and 5b) nominally correspond to the same density. The fact that $B_{\rm sat}$ is a cool down dependent parameter proves that the MR in strong parallel fields is not solely related to spin-polarization of mobile electrons. The fact that the variations arise in strong fields $g\mu_B B_\parallel \sim E_F$ hints that a deep tail of localized states located near the bottom of the conduction band (or near the bottom of the upper spin subband) [14, 15, 27, 28] is responsible for the magnetoresistance variations. We note that, at low temperatures T = 0–1 K $\ll T_F$ and at zero field, temperature activation of carriers from the tail of localized states to the Fermi level (across the energy gap $\sim E_F$) is negligibly weak and could not affect the data shown in Fig. 2. In contrast, at higher temperatures $T \sim T_F$, the thermal activation of carriers from the tail of localized states to the Fermi level produces noticeable effects, which are detected in the Hall voltage [29].

It is worth mentioning that the influence of variable disorder on transport and magnetotransport in Si-MOSFETs has been studied earlier. Both temperature dependence $\rho(T)$ and magnetoresistance $\rho(B_\parallel)$ were found to be different in samples with different mobility [20, 19], in samples cooled down with different values of the substrate bias voltage [30], and with an intentionally varied oxide charge [23]. In contrast, in our studies, we kept constant the scattering time, the quantum time, the phase breaking time, the interface charge, and the parameters relevant to electron–electron interaction. Even under such conditions, strong nonuniversal variations in $\rho(T)$ and $\rho(B_\parallel)$ occur.

In order to elucidate the origin of the observed variations in disorder, we have analyzed SdH oscillations at weak perpendicular magnetic fields versus temperature and in-plane magnetic field. Figure 6 shows typical $\rho_{xx}(B_\perp)$ curves for six temperatures that were measured during the same cool down for a fixed gate voltage value. The ρ_{xx} minima occur when the Fermi energy coincides with the middle of the energy gap. The upper-left inset clearly shows that the minima of the oscillations shift with temperature, thereby, providing evidence for the changes in the density of mobile carriers. The lower-left inset demonstrates that the shift of the ρ_{xx} minima is not caused by variations in the residual field of the superconducting magnet (maintained at 4 K).

We fitted the total oscillatory picture with the theoretical dependence (similar to that in [9]) using the frequency of oscillations $n_{\rm SdH}$ for each of the curves as a fitting parameter. The resulting temperature dependences of $n_{\rm SdH}$ are shown in Fig. 7. The error bars on the figure correspond to relative changes of the frequency

with temperature; the absolute frequencies are determined with about three times lower precision. At a higher density $n > 5 \times 10^{11}$ cm^{-2}, the oscillation frequency was independent of temperature within the 0.5% uncertainty. The $n_{SdH}(T)$ changes become noticeable at a density of $\leq 4 \times 10^{11}$ cm^{-2} (which is four times larger than the critical density of the 2D MIT); they increase progressively with in-plane field, as Figs. 7a and 7b show.

DISCUSSION

The measured density values n_{SdH} refer to the density of mobile electrons that participate in the Landau quantization. The weak temperature variations of the density of mobile electrons do not involve a large energy scale on the order of $E_F \sim 8$ K. This points to the presence of resonant localized states at the Fermi level, which are separated spatially and by a small energy barrier from the mobile states. The density variations are then caused by exchange of electrons between those two states via either overbarrier transitions, as schematically drawn in Fig. 7, or via tunneling.

temperature induced exchange of electrons between the bands of mobile and localized states, by itself, can not produce a significant effect on the resistivity. Indeed, the appearance (disappearance) of ~10^9 cm^{-2} charged scatterrers (which corresponds to the 0.5% variations of density in Fig. 7) may, correspondingly, cause 0.5% changes in ρ. However, in the critical regime, the localized states may be expected to occupy a significant share of the total 2D layer by forming clusters. The periphery of the cluster is expected to consist of the resonant states that may emit and absorb electrons. As a result, the overall area available for motion of mobile electrons changes with temperature, similar to that in the known percolation picture [31]. The resonant states, thus, may control transport through the saddle points separating neighboring areas occupied by mobile electrons and, thus, indirectly trigger the strong changes in ρ.

In the $B_\parallel$ field, the resonant state should split and move relative to the band of mobile states. This model can potentially explain both the strong variations of $\rho(T)$ and $\rho(B_\parallel)$ in the critical regime and the weak changes in the mobile carrier density. Formation of the two-phase state may be caused by either disorder or electron–electron interactions. Spontaneous formation of the heterophase state in the vicinity of the phase transition was established for quasi-1D systems [32]; in 2D electron systems, the two-phase state is also intrinsic to some theoretical models [33].

To summarize, by cooling the same high-mobility Si-MOS sample at different fixed values of the gate voltage, we tested the universality of temperature and magnetic-field dependences of the resistivity near the 2D MIT. An important advantage of this approach is that the different cool down procedures do not alter the interaction effects between mobile carriers. It has been found that, in the vicinity of the transition ($\rho \sim h/e^2$), the specific cool down effects strongly affect $\rho(T)$; these effects vanish only when ρ decreases below ~$0.1h/e^2$

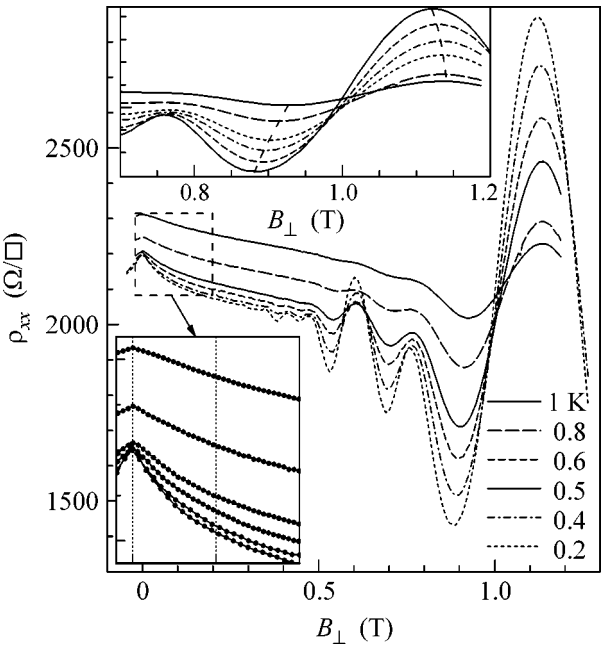


Fig. 6. Typical Shubnikov–de Haas oscillations at six temperatures (indicated on the main panel) and in in-plane field $B_\parallel = 0.02$ T. The nominal density value is $n \approx 2 \times 10^{11}$ cm^{-2}. The lower left inset demonstrates the precise control of the zero magnetic field position, and the upper inset magnifies one of the oscillations to show its shifting with *T*.

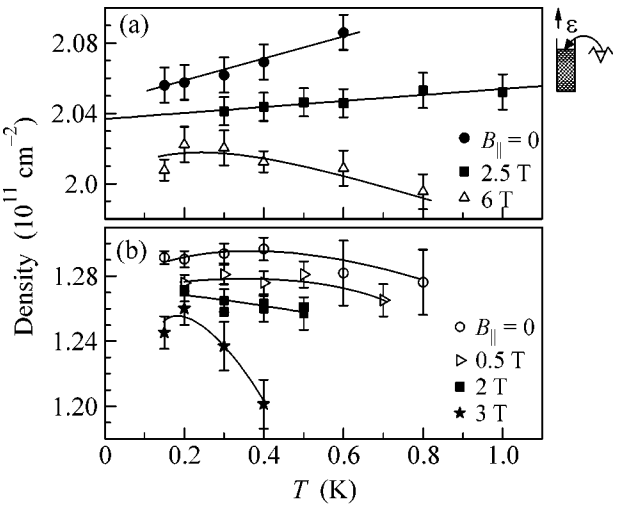


Fig. 7. Temperature dependences of the frequency of SdH oscillations at two fixed values of the gate voltage. The different curves within each panel are offset shifted for clarity. The diagrams on the right schematically show the DOS for the band of mobile electrons and a resonant state. The arrow shows an overbarrier transition.

with increasing electron density; they also vanish as T decreases. The nonuniversal behavior is especially dramatic in strong $B_\parallel$ fields, where it extends to much higher electron densities (we observed pronounced nonuniversality of $R(B_\parallel)$ over a range of $n = (1–3) \times 10^{11}$ cm^{-2}).

Our results reveal the existence of the resonant (shallow) localized states near the Fermi energy. The observed temperature variation of the frequency of Shubnikov–de Haas oscillations demonstrates a weak exchange of electrons between the reservoirs of mobile and resonant localized states. The large changes of $\rho(T)$ at elevated temperature signify the development of a spatial inhomogeneity of the 2D system, which may result from either interactions between electrons or disorder.

One of the authors (V.P.) acknowledges discussions with V.I. Kozub and Yu.M. Galperin. This work was supported by the NSF, the Russian Foundation for Basic Research, Programs of the Presidium and of the Division of Physical Sciences of the RAS, and the presidential program "State Support of Leading Scientific Schools."

REFERENCES

1. E. Abrahams, S. Kravchenko, and M. P. Sarachik, Rev. Mod. Phys. **73**, 251 (2001).
2. For a review, see: B. L. Altshuler and A. G. Aronov, in *Electron–Electron Interactions in Disordered Systems*, Ed. by A. L. Efros and M. Pollak (Elsevier, Amsterdam, 1985); P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
3. G. Zala, B. N. Narozny, and I. L. Aleiner, Phys. Rev. B **64**, 214204 (2001); **65**, 020201(R) (2002).
4. I. V. Gornyi and A. D. Mirlin, Phys. Rev. B **69**, 045313 (2004).
5. S. Das Sarma and E. H. Hwang, Phys. Rev. Lett. **83**, 164 (1999).
6. A. M. Finkelstein, in *Soviet Scientific Reviews*, Ed. by I. M. Khalatnikov (Harwood Academic, London, 1990), Vol. 14, p. 3.
7. C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984); C. Castellani, G. Kotliar, and P. A. Lee, Phys. Rev. Lett. **59**, 323 (1987); C. Castellani, C. Di Castro, H. Fakuyama, *et al.*, Phys. Rev. B **33**, 7277 (1986); C. Castellani, C. Di Castro, and P. A. Lee, Phys. Rev. B **57**, R9381 (1998).
8. A. Punnoose and A. M. Finkelstein, Phys. Rev. Lett. **88**, 016802 (2002).
9. V. M. Pudalov, M. Gershenson, H. Kojima, *et al.*, Phys. Rev. Lett. **88**, 196404 (2002).
10. Y. Y. Proskuryakov, A. K. Savchenko, S. S. Safonov, *et al.*, Phys. Rev. Lett. **89**, 076406 (2002).
11. V. M. Pudalov, M. Gershenson, H. Kojima, *et al.*, Phys. Rev. Lett. **91**, 126403 (2003).
12. S. A. Vitkalov, K. James, B. N. Narozhny, *et al.*, Phys. Rev. B **67**, 113310 (2003).
13. J. Zhu, H. L. Stormer, L. N. Pfeiffer, *et al.*, Phys. Rev. Lett. **90**, 056805 (2003).
14. N. F. Mott, *Metal–Insulator Transitions* (Taylor and Francis, London, 1974; Nauka, Moscow, 1979).
15. V. I. Kozub and N. V. Agrinskaya, Phys. Rev. B **64**, 245103 (2001).
16. T. M. Klapwijk and S. Das Sarma, Solid State Commun. **110**, 581 (1999).
17. B. L. Altshuler and D. L. Maslov, Phys. Rev. Lett. **83**, 2092 (1999).
18. V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, cond-mat/0103087.
19. V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, Phys. Rev. Lett. **88**, 076401 (2002).
20. V. M. Pudalov, G. Brunthaler, A. Prinz, and G. Bauer, JETP Lett. **68**, 442 (1998).
21. V^{cool} determines the depth of the confining potential well and, simultaneously, the number of interface traps sunk under the Fermi level. At low temperatures, as V_g is varied, the potential well remains almost unchanged and memorizes an imprint of the disorder formed during its cooling down.
22. Sample Si6-14 with 190 nm thick gate oxide was fabricated on (001)-Si wafer and had a rectangular channel 2.5×0.25 mm oriented along [010].
23. For a review see, T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
24. A. Gold and W. Götze, Phys. Rev. B **33**, 2495 (1986).
25. B. L. Altshuler, D. L. Maslov, and V. M. Pudalov, Physica E (Amsterdam) **9**, 209 (2001).
26. O. Prus, M. Reznikov, U. Sivan, and V. Pudalov, Phys. Rev. Lett. **88**, 016801 (2002).
27. S. A. Vitkalov, M. P. Sarachik, and T. M. Klapwijk, Phys. Rev. B **65**, 201106 (2002).
28. A. Gold and V. T. Dolgopolov, Phys. Rev. Lett. **89**, 129701 (2002).
29. A. Yu. Kuntsevich, D. A. Knyazev, V. I. Kozub, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 502 (2005) [JETP Lett. **81**, 409 (2005)].
30. A. Lewalle, M. Pepper, C. J. B. Ford, *et al.*, cond-mat/0108244.
31. Y. Meir, Phys. Rev. Lett. **83**, 3506 (1999).
32. A. V. Kornilov, V. M. Pudalov, Y. Kitaoka, *et al.*, Phys. Rev. B **69**, 224404 (2004).
33. B. Spivak, Phys. Rev. B **64**, 085317 (2001).