Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 723–747. Translated from Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 755–779.


# Nonperturbative Phenomena in QCD at Finite Temperature

**N. O. Agasian***

*Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received July 26, 2004

**Abstract**—The main objective of the present study is to analyze various nonperturbative phenomena in QCD both at low, $T < T_c$, and at high, $T > T_c$, temperatures. New methods are developed that make it possible, on one hand, to describe data obtained by numerically simulating QCD on a lattice and, on the other hand, to study new physical phenomena in QCD at finite temperature. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

The physics of strong interactions, QCD, has been flourishing over the past three decades. For many years, investigations into the behavior of strongly interacting matter under various external effects have been of great topical interest. In the real world, these are primarily temperature and the baryon density. Interest in the behavior of matter under extreme conditions (high temperatures commensurate with the characteristic QCD scale, $T \sim 200$ MeV, and high baryon densities, $n > n_0 \simeq 0.17\,\text{fm}^{-3}$, where $n_0$ is the normal nuclear density) is motivated by the fact that an increase in energy was achieved in experiments aimed at studying heavy-ion collisions. In view of this, it is expected that densities and temperatures at which a phase transition to a new state of strongly interacting matter, quark–gluon plasma [1, 2], is possible are reached in such experiments.[1)]

Extreme conditions existed at the initial stage of expansion of the Universe. Within the time interval $t \sim 10^{-6}$–$10^{-5}$ s after the Big Bang, the Universe passed the stage of a strong phase transition in which the system transformed into a hadronic phase dominated by the essentially nonperturbative phenomena of confinement and spontaneous chiral-symmetry breaking. Extremely high baryon densities ($n \sim 10n_0$) also exist in central regions of neutron stars, where the color-superconductivity phase, which was predicted theoretically in recent years (for an overview, see [3–5]), can be realized.

Quantum chromodynamics is a quantum theory of Yang–Mills gauge fields interacting with quark fermion fields. The asymptotic-freedom phenomenon [6, 7] and a topologically nontrivial structure of the vacuum of non-Abelian gauge theories [8–12] were discovered in the 1970s. A further development of the theory revealed that it is precisely the complicated nonperturbative structure of the vacuum (nonperturbative fluctuations of vacuum fields) that is responsible for the phenomena of confinement and spontaneous chiral-symmetry breaking and, thereby, for the formation of the physical hadron spectrum. Thus, the development of new theoretical approaches was required for describing nonperturbative phenomena in quantum field theories.

Investigations of the vacuum state at finite temperature and finite values of the chemical potential and external fields lead to new interesting phenomena, including various phase transitions. Accordingly, there arises the need for developing a theoretical formalism for exploring the behavior of a quantum-field system and its vacuum state under external effects.

Quantum field theory at finite temperature and a finite value of the chemical potential has been studied predominantly along three lines:

(i) Perturbative calculations have been performed, and various resummation schemes for perturbation-theory series have been constructed. Advances have predominantly been made in developing the hard-thermal-loop (HTL) and hard-dense-loop (HDL) approximations (for an overview, see [13, 14]).

(ii) Various effective models that describe one nonperturbative phenomenon in actual QCD or another have been constructed. For example, various approaches based on employing sigma models to

*e-mail: agasian@heron.itep.ru

1)It should be emphasized from the outset that there is presently no consensus on the theoretical interpretation of the experiments in question. The main reason is that it is not quite clear whether the initial stage of a heavy-ion collision can be justifiably described in terms of a steady-state thermodynamically equilibrium system in the phase of quark–gluon plasma.

describe the chiral order parameter in QCD have been developed; investigations within effective chiral theory at $T \neq 0$ have been performed; and various generalizations of the Nambu–Jona-Lasinio model to the case of finite temperature, a finite value of the chemical potential, and finite external fields have been introduced. Other effective models have also been developed (for an overview, see [3, 5]).

(iii) The complex nonperturbative structure of the vacuum of non-Abelian gauge theories—in particular, QCD at finite $T$ and $n$—has been studied by means of numerical simulations with the aid of computers. Interesting and important results have been obtained along this line, and it is expected that an increase in the power of computers and the creation of new computational schemes would make it possible to obtain deeper insights into the nonperturbative dynamics of QCD (for an overview, see [15–17]).

It should be noted that the high-temperature phase ($T > T_c$), which involves restored chiral symmetry and deconfined quarks and gluons, is described in a conventional way as a quark–gluon plasma by analogy with an electromagnetic plasma. In this approach, the approximation of noninteracting quark and gluon gases is used for a zero-order approximation; further, the interaction is taken into account in the form of a series in powers of the QCD running coupling constant $g(T)$. In view of the asymptotic-freedom phenomenon in QCD, the coupling constant must decrease with temperature. In order to improve the convergence of a relevant series—for example, in calculating pressure, which is among the main thermodynamic quantities that characterize the system—use is made of the method for summing hard thermal loops, which effectively reduces to introducing the so-called magnetic gluon mass. The presence of a nonperturbative magnetic mass, $\propto g^2(T)T$, in the theory is necessary for removing well-known infrared singularities [18]. At the same time, the QCD interaction remains strong, $g(5T_c) \sim 1$, even at temperatures $T \sim 5T_c$, so that the use of perturbation theory, at least as a standard means for taking into account small corrections in the interaction to the zero-order approximation, is not quite legitimate. Moreover, the magnetic-confinement phenomenon in non-Abelian gauge theories, which occurs at all temperatures, inevitably requires the presence of fluctuating nonperturbative chromomagnetic fields [19–22] (see also [23–25]). Thus, QCD thermodynamics above the critical temperature must be described in terms of a phase that involves a strong chromomagnetic condensate, against whose background there arise excitations of hot quarks and gluons.

Relations that are obtained as corollaries of the symmetry properties of the theory play an important role in quantum field theory. Searches for symmetries and constraints that these symmetries impose on physical characteristics of the system are of particular importance in QCD, which is a theory that involves confinement and where composite states (hadrons) appear as observables. Low-energy theorems, or Ward identities (scale and chiral ones), play a crucial role in the understanding of nonperturbative vacuum properties of QCD. Stricly speaking, low-energy theorems were formulated almost simultaneously with the beginning of the application of quantum-field methods in particle physics (see, for example, Low theorems [26]). In QCD, they were obtained in the early 1980s [27]. Low-energy theorems in QCD, which follow from general symmetry properties and which are independent of the details of the confinement mechanism, make it possible to obtain information that sometimes cannot be deduced in any other way; they can also be used as "physically reasonable" constraints in constructing effective theories and various models of the QCD vacuum. In the present study, we develop a method that makes it possible to generalize low-energy QCD theorems to the case of finite temperature. The nonperturbative QCD vacuum and condensates at $T \neq 0$ are studied by using this method.

As was indicated above, nonperturbative fluctuations of gluon fields play a crucial role in the vacuum of non-Abelian gauge theories and determine many features of QCD. The instanton-liquid model, which was proposed in [28, 29], is an elaborate theory that explains quite successfully some phenomena in QCD. In this pattern, well-separated and not very strongly interacting instantons and anti-instantons (from here, the name "liquid" comes) are the main nonperturbative fields. The instanton density is approximately $N/V_4 = 1$ fm$^{-4}$. This model makes it possible to solve a number of problems in QCD—in particular, the phenomenon of spontaneous chiral-symmetry breaking naturally arises there and the $\eta'$-meson mass is explained within it. Nevertheless, the model in question possesses a number of serious drawbacks—namely, it is not known how the instanton–anti-instanton ensemble is stabilized (problem of an infrared inflation of instantons) and it is impossible to explain the confinement phenomenon within the instanton-liquid model. However, the vacuum involves, in addition to semiclassical instantons, other nonperturbative fields, which enables one, among other things, to solve the infrared problem of instantons.

The nonperturbative QCD vacuum can be parameterized by a set of nonlocal gauge-invariant vacuum expectation values of gluon-field strengths [30–32]. It appears that, in this way, one can describe well a large number of hadron-physics phenomena (see the review article of Di Giacomo *et al.* [33]). In order to

explain qualitatively relevant effects, it is sufficient, in many cases, to take into account only a bilocal correlation function (stochastic-vacuum model). Moreover, there exist indications that corrections that arise owing to the inclusion of higher correlation functions are moderate, being about a few percent in some cases [33]. In this approach, a nonzero string tension arises naturally. It is expressed in terms of the vacuum expectation values of the strengths of chromoelectric fields [31]. Within the stochastic-vacuum model, the problem of an infrared inflation of instantons can also be solved in a natural way. Investigations into the effect of nonperturbative quantum fluctuations on instantons were initiated in [34, 35]. Later, it was shown that, in the nonperturbative stochastic vacuum, standard perturbation theory changes and that the contribution of large-size instantons to physical quantities becomes finite [36]. Further, it was proven in [37] that the direct interaction of an instanton with nonperturbative vacuum fields leads to the stabilization of the instanton at scales on the order on the correlation length in the vacuum condensate. The size distribution of instantons that has a maximum at $\rho_c \simeq 0.3$ fm was obtained and was found to agree well with data from lattice calculations.

The scaling of the string tension between sources in different representations of the color group [38] is observed in lattice calculations and is an important property of the vacuum. It was shown in [39] that the instanton contribution to the potential of heavy-quark interaction in different representations violates Casimir scaling, whence one obtains a constraint on the instanton density.

The main objective of the present study is to analyze various nonperturbative phenomena in QCD both at low ($T < T_c$) and at high ($T > T_c$) temperatures. New methods are developed here that make it possible to describe available lattice data, on one hand, and to study new physical phenomena in QCD at finite temperature, on the other hand. A separate section is devoted to studying topologically nontrivial field configurations (calorons) and their contribution to the bilocal correlation function in QCD at finite temperature.

The ensuing exposition is organized as follows. Section 2 is devoted to studying thermodynamic properties of the nonperturbative QCD vacuum in the hadron phase—that is, at temperatures below the temperature of the quark–hadron phase transition. An expression that relates the anomalous contribution to the trace of the energy–momentum tensor to the thermodynamic pressure in light-quark QCD is obtained without any model assumptions by considering the QCD partition function and its renormalization-group properties. The generalization of low-energy QCD theorems to the case of finite temperature is given. Analytic expressions are derived for the temperature dependences of the quark and gluon condensates at low temperatures, $T < M_\pi$. A low-temperature relation is derived for QCD: it is shown that the temperature derivatives of the anomalous and normal (quark-mass term) contributions to the trace of the energy–momentum tensor in QCD are equal to each other in the low-temperature ($T < M_\pi$) region. Leading corrections to this relation that are associated with the excitation of massive $K$ and $\eta$ mesons are obtained. In the temperature region $T < 140$ MeV, these corrections are less than 4%. The strange quark condensate $\langle \bar{s}s \rangle$ is studied, and it is shown that, with increasing temperature, it decreases much more slowly than the light-quark condensate $\langle \bar{u}u \rangle = \langle \bar{d}d \rangle$.

In Section 3, expressions for the temperature dependences of the quark and gluon condensates up to the critical temperature $T_c$ are derived within the hadron-resonance-gas model. It is shown that, upon taking into account the chromoelectric shift of the hadron mass, the quark condensate and only half of the gluon condensate (chromoelectric component, which is responsible for confinement and string formation) evaporate at the same temperature, $T_c \sim 190$ MeV. In the vicinity of this temperature, the energy density is $\varepsilon(T \sim T_c) \sim 1$–$1.5$ GeV/fm$^3$.

In Section 4, attention is given to studying the properties of the magnetic sector of $SU(N)$ Yang–Mills theory at finite temperature [22] and to considering the interesting phenomenon of magnetic confinement. It is shown that the finite-temperature behavior of physical quantities in the magnetic sector is qualitatively different in two temperature regions. The temperature dependence of the gauge-invariant bilocal correlation function for chromomagnetic-field strengths and the spacelike string tension $\sigma_s(T)$ is derived analytically. It is revealed that, in the region $T < 2T_c$, the chromomagnetic condensate grows slowly with temperature—that is, $\langle H^2 \rangle(T) = \langle H^2 \rangle \coth(M/2T)$, where $M = 1/\xi_m \simeq 1.5$ GeV is the inverse magnetic correlation length, which is independent of temperature for $T < 2T_c$. In the region $T > 2T_c$, the amplitude of the chromomagnetic correlation function increases with increasing temperature, $\langle H^2 \rangle(T) \propto g^8(T)T^4$, while the correlation length decreases, $\xi_m(T) \propto 1/(g^2(T)T)$. This behavior of the chromomagnetic correlation function explains the magnetic-confinement phenomenon within $4d$ $SU(N)$ Yang–Mills theory at finite temperature. The resulting temperature dependence of the spacelike string tension agrees perfectly with lattice data over the entire temperature range. The temperature range is found within which one can go over to the reduced $3d$ theory—that is, to a description of the

dynamics of the chromomagnetic sector in terms of static correlation functions. The relative contribution of nonzero Matsubara modes to the spacelike string tension is calculated. At $T = 2T_c$, this contribution is about 5%.

In Section 5, the bilocal correlation function in gluodynamics at finite temperature is calculated within the instanton-gas model. It is shown that the correlation length in the instanton vacuum decreases with increasing temperature. Our results are compared with lattice data for the bilocal correlation function at $T \neq 0$. The problem of the instanton density and the possible structures of the nonperturbative vacuum are discussed.

The main results of the present study are formulated in the Conclusion.

## 2. NONPERTURBATIVE QCD VACUUM AT FINITE TEMPERATURE

At low temperatures, $T < T_c$ ($T_c$ is the temperature of the quark–hadron phase transition), QCD dynamics is essentially nonperturbative and is characterized by the phenomena of confinement and spontaneous chiral-symmetry breaking. In the hadronic phase, realized at low temperatures, the partition function is dominated by the contribution of the lightest particles in the physical spectrum. The $\pi$ meson, which is a Goldstone excitation in the chiral condensate in the limit of two massless quarks, plays this role in QCD. Therefore, effective chiral theory [40–42], which is often referred to as chiral perturbation theory, is used as a standard method in low-temperature QCD physics.

An analysis of the behavior of the order parameter of the chiral phase transition (quark condensate $\langle\bar{q}q\rangle$) with increasing temperature is an important problem in finite-temperature QCD. In the ideal-gas approximation (this corresponds to the one-loop level in chiral perturbation theory), the contribution of massless $\pi$ mesons to the quark condensate is proportional to $T^2$ [43–45]. Within chiral perturbation theory, two-loop ($\sim T^4$) and three-loop ($\sim T^6$) contributions to the quark condensate $\langle\bar{q}q\rangle$ were calculated in [44–46] and [47, 48], respectively. In the general case of $N_f$ massless quarks, the low-temperature expansion of the quark condensate in chiral perturbation theory has the form

$$\frac{\langle\bar{q}q\rangle(T)}{\langle\bar{q}q\rangle} = 1 - \frac{N_f^2-1}{N_f}\frac{T^2}{12F_\pi^2} \qquad (1)$$
$$- \frac{N_f^2-1}{2N_f^2}\left(\frac{T^2}{12F_\pi^2}\right)^2 - N_f(N_f^2-1)\left(\frac{T^2}{12F_\pi^2}\right)^3$$
$$\times \ln\frac{\Lambda_q}{T} + O(T^8),$$

where $F_\pi \simeq 93$ MeV is the axial coupling constant and $\Lambda_q \simeq 470$ MeV ($N_f = 2$) [48]. It is well known that, owing to the smallness of the $\pi$-meson mass with respect to the characteristic mass scale of strong interactions, $M_\pi \ll 1$ GeV, the pion stands out among other strongly interacting particles. Therefore, the chiral limit, $M_\pi \to 0$, is a good approximation for many QCD problems arising at zero temperature. On the other hand, it is obvious that, in finite-temperature QCD, there arises a new mass scale that is determined by the critical (phase-transition) temperature $T_c$, which separates two fundamentally different phases of strongly interacting matter—the hadronic phase, which involves confinement and spontaneous chiral-symmetry breaking, and the quark–gluon phase, where there is no color confinement and where chiral invariance is restored. Numerically, the critical temperature appears to be close to the pion mass, $T_c \approx M_\pi$.[2)] In the hadronic phase, there is therefore no small parameter $M_\pi/T$, so that there is no interval $M_\pi \ll T < T_c$, where the expansion in (1) would be valid for thermal pion excitations. However, hadron states are heavier than the $\pi$ meson, having masses severalfold greater than $T_c$; therefore, their contribution to thermodynamic quantities in the low-temperature hadronic phase is suppressed by the Boltzmann factor $\sim \exp\{-m_h/T\}$. Thus, the thermodynamics of the low-temperature hadronic phase, $T < M_\pi$, is described mainly in terms of thermal excitations of relativistic massive $\pi$ mesons. For the temperature dependence of the quark condensate, the inclusion of a nonzero quark mass was studied in [48] within chiral perturbation theory, where the $u$- and $d$-quark masses are considered as excitations. Accordingly, the thermodynamic quantities [pressure, $\langle\bar{q}q\rangle(T)$] are represented in the form of series in powers of $T$ and $m_q$. Owing to the fact that the spectrum of an unperturbed system involves the massless pion, the quark mass term leads to infrared divergences for $T \to 0$—that is, there appear terms in the quark condensate $\langle\bar{q}q\rangle(T)$ that involve negative powers of $T(\sim m_q/T)$ [48]. A scheme of resummation of these divergences was proposed in [48], and the quark condensate $\langle\bar{q}q\rangle(T)$ was represented there as a graph of a function of $T$.

For the gluon condensate $\langle G^2\rangle \equiv \langle(gG^a_{\mu\nu})^2\rangle$, the situation is substantially different. The gluon condensate is not an order parameter for a phase transition and does not lead to any spontaneous symmetry breaking. At the quantum level, an anomaly in the trace of the energy–momentum tensor leads to a violation of scaling, and this phenomenon is described by

[2)] According to lattice calculations, the temperature of the deconfining phase transition is $T_c(N_f = 2) \simeq 173$ MeV or $T_c(N_f = 3) \simeq 154$ MeV [49].

a nonzero value of $\langle G^2\rangle$. However, this is not a spontaneous symmetry breaking; therefore, this does not generate Goldstone particles. In gluodynamics, the mass of the lightest excitation of the $0^{++}$ glueball—it can be identified with a dilaton—is directly related to the gluon condensate, $m_D \propto (\langle G^2\rangle)^{1/4}$. Thus, we can state that, in the gluon sector of QCD, thermal excitations of massive glueballs ($M_{0^{++}} \sim 1.5$ GeV) are suppressed in proportion to $\exp\{-M_{\rm gl}/T\}$, so that their contribution to the finite-temperature shift of the gluon condensate is small, $\Delta\langle G^2\rangle/\langle G^2\rangle \sim 0.1\%$ at $T = 200$ MeV [50].

Further, we note that, in the one-loop approximation of chiral perturbation theory, the pions are described as a gas of noninteracting massless particles. Obviously, this system is scale-invariant; therefore, it does not contribute to the trace of the energy–momentum tensor and, accordingly, to $\langle G^2\rangle$. It was demonstrated in [51] that the gluon condensate appears to be temperature dependent only at the three-loop level of chiral perturbation theory upon taking into account the interaction between Goldstone bosons.

In this section, we develop a method that makes it possible to calculate the temperature dependences of quark and gluon condensates beyond chiral perturbation theory. On the basis of general renormalization-group relations at $T \neq 0$ and $m_q \neq 0$, we derive an equation that relates the anomalous contribution to the trace of the energy–momentum tensor (gluon condensate) to thermodynamic pressure in QCD including light quarks. Methodologically, this approach is a generalization of low-energy QCD theorems to the case of finite temperature. The proposed method is quite general and model-independent and can be used to study various nonperturbative phenomena in QCD at finite temperature and in external fields. In particular, an investigation of nonperturbative QCD in a magnetic field (at $T = 0$ and $T \neq 0$) on the basis of this method makes it possible to predict a nontrivial behavior of the quark condensate (freezing of the order parameter for the chiral phase transition [52, 53]) and of the gluon condensate in response to variations in the magnetic field [54–56].

### *2.1. Relation between Thermodynamic Pressure and Condensates in QCD: Renormalization-Group Equations and Low-Energy Theorems at $T \neq 0$*

At a nonzero quark mass ($m_q \neq 0$), scale invariance is broken even at the classical level. This implies that, in the ideal gas approximation, thermal excitations of $\pi$ mesons will change the gluon condensate with increasing temperature. In order to determine this dependence, we will use general renormalizability properties and renormalization-group relations for the QCD partition function. In the ensuing exposition, we will follow the approach developed in [57, 58].

The QCD partition function normalized to perturbation theory at $T = 0$ with two sorts of quarks (the generalization to the case of arbitrary $N_f$ is trivial) can be written in the form ($\beta = 1/T$)

$$Z = \frac{1}{Z_{PT}}\int [DA]\prod_{q=u,d}[D\bar{q}][Dq] \times \exp\left\{-\int\limits_0^\beta dx_4\int\limits_V d^3x\mathcal{L}\right\}, \tag{2}$$

where, as usual, we have imposed periodic boundary conditions on boson (gauge) fields and antiperiodic boundary conditions on fermion fields. Here, the QCD Lagrangian has the form

$$\mathcal{L} = \frac{1}{4g_0^2}(G^a_{\mu\nu})^2 + \sum_{q=u,d}\bar{q}\left[\gamma_\mu\left(\partial_\mu - i\frac{\lambda^a}{2}A^a_\mu\right) + m_{0q}\right]q, \tag{3}$$

where we have omitted the ghost and gauge-fixing terms since they are immaterial for our further considerations. The free-energy density is given by

$$F(T, m_{0u}, m_{0d}) = -\frac{1}{\beta V}\ln Z.$$

From Eq. (2), we obtain the expression for the gluon condensate in the form

$$\langle G^2\rangle(T, m_{0u}, m_{0d}) = 4\frac{\partial F}{\partial(1/g_0^2)}. \tag{4}$$

The system described by the partition function (2) is characterized by the set of dimensional parameters $M$, $T$, and $m_{0q}(M)$ and by the dimensionless charge $g_0^2(M)$, where the bare quark masses $m_{0q}$ and the coupling constant $g_0^2$ are determined at the ultraviolet-cutoff mass $M$. On the other hand, one can go over to a renormalized free energy $F_R$ and, by using the renormalization-invariance properties of $F_R$, recast (4) into a form that involves derivatives with respect to physical parameters, $T$ and the renormalized masses $m_q$. For this, we will use a dimensional transmutation, which is a well-known property of QCD. The dimensional transmutation leads to the appearance of the dimensional nonperturbative parameter

$$\Lambda = M\exp\left\{\int\limits_{\alpha_s(M)}^{\infty}\frac{d\alpha_s}{\beta(\alpha_s)}\right\}, \tag{5}$$

where $\alpha_s = g_0^2/4\pi$ and $\beta(\alpha_s) = d\alpha_s(M)/d\ln M$ is the Gell-Mann–Low function. Furthermore, it is well known that the quark mass $m_0$ has an anomalous dimension $\gamma_m$ and depends on the scale $M$. The renormalization-group equation for the running mass $m_0(M)$ has the form $d\ln m_0/d\ln M = -\gamma_m$, and we use the modified-subtraction ($\overline{MS}$) scheme, where $\beta$ and $\gamma_m$ are independent of the quark mass [59]. The renormalization-invariant mass is then given by

$$m_q = m_{0q}(M)\exp\left\{\int^{\alpha_s(M)} \frac{\gamma_{m_q}(\alpha_s)}{\beta(\alpha_s)}d\alpha_s\right\}, \quad (6)$$

where the indefinite integral is taken at the point $\alpha_s(M)$. Since the physical (renormalized) free energy is a renormalization-invariant quantity, its anomalous dimension is zero. Thus, the quantity $F_R$ has only a normal (canonical) dimension, which is equal to four. By using the renormalization invariance of the quantity $\Lambda$, we can represent $F_R$ in the most general form

$$F_R = \Lambda^4 f\left(\frac{T}{\Lambda}, \frac{m_u}{\Lambda}, \frac{m_d}{\Lambda}\right), \quad (7)$$

where $f$ is a function of the ratios $T/\Lambda, \ldots$. From relations (5), (6), and (7), we obtain

$$\frac{\partial F_R}{\partial(1/g_0^2)} = \frac{\partial F_R}{\partial\Lambda}\frac{\partial\Lambda}{\partial(1/g_0^2)} + \sum_q \frac{\partial F_R}{\partial m_q}\frac{\partial m_q}{\partial(1/g_0^2)}, \quad (8)$$

$$\frac{\partial m_q}{\partial(1/g_0^2)} = -4\pi\alpha_s^2 m_q \frac{\gamma_{m_q}(\alpha_s)}{\beta(\alpha_s)}. \quad (9)$$

With allowance for expression (4), we find for the gluon condensate in QCD that [57][3)]

$$\langle G^2\rangle(T, m_u, m_d) = \frac{16\pi\alpha_s^2}{\beta(\alpha_s)} \quad (10)$$
$$\times\left(4 - T\frac{\partial}{\partial T} - \sum_q (1+\gamma_{m_q}) m_q \frac{\partial}{\partial m_q}\right) F_R.$$

In the one-loop approximation, we have $\beta(\alpha_s) \to -b\alpha_s^2/2\pi$ and $1+\gamma_m \to 1$, where $b = (11N_c - 2N_f)/3$. Thus, we arrive at the following expressions for the condensates:

$$\langle G^2\rangle(T) = -\frac{32\pi^2}{b} \quad (11)$$
$$\times\left(4 - T\frac{\partial}{\partial T} - \sum_q m_q\frac{\partial}{\partial m_q}\right) F_R \equiv -\hat{D}F_R,$$

$$\langle\bar{q}q\rangle(T) = \frac{\partial F_R}{\partial m_q}. \quad (12)$$

By using the above relations, one can obtain low-energy QCD theorems at finite temperature.[4)] For the sake of simplicity, we consider the chiral limit $m_q = 0$. We introduce the field $\sigma(\tau = x_4, \mathbf{x})$,

$$\sigma(\tau, \mathbf{x}) = -\frac{b_0}{32\pi^2}(G^a_{\mu\nu}(\tau, \mathbf{x}))^2, \quad (13)$$

and the operator $\hat{D}$,

$$\hat{D} = 4 - T\frac{\partial}{\partial T}. \quad (14)$$

By differentiating (4) with respect to $1/g_0^2$ $n$ times and taking into account Eqs. (7), (13), and (14), we obtain

$$\hat{D}^{n+1}F = \hat{D}^n\langle\sigma(0,\mathbf{0})\rangle \quad (15)$$
$$= \int d\tau_n d^3x_n \ldots \int d\tau_1 d^3x_1$$
$$\times\langle\sigma(\tau_n, \mathbf{x}_n)\ldots\sigma(\tau_1, \mathbf{x}_1)\sigma(0,\mathbf{0})\rangle_c.$$

The subscript $c$ in (15) indicates that only connected diagrams are included in the vacuum expectation value. A similar argument applies to an arbitrary operator $\hat{O}(x)$ constructed from quark fields and gluon fields or from both of them. We have

$$\left(T\frac{\partial}{\partial T} - d\right)^n\langle\hat{O}\rangle \quad (16)$$
$$= \int d\tau_n d^3x_n \ldots \int d\tau_1 d^3x_1$$
$$\times\langle\sigma(\tau_n, \mathbf{x}_n)\ldots\sigma(\tau_1, \mathbf{x}_1)\hat{O}(0,\mathbf{0})\rangle_c,$$

where $d$ is a canonical dimension of the operator $\hat{O}$. If the operator $\hat{O}$ has an anomalous dimension as well, the corresponding $\gamma$ function must be taken into account.

The above relations (11) and (12) are general and are applicable at any temperature $T$. They relate the nonperturbative condensates to the free energy or, which is the same, to the sign-reversed pressure. The free energy of the system is determined by the spectrum of possible quantum states in a specific phase. Thus, finding the behavior of the condensates versus temperature reduces to evaluating the free energy of the phase being studied. Further, we will explore the hadronic state of strongly interacting matter—that is, the confining phase.

[3)] An anomalous contribution to the trace of the energy–momentum tensor in QCD is related to the gluon condensate by the equation $\langle\theta^g_{\mu\mu}\rangle = (\beta(\alpha_s)/(16\pi\alpha_s^2))\langle G^2\rangle(T, m_u, m_d)$.

[4)] Low-energy theorems in gluodynamics at finite temperature were considered in [60].

### 2.2. Quark and Gluon Condensates at Low Temperature

The effective pressure (we normalize the partition function to perturbation theory at $T = 0$), from which the condensates $\langle\bar{q}q\rangle(T)$ and $\langle G^2\rangle(T)$ can be extracted with the aid of relations (11) and (12), is given by

$$P_{\rm eff}(T) = -\varepsilon_{\rm vac} + P_h(T), \tag{17}$$

where $\varepsilon_{\rm vac} = \langle\theta_{\mu\mu}\rangle/4$ is the nonperturbative vacuum-energy density at $T = 0$ and

$$\langle\theta_{\mu\mu}\rangle = -\frac{b}{32\pi^2}\langle G^2\rangle + \sum_{q=u,d} m_q\langle\bar{q}q\rangle \tag{18}$$

is the trace of the energy–momentum tensor. In Eq. (17), $P_h$ is the thermodynamic pressure in the hadronic phase; for a gas of massive pions, we have

$$P_\pi = -3T\int\frac{d^3p}{(2\pi)^3}\ln(1-\exp(-\sqrt{\mathbf{p}^2+M_\pi^2}/T)). \tag{19}$$

The quark and gluon condensates are then given by

$$\langle\bar{q}q\rangle(T) = -\partial P_{\rm eff}/\partial m_q, \tag{20}$$

$$\langle G^2\rangle(T) = \hat{D}P_{\rm eff}, \tag{21}$$

where the operator $\hat{D}$ was defined by Eq. (11),

$$\hat{D} = \frac{32\pi^2}{b}\left(4 - T\frac{\partial}{\partial T} - \sum_q m_q\frac{\partial}{\partial m_q}\right). \tag{22}$$

Let us consider the case of zero temperature, $T = 0$. By using the low-energy theorem for the derivative of the gluon condensate with respect to the quark mass [27],

$$\frac{\partial}{\partial m_q}\langle G^2\rangle = \int d^4x\langle G^2(0)\bar{q}q(x)\rangle = -\frac{96\pi^2}{b}\langle\bar{q}q\rangle + O(m_q), \tag{23}$$

where $O(m_q)$ is a term that is linear in the light-quark mass, we arrive at the relation [5)]

$$\frac{\partial\varepsilon_{\rm vac}}{\partial m_q} = -\frac{b}{128\pi^2}\frac{\partial}{\partial m_q}\langle G^2\rangle + \frac{1}{4}\langle\bar{q}q\rangle = \frac{3}{4}\langle\bar{q}q\rangle + \frac{1}{4}\langle\bar{q}q\rangle = \langle\bar{q}q\rangle. \tag{24}$$

We note that three-fourths of the quark condensate stems from the gluon part of the nonperturbative vacuum energy density. Following the same lines of reasoning, we obtain

$$-\hat{D}\varepsilon_{\rm vac} = -\frac{32\pi^2}{b}\left(4 - \sum_q m_q\frac{\partial}{\partial m_q}\right) \times\left(-\frac{b_0}{128\pi^2}\langle G^2\rangle + \frac{1}{4}\sum_q m_q\langle\bar{q}q\rangle\right) = \langle G^2\rangle. \tag{25}$$

In order to find the dependence of the quark and gluon condensates on $T$, we use the Gell-Mann–Oakes–Renner (GMOR) relation [61] ($\Sigma \equiv -\langle\bar{q}q\rangle = |\langle\bar{u}u\rangle| = |\langle\bar{d}d\rangle|$):

$$F_\pi^2M_\pi^2 = -\frac{1}{2}(m_u+m_d)\langle\bar{u}u+\bar{d}d\rangle = (m_u+m_d)\Sigma. \tag{26}$$

We then have

$$\frac{\partial}{\partial m_q} = \frac{\Sigma}{F_\pi^2}\frac{\partial}{\partial M_\pi^2}, \tag{27}$$

$$\sum_q m_q\frac{\partial}{\partial m_q} = (m_u+m_d)\frac{\Sigma}{F_\pi^2}\frac{\partial}{\partial M_\pi^2} = M_\pi^2\frac{\partial}{\partial M_\pi^2}, \tag{28}$$

$$\hat{D} = \frac{32\pi^2}{b}\left(4 - T\frac{\partial}{\partial T} - M_\pi^2\frac{\partial}{\partial M_\pi^2}\right). \tag{29}$$

Bringing together all of the above formulas, we finally represent the condensates as [57, 58]

$$\frac{\Sigma(T)}{\Sigma} = 1 - \frac{1}{F_\pi^2}\frac{\partial P_h(T)}{\partial M_\pi^2}, \tag{30}$$

$$\langle G^2\rangle(T) = \langle G^2\rangle + \frac{32\pi^2}{b}\left(4 - T\frac{\partial}{\partial T} - M_\pi^2\frac{\partial}{\partial M_\pi^2}\right)P_h(T). \tag{31}$$

Let us consider the thermodynamics of the hadronic phase at low temperatures, $T \lesssim M_\pi$. It was indicated above that, in this region, the main contribution to thermodynamic quantities comes from thermal excitations of massive $\pi$ mesons. The pion-gas density at a temperature $T$ is given by

$$n_\pi(T) = 3\int\frac{d^3p}{(2\pi)^3}\frac{1}{\exp(\sqrt{\mathbf{p}^2+M_\pi^2}/T)-1}. \tag{32}$$

For the mean distance between the particles of a gas, we then have $L_\pi \simeq n_\pi^{-1/3}$. The dilute-gas approximation is applicable if the mean distance between the particles, $L_\pi$, is much shorter than the mean free path calculated with allowance for collisions, $L_\pi \ll \lambda_{\pi\pi}$. For a hadron gas, the mean free path is well known [62–64] at low temperatures, in which case the gas

[5)] Equation (24) includes corrections proportional to $m_q\partial\langle\bar{q}q\rangle/\partial m_q \sim O(m_q)$, which are small for light $u$ and $d$ quarks.

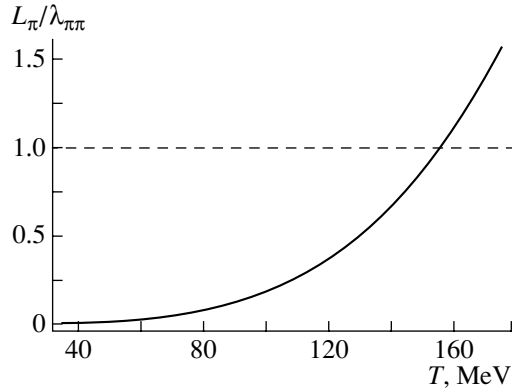


**Fig. 1.** Ratio of the mean distance between particles, $L_\pi$, and the free path $\lambda_{\pi\pi}$ in the pion gas as a function of temperature.

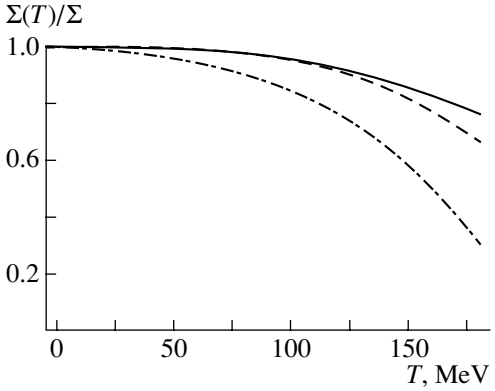


**Fig. 2.** Quark condensate $\Sigma(T)/\Sigma$ as a function of temperature: (solid curve) results of the calculation by formula (34); (dash-dotted curve) three-loop result within chiral perturbation theory (1) in the chiral limit for $N_f = 2$, and (dashed curve) numerical three-loop result of chiral perturbation theory at a nonzero quark mass ($M_\pi = 140$ MeV) [48].

consists almost completely of pions. An analysis of pion–pion scattering reveals that, in the temperature range $50 < T < 140$ MeV, the relation $\lambda_{\pi\pi} \simeq 12F_\pi^4/T^5$ [64] holds for the mean free path to within 20% [65]. Figure 1 shows the ratio $L_\pi/\lambda_{\pi\pi}$ as a function of temperature $T$. Thus, one can see that the gas approximation is correct for pions at low temperatures ($T \lesssim M_\pi$).

The pressure of a relativistic massive pion gas is given by

$$P_\pi(T) = -\frac{3T^4}{2\pi^2}\int\limits_0^\infty x^2 dx \ln(1 - e^{-\omega_\pi(x)}) \qquad (33)$$
$$= \frac{3T^4}{2\pi^2}\sum_{n=1}^\infty \frac{1}{n}\int\limits_0^\infty x^2 dx e^{-n\omega_\pi(x)}$$
$$= \frac{3M_\pi^2 T^2}{2\pi^2}\sum_{n=1}^\infty \frac{1}{n^2}K_2\left(n\frac{M_\pi}{T}\right),$$

where $\omega_\pi(x) = \sqrt{x^2 + M_\pi^2/T^2}$ and $K_2$ is the modified Bessel function of second order. By using relations (20), (24), (27), and (33), one can derive the following expression for the temperature dependence of the quark condensate [57]:

$$\frac{\Sigma(T)}{\Sigma} = 1 - \frac{3T^2}{4\pi^2 F_\pi^2}\int\limits_0^\infty \frac{x^2 dx}{\omega_\pi(x)(e^{\omega_\pi(x)} - 1)} \qquad (34)$$
$$= 1 - \frac{3M_\pi T}{4\pi^2 F_\pi^2}\sum_{n=1}^\infty \frac{1}{n}K_1\left(n\frac{M_\pi}{T}\right).$$

In the chiral limit, $M_\pi \ll T < T_c$, we find that the shift of the quark condensate is determined by the following standard relation of chiral perturbation theory:

$$\frac{\Delta\Sigma}{\Sigma} = -\frac{3T^2}{4\pi^2 F_\pi^2}\sum_{n=1}^\infty \frac{1}{n^2} = -\frac{T^2}{8F_\pi^2}. \qquad (35)$$

In the opposite limit case of low temperatures, $T \ll M_\pi$, we obtain

$$\frac{\Delta\Sigma}{\Sigma} = -\frac{3M_\pi T}{4\pi^2 F_\pi^2}K_1\left(\frac{M_\pi}{T}\right) \qquad (36)$$
$$\to -\frac{3M_\pi^{1/2}T^{3/2}}{2^{5/2}\pi^{3/2}F_\pi^2}e^{-M_\pi/T}.$$

The quark condensate as a function of temperature is shown in Fig. 2. One can see that, for relativistic massive pions, the gas-approximation formula (34) agrees well with the results of the numerical calculations of $\Sigma(T)$ that were performed in [48] in the three-loop approximation of chiral perturbation theory at a nonzero quark mass. At $T = M_\pi$, the deviation is as small as 2.5%.

Let us consider the temperature dependence of the gluon condensate within the above approach. The use of Eqs. (21), (25), and (29) leads to the following result [57]:

$$\frac{\langle G^2\rangle(T)}{\langle G^2\rangle} = 1 - \frac{24}{b}\frac{M_\pi^2 T^2}{\langle G^2\rangle} \qquad (37)$$
$$\times \int\limits_0^\infty \frac{x^2 dx}{\omega_\pi(x)(e^{\omega_\pi(x)} - 1)} = 1 - \frac{24}{b}\frac{M_\pi^3 T}{\langle G^2\rangle}$$
$$\times \sum_{n=1}^\infty \frac{1}{n}K_1\left(n\frac{M_\pi}{T}\right).$$

For $M_\pi \ll T$, we have

$$\frac{\Delta\langle G^2\rangle}{\langle G^2\rangle} = -\frac{24}{b}\frac{M_\pi^2 T^2}{\langle G^2\rangle}\sum_{n=1}^{\infty}\frac{1}{n^2} = -\frac{4\pi^2}{b}\frac{M_\pi^2 T^2}{\langle G^2\rangle}. \tag{38}$$

In the chiral limit, $M_\pi = 0$, we have $\Delta\langle G^2\rangle(T) = 0$ in accordance with the fact that a free gas of massless particles is conformally invariant. In the nonrelativistic (low-temperature) limit, $T \ll M_\pi$, we obtain

$$\frac{\Delta\langle G^2\rangle}{\langle G^2\rangle} = -\frac{24}{b}\frac{M_\pi^3 T}{\langle G^2\rangle}K_1\left(\frac{M_\pi}{T}\right) \tag{39}$$
$$\to -\frac{3\cdot 2^{5/2}\pi^{1/2}}{b}\frac{M_\pi^{5/2}T^{3/2}}{\langle G^2\rangle}e^{-M_\pi/T}.$$

At the same time, we indicated above that, within chiral perturbation theory at zero quark mass, the temperature dependence of the gluon condensate arises only at the three-loop level because of the interaction between massless pions [51]. In [47, 48], the pressure of $N_f$ massless quarks was found in the three-loop approximation of chiral perturbation theory. The result is

$$P(T)^{N_f=2}_{\mathrm{ChPT}} \tag{40}$$
$$= \frac{\pi^2}{30}T^4\left\{1 + 4\left(\frac{T^2}{12F_\pi^2}\right)^2\ln\frac{\Lambda_p}{T}\right\} + O(T^{10}),$$

where the numerical value of $\Lambda_p \simeq 275$ MeV [48] was obtained at $m_u = m_d = 0$ and fixed $m_s$. By using Eqs. (21), (22), and (27), the following expression (Leutwyler [51]) can be obtained for the gluon condensate:

$$\frac{\langle G^2\rangle(T)_{\mathrm{ChPT}}}{\langle G^2\rangle} = 1 - \frac{16\pi^4}{135b}\frac{T^8}{F_\pi^4\langle G^2\rangle}\left(\ln\frac{\Lambda_p}{T} - \frac{1}{4}\right). \tag{41}$$

Figure 3 shows the gluon condensate as a function of temperature at $\langle G^2\rangle = 0.5$ GeV$^4$.[6)] One can see that, in the low-temperature region, thermal excitations of $\pi$ mesons change the gluon condensate only slightly. From the physical point of view, the smallness of the thermal shift $\Delta\langle G^2\rangle(T)$ can be associated with a great numerical value of $\langle G^2\rangle$ at $T = 0$ in relation to the characteristic parameters in the finite-temperature hadronic phase—that is, $\Delta\langle G^2\rangle/\langle G^2\rangle \propto M_\pi^2 T_c^2/\langle G^2\rangle \simeq 10^{-3}$ at $M_\pi = 0.14$ GeV, $T_c = 0.17$ GeV, and $\langle G^2\rangle = 0.5$ GeV$^4$.

[6)]This corresponds to the standard value of $\langle\frac{\alpha_s}{\pi}F_{\mu\nu}^2\rangle = 0.012$ GeV$^4$ [66].

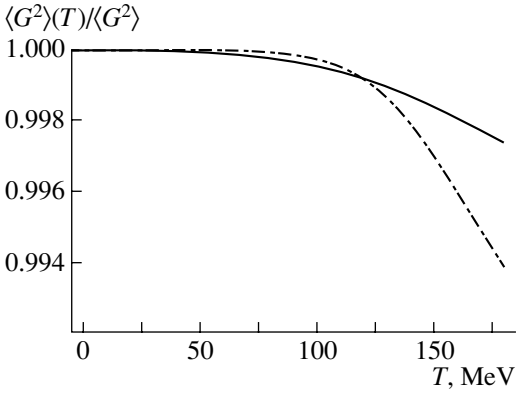


**Fig. 3.** Gluon condensate $\langle G^2\rangle(T)/\langle G^2\rangle$ as a function of temperature: (solid curve) results obtained by formula (37) at $\langle G^2\rangle = 0.5$ GeV$^4$ and (dash-dotted curve) three-loop results of chiral perturbation theory in the chiral limit [see Eq. (41)].

### *2.3. Low-Temperature Relations in QCD*

Within the approach proposed above, one can derive thermodynamic relations for the anomalous and normal (quark-mass term) contributions to the trace of the energy–momentum tensor in QCD [57, 58]. At low temperatures, the main contribution to pressure comes from thermal excitations of $\pi$ mesons. In general, the pressure of the gas formed by particles of mass $M_\pi$ can be written in the form

$$P_\pi = T^4\varphi(M_\pi/T), \tag{42}$$

where $\varphi$ is a function of the ratio $M_\pi/T$; in the case of a dilute gas, it is given by expression (33). The following relation then holds:

$$\left(4 - T\frac{\partial}{\partial T} - M_\pi^2\frac{\partial}{\partial M_\pi^2}\right)P_\pi = M_\pi^2\frac{\partial P_\pi}{\partial M_\pi^2}. \tag{43}$$

With the aid of Eqs. (20), (21), (24), (25), and (43), one can obtain

$$\Delta\langle\bar{q}q\rangle = -\frac{\partial P_\pi}{\partial m_q}, \quad \Delta\langle G^2\rangle = \frac{32\pi^2}{b}M_\pi^2\frac{\partial P_\pi}{\partial M_\pi^2}, \tag{44}$$

where $\Delta\langle\bar{q}q\rangle = \langle\bar{q}q\rangle(T) - \langle\bar{q}q\rangle$ and $\Delta\langle G^2\rangle = \langle G^2\rangle(T) - \langle G^2\rangle$. Taking into account relation (28), we can recast (44) into the form

$$\Delta\langle G^2\rangle = -\frac{32\pi^2}{b}\sum_q m_q\Delta\langle\bar{q}q\rangle. \tag{45}$$

By differentiating expression (45) with respect to $T$, we arrive at

$$\frac{\partial\langle G^2\rangle}{\partial T} = -\frac{32\pi^2}{b}\sum_q m_q\frac{\partial\langle\bar{q}q\rangle}{\partial T}. \tag{46}$$

This equation can be rewritten in the form [58]

$$\frac{\partial\langle\theta^g_{\mu\mu}\rangle}{\partial T} = \frac{\partial\langle\theta^q_{\mu\mu}\rangle}{\partial T}, \tag{47}$$

where $\langle\theta^q_{\mu\mu}\rangle = \sum m_q\langle\bar{q}q\rangle$ and $\langle\theta^g_{\mu\mu}\rangle = (\beta(\alpha_s)/(16\pi\times\alpha_s^2))\langle G^2\rangle$ are, respectively, the quark and the gluon contribution to the trace of the energy–momentum tensor.

As was mentioned above, the $\pi$ meson plays a special role in QCD thermodynamics because its mass is numerically close to the phase-transition temperature, while the masses of the remaining hadrons are severalfold larger than $T_c$. The effect of thermal excitations of massive hadrons on the features of the quark and gluon condensates was investigated in [67] within the conformally generalized nonlinear $\sigma$ model. It was shown that, at low temperatures, $T < 100$ MeV, the contribution to $\langle\bar{q}q\rangle$ from massive states is below 5%. At $T = M_\pi$, this contribution is approximately 10%.

Let us now consider leading corrections to the low-temperature relations (45)–(47). It is obvious that these leading corrections are associated with the pion–pion interaction, since, in this case, its contribution to pressure is proportional to the square of the pion density, $\propto e^{-2M_\pi/T}$. The inclusion of the $s$ quark also leads to contributions to pressure that are proportional to $e^{-M_K/T}$ and $e^{-M_\eta/T}$ and which are associated with thermal excitations of, respectively, $K$ and $\eta$ mesons. The pressure in the hadron phase can then be represented in the form

$$P_h(T) = P_g(T) + P_{\pi\pi}(T), \tag{48}$$

$$P_g(T) = \sum_{i=\pi,K,\eta} P_i(T), \tag{49}$$

where $P_i(T) = g_i T^4\varphi(M_i/T)$ is the pressure of the $i$-meson gas ($i = \pi, K, \eta$) and $g_i$ is the number of degrees of freedom of the $i$th state with $g_\pi = 3$, $g_K = 4$, and $g_\eta = 1$. The pressure associated with the pion–pion interaction can be written in a general form as

$$P_{\pi\pi} = T^4\frac{M_\pi^2}{F_\pi^2}f\left(\frac{M_\pi}{T}\right), \tag{50}$$

where $f$ is a function of the ratio $M_\pi/T$ and the factor $M_\pi^2/F_\pi^2$ stems from the vertex corresponding to the pion–pion interaction. Following the same lines of reasoning as in deriving relation (43), we can find, with the aid of the Gell-Mann–Oakes–Renner formula for three quark flavors, that

$$\hat{D}P_g(T) \tag{51}$$
$$= \frac{32\pi^2}{b}\left(4 - T\frac{\partial}{\partial T} - \sum_{q=u,d,s} m_q\frac{\partial}{\partial m_q}\right)P_g(T)$$
$$= \frac{32\pi^2}{b}\sum_{i=\pi,K,\eta} M_i^2\frac{\partial P_i}{\partial M_i^2}.$$

For the temperature shift of the condensates, we then obtain [68]

$$\frac{\Delta\Sigma(T)}{\Sigma} = \frac{\partial P_g}{\partial m_u} \tag{52}$$
$$= \frac{1}{F_\pi^2}\left(\frac{\partial P_\pi}{\partial M_\pi^2} + \frac{\partial P_K}{\partial M_K^2} + \frac{1}{3}\frac{\partial P_\eta}{\partial M_\eta^2}\right),$$

$$\frac{\Delta\Sigma_s(T)}{\Sigma_s} = \frac{\partial P_g}{\partial m_s} \tag{53}$$
$$= \frac{1}{F_\pi^2}\left(\frac{\partial P_K}{\partial M_K^2} + \frac{4}{3}\frac{\partial P_\eta}{\partial M_\eta^2}\right).$$

We note that, since the $\pi$ meson does not carry strangeness, it does not participate in the evaporation of the strange condensate ($\langle\bar{s}s\rangle$). The leading contribution to $\Delta\langle\bar{s}s\rangle(T)$ is associated with the excitation of the lightest particle carrying strangeness—that is, with the $K$ meson. The $K$-meson mass ($M_K = 493$ MeV) is severalfold larger than the pion mass $M_\pi$. The contribution of the $K$ meson to $\Delta\Sigma_s(T)$ can easily be derived on the basis of the low-temperature expression for the light-quark condensate $\Delta\Sigma(T)$ [see (36)] upon the obvious substitutions $M_\pi \to M_K$ and $F_\pi \to F_K$ and the multiplication by a factor of 4/3. For the ratio of the condensates, we then have [68]

$$\frac{\Delta\Sigma_s(T)/\Sigma_s}{\Delta\Sigma(T)/\Sigma} = \frac{4}{3}\left(\frac{M_K}{M_\pi}\right)^{1/2}\left(\frac{F_\pi}{F_K}\right)^2 e^{(M_\pi - M_K)/T}, \tag{54}$$

and this ratio is about 0.13 at $T \sim 140$ MeV. Employing the same procedure as in deriving Eqs. (45) and taking into account Eqs. (50)–(53), we further have

$$-\frac{b}{32\pi^2}\Delta\langle G^2\rangle = \sum_{q=u,d,s} m_q\Delta\langle\bar{q}q\rangle + 2P_{\pi\pi} \tag{55}$$
$$+ \frac{1}{2}M_\pi^2\frac{\partial P_K}{\partial M_K^2}.$$

Let us introduce the functions

$$\theta^\pm(T) = \langle\theta^g_{\mu\mu} \pm \theta^q_{\mu\mu}\rangle(T) - \langle\theta^g_{\mu\mu} \pm \theta^q_{\mu\mu}\rangle(0). \tag{56}$$

Here, $\theta^+(T)$ is the finite-temperature part of the total trace of the energy–momentum tensor; that is, $\theta^+(T) = \Delta\langle\theta^{\rm tot}_{\mu\mu}\rangle(T) = \varepsilon - 3P$, where $\varepsilon = TdP/dT - P$ is the energy density. The function

$$\delta_\theta(T) = \frac{\theta^-(T)}{\theta^+(T)} \tag{57}$$
$$= \left(2P_{\pi\pi} + \frac{1}{2}M_\pi^2\frac{\partial P_K}{\partial M_K^2}\right)\Big/(\varepsilon - 3P)$$

can then be considered as a measure of the deviation from the low-temperature relation (45). Let us estimate this quantity numerically. For $P_{\pi\pi}$, we have [44]

$$P_{\pi\pi} = -\frac{1}{6}\left(\frac{M_\pi^2}{F_\pi^2}\right)[g_1(M_\pi/T)]^2, \qquad (58)$$

$$g_1 = \int \frac{d^3p}{(2\pi)^3}\frac{1}{\omega_\pi\left(e^{\omega_\pi/T}-1\right)}, \qquad (59)$$

$$\omega_\pi = \sqrt{\mathbf{p}^2 + M_\pi^2}.$$

The pressure of the $i$-meson gas can be represented as

$$P_i = -g_i T\int \frac{d^3p}{(2\pi)^3}\ln(1-\exp(-\sqrt{\mathbf{p}^2+M_i^2}/T)). \qquad (60)$$

Under the assumption that $M_\pi = 140$ MeV, $M_K = 493$ MeV, and $F_\pi = 93$ MeV, numerical calculations yield

$$\delta_\theta(T < 140\ \text{MeV}) < 0.04. \qquad (61)$$

## 3. HADRON RESONANCE GAS AND NONPERTURBATIVE QCD VACUUM

Lattice calculations in finite-temperature QCD show that deconfinement and the restoration of chiral invariance occur at the same temperature and that, in the case of two light quark flavors ($N_f = 2$), the critical temperature falls within the interval $T_c \simeq 170-190$ MeV [49, 69]. From an analysis of QCD thermodynamics on a lattice and from experimental data on high-energy collisions, one can also deduce that the energy density of the system at the point of the quark–hadron phase transition is $\varepsilon(T_c) \sim 1-1.5\ \text{GeV/fm}^3$.

Recent numerical simulations on a lattice for $SU(3)$ gauge theory featuring no quarks and for $N_f = 2$ QCD revealed a strong suppression of the chromoelectric component and a weak growth of the chromomagnetic component of the gluon condensate with temperature above the critical temperature $T_c$ [70]. The temperature dependence of the gauge-invariant bilocal correlation function for the chromomagnetic-field strengths and the spacelike string tension $\sigma_s(T)$ was determined analytically in [22] (see Section 4 below). As a consequence, it was found that, in the region $T < 2T_c$, the chromomagnetic condensate grows slowly with temperature, $\langle H^2\rangle(T) = \langle H^2\rangle\coth(M/2T)$, where $M = 1/\xi_m \simeq 1.5$ GeV is the inverse magnetic correlation length, which is independent of temperature for $T < 2T_c$. In the region $T > 2T_c$, the amplitude of the chromomagnetic correlation function grows with temperature, $\langle H^2\rangle(T) \propto g^8(T)T^4$, while the correlation length decreases, $\xi_m(T) \propto 1/(g^2(T)T)$. This behavior of the magnetic correlation function explains the magnetic-confinement phenomenon within the stochastic-vacuum model. The resulting temperature dependence of the spacelike string tension agrees perfectly with lattice data [71–73] over the entire temperature range.

In view of all of the facts listed above, it is necessary that a unified approach within $N_f = 2$ QCD ensure the following at the critical point $T_c \sim 175-190$ MeV: an energy-density value of $\varepsilon_c \sim 1-1.5\ \text{GeV/fm}^3$; the vanishing of the quark condensate $\langle\bar{q}q\rangle(T)$; and the evaporation of only half of the gluon condensate (its chromoelectric component responsible for string formation and confinement), this being required for the magnetic-confinement phenomenon to be preserved.

In this section, we study the temperature properties of the quark and gluon condensates within the approach based on the description of the confining phase in terms of a hadron resonance gas and show that the aforementioned phenomena can be quantitatively explained within this approach upon appropriately taking into account the temperature shift of hadron masses [74].

We will consider QCD involving two light quark flavors. It was shown above that, if the pressure in the hadron phase, $P_h(T)$, is known, the temperature dependence of the quark condensate can be determined by using the Gell-Mann–Oakes–Renner relation. The result is

$$\frac{\langle\bar{q}q\rangle(T)}{\langle\bar{q}q\rangle} = 1 - \frac{1}{F_\pi^2}\frac{\partial P_h(T)}{\partial M_\pi^2}, \qquad (62)$$

where $F_\pi = 93$ MeV is the axial pion decay constant. The respective temperature dependence for the gluon condensate, $\langle G^2\rangle(T) \equiv \langle(gG^a_{\mu\nu})^2\rangle(T)$, was obtained above (see Subsections 2.1 and 2.2) in the form

$$\langle G^2\rangle(T) = \langle G^2\rangle + \frac{32\pi^2}{b}\left(4 - T\frac{\partial}{\partial T} - M_\pi^2\frac{\partial}{\partial M_\pi^2}\right)P_h(T),$$

where $b = 11N_c/3 - 2N_f/3 = 29/3$ and $\langle G^2\rangle \sim 0.5-1\ \text{GeV}^4$. The expressions for $\langle\bar{q}q\rangle(T)$ and $\langle G^2\rangle(T)$ in $N_f = 3$ QCD were obtained in [68]. Thus, we can state that, if the pressure $P_h$ as a function of temperature and the $\pi$-meson mass is known, one can find the temperature dependence of the quark and gluon condensates in the hadronic phase.

In order to describe QCD thermodynamics in the confining phase, we use the model of a hadron

resonance gas.[7] In this approach, the thermodynamic properties of the system are determined by the total pressure of relativistic Bose and Fermi gases describing thermal excitations of massive hadrons. The main motivation of employing the approach in question is that all significant degrees of freedom of strongly interacting matter have been thereby included in the consideration. Moreover, the use of the hadron-resonance spectrum takes effectively into account interaction between stable particles. At the same time, the description of multiparticle production in heavy-ion collisions within the hadron-resonance-gas model [76–78] leads to good agreement with experimental data. Thus, the pressure in the confining phase can be represented in the form

$$P_h = T\sum_i g_i\eta_i \int \frac{d^3p}{(2\pi)^3}\ln\left(1+\eta_i e^{-\omega_i/T}\right), \quad (63)$$

$$\omega_i = \sqrt{p^2+m_i^2},$$

$$\eta_i = \begin{cases} +1, & \text{fermions,} \\ -1, & \text{bosons,} \end{cases} \quad (64)$$

where $g_i$ is the spin–isospin degeneracy factor (for example, $g_\pi = 3$, $g_N = 8$, ...). The energy density $\varepsilon_h = T\partial P_h/\partial T - P_h$ in the hadronic phase is given by

$$\varepsilon_h = \sum_i g_i \int \frac{d^3p}{(2\pi)^3}\frac{\omega_i}{\exp(\omega_i/T)+\eta_i}. \quad (65)$$

In order to study the condensates in the confining phase quantitatively, it is necessary to know the pressure $P_h$ as a function of the light-quark mass (in the case of $N_f = 2$) or, which is the same, as a function of the $\pi$-meson mass. Within the hadron-resonance-gas model, this is equivalent to knowing the dependence of the masses of all resonances on the $\pi$-meson mass. This dependence was studied numerically in lattice calculations, and the five-parameter formula

$$m_i = N_u a_1 x\sqrt{\sigma} + \frac{m_h}{1+a_2x+a_3x^2+a_4x^3+a_5x^4}, \quad (66)$$

$$x \equiv M_\pi/\sqrt{\sigma},$$

$$a_1 = 0.51, \quad a_2 = a_1 N_u\sqrt{\sigma}/m_h,$$

$$a_3 = 0.115, \quad a_4 = -0.0223, \quad a_5 = 0.0028,$$

where $m_h$ is the physical hadron mass, $N_u$ is the number of light quark flavors ($N_u = 2$ for mesons, and $N_u = 3$ for baryons), and $\sigma = (0.42\ \text{GeV})^2$ is the string tension, was proposed in [79]. For a specific choice of parameters, this formula describes well the masses of all particles considered in [79].

One must also consider that the hadron masses change with increasing temperature. Within the finite-temperature conformally generalized nonlinear $\sigma$ model involving light and massive hadrons [67], it was shown that the temperature shift of the hadron masses can be taken into account with the aid of the substitutions

$$m_h \to m_h(\chi_T/\chi_0), \quad M_\pi \to M_\pi\sqrt{\chi_T/\chi_0}, \quad (67)$$

$$\chi_T/\chi_0 = \left(\langle G^2\rangle(T)/\langle G^2\rangle\right)^{1/4},$$

where $\chi$ is the dilaton field. That the dependence of the $\pi$-meson mass differs from those for other particles is a manifestation of its Goldstone nature.[8] In the chiral limit, $m_q \to 0$, the above relation for the hadron masses is a rigorous corollary of low-energy QCD theorems [27].

Formulas (31) and (62)–(67) determine the thermodynamic properties of the system in the hadronic phase and make it possible to calculate the quark and gluon condensates over the entire temperature range below the critical temperature $T_c$.

We take into account all hadron states of mass below 2.5 GeV for mesons and below 3.0 GeV for baryons. In all, there are 2078 states (with allowance for the degeneracy factors $g_i$). It is obvious that, at temperatures below the pion mass, $T < M_\pi = 140$ MeV, the main contribution to thermodynamic quantities comes from thermal excitations of $\pi$ mesons, because the remaining states are considerably heavier and are therefore exponentially suppressed by a Boltzmann factor proportional to $\exp\{-m_h/T\}$. At $T > M_\pi$, however, a large number of heavy states begin exerting a significant effect on the thermodynamics of the system. In Fig. 4, the pion contribution is depicted by the dashed curve. One can see that, up to the temperature of $T = 120$ MeV, the main contribution to the pressure $P_h$ does indeed come from pions. At higher temperatures, the main contribution to the pressure is due to all of the remaining hadron states. In Fig. 4, we also present lattice data from [81] for the pressure $P_h$ in $N_f = 2$ QCD. One can see that, in the temperature region $T < T_c$, the model of a hadron resonance gas describes correctly, with allowance for the chromoelectric mass shift, the growth of the pressure with temperature.

Figure 5 shows the energy density $\varepsilon_h$ as a function of temperature. The value of 1 GeV/fm$^3$, which corresponds to the estimates of the energy density upon the quark–hadron phase transition, is attained

[7] The model of a hadron resonance gas was proposed by Hagedorn back in 1965 [75] for describing hot strongly interacting matter.

[8] At a finite baryon density, similar relations were used in [80].

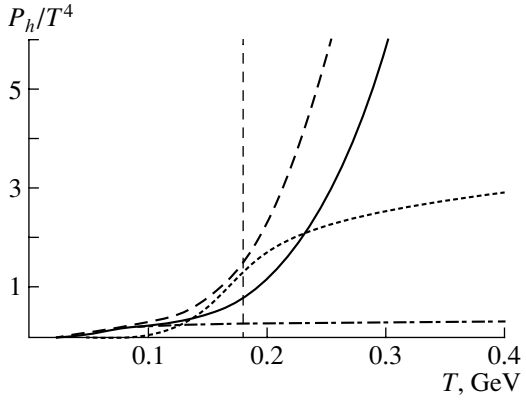


**Fig. 4.** Pressure $P_h/T^4$ as a function of temperature: (solid curve) zero-temperature hadron spectrum; (dashed curve) spectrum with allowance for the chromoelectric shift, $\chi_T/\chi_0 = 0.84$; (dash-dotted curve) results obtained with allowance for only pion excitations; and (dotted curve) lattice data borrowed from [81]. The vertical straight line corresponds to the critical temperature.

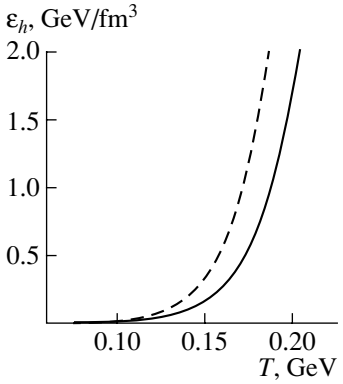


**Fig. 5.** Energy density $\varepsilon_h$ as a function of temperature: (solid curve) zero-temperature hadron spectrum and (dashed curve) spectrum with allowance for the chromoelectric shift, $\chi_T/\chi_0 = 0.84$.

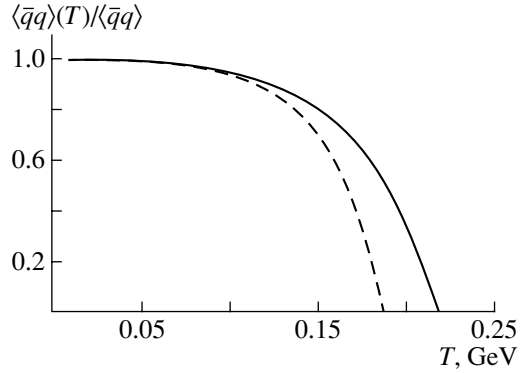


**Fig. 6.** Quark condensate $\langle\bar{q}q\rangle(T)/\langle\bar{q}q\rangle$ as a function of temperature: (solid curve) zero-temperature hadron spectrum and (dashed curve) spectrum with allowance for the chromoelectric shift, $\chi_T/\chi_0 = 0.84$.

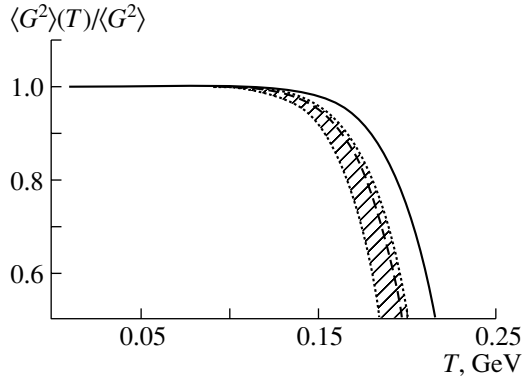


**Fig. 7.** Gluon condensate $\langle G^2\rangle(T)/\langle G^2\rangle$ as a function of temperature: (solid curve) zero-temperature hadron spectrum and (dashed curve) spectrum with allowance for the chromoelectric shift, $\chi_T/\chi_0 = 0.84$, at $\langle G^2\rangle = 0.87$ GeV$^4$ [82]. The shaded band corresponds to the values of $\langle G^2\rangle = 0.5$–$1.0$ GeV$^4$.

at $T \simeq 175$ MeV—that is, in the region of the phase-transition temperatures obtained in lattice calculations [16].

Figures 6 and 7 show the temperature dependences of, respectively, the quark and the gluon condensate. It is of importance that the temperature at which the quark condensate vanishes is identical to that at which half of the gluon condensate evaporates; with allowance for the chromoelectric shift of the hadron masses, this temperature is $T \simeq 190$ MeV.

Strictly speaking, the temperature shift of the gluon condensate must be found self-consistently (with the aid of the effective dilaton Lagrangian at $T \neq 0$) with allowance for the shift of the hadron masses (see [67]). However, the results of numerical calculations show that, up to temperatures of $T \sim M_\pi$, the gluon condensate decreases very slowly; at $T = M_\pi$, the temperature shift is $\Delta\langle G^2\rangle(T) \approx 0.02\langle G^2\rangle$. With increasing temperature, $T > M_\pi$, the gluon condensate decreases sharply; in a rather narrow temperature interval, $\Delta T \sim 50$ MeV, there occurs a main change in the gluon condensate by about 50%. In accordance with the above, we present numerical results obtained with allowance for the temperature shift of the hadron masses by 16% ($\chi_T/\chi_0 = 0.84 \simeq (0.5)^{1/4}$). We note that, even without allowance for the decrease in the mass $m_h$ with temperature, the quark condensate and half of the gluon condensate evaporate at the same temperature of $T \sim 215$ MeV.

## 4. MAGNETIC CONFINEMENT IN FINITE-TEMPERATURE $SU(N)$ YANG–MILLS THEORY

It is well known that, in $SU(N)$ non-Abelian gauge theory, the deconfining phase transition from

the low-temperature glueball phase to the hot-gluon-matter phase occurs at a finite temperature. Upon the phase transition at the critical point $T_c$, thermodynamic characteristics of the system—such as the energy density $\varepsilon$, the heat capacity, and the nonideality parameter $(\varepsilon - 3P)/T^4$—change their behavior substantially [15, 73, 83]. Moreover, it is known that, in non-Abelian gauge theories, a rearrangement of the nonperturbative gluon vacuum occurs upon the phase transition. For $SU(3)$ gauge theory and for QCD involving two quark flavors, the behavior of the gauge-invariant two-point correlation function for the gluon-field strengths in the temperature region around the phase-transition point was studied in [70] by means of numerical simulations on a lattice. The data obtained in this way clearly demonstrated a strong suppression of the electric component of the correlation function and the invariability of its magnetic component. The magnetic correlation function and the magnetic gluon condensate, which is related to it, undergo virtually no changes upon the transition from the confining ($T < T_c$) to the deconfining phase ($T > T_c$). In contrast to this, the electric component of the correlation function decreases fast at temperatures above $T_c$, so that the electric gluon condensate vanishes abruptly above the critical point $T_c$. This behavior was predicted theoretically in [23, 24].

At the phase-transition point, $T = T_c$, the physical string tension $\sigma(T)$ vanishes. However, it is well known that, in non-Abelian gauge theories, the Wilson loop $W(C)$ behaves, at any finite temperature, according to the area law for large spacelike contours $C$. This phenomenon is known as "magnetic" confinement, and a nonzero value of the spacelike string tension $\sigma_s(T)$ is associated with it [84]. The temperature dependence of the spacelike string tension, $\sigma_s(T)$, was studied on a lattice for $SU(2)$ [71] and $SU(3)$ [72, 73] purely gauge theories. In those studies, it was found that $\sigma_s(T)$ smoothly passes the phase-transition point, growing with increasing temperature. At temperatures above $2T_c$, the spacelike string tension approaches the scaling behavior $\sqrt{\sigma_s(T)}/g^2T \sim \text{const}$ [71–73], which is determined by the magnetic nonperturbative scale of about $g^2T$ [18, 85]. In addition, it is well established and is corroborated by lattice calculations that, at high temperatures, $4d$ $SU(N)$ gauge theory is described by an effective $3d$ gauge theory involving Higgs fields in the adjoint representation—this is the so-called dimensional reduction [86–91]. Moreover, numerous lattice calculations [71, 72, 91–94] revealed that scalar Higgs fields have but a slight effect on the physical properties of the magnetic sector of gauge theories. Thus, the behavior of some chromomagnetic quantities in the respective $3d$ theory is equivalent to the behavior of these quantities in $4d$ gauge theory at high temperatures. In particular, the spacelike string tension $\sigma_s(T)$ at high temperatures, $T \gg T_c$, agrees, to a high accuracy, with the string tension $\sigma_3$ calculated on a lattice in $3d$ gauge theory [72, 95].

In this section, we study the properties of the magnetic sector of $4d$ $SU(N)$ Yang–Mills theory at finite temperature. It is shown that the finite-temperature behavior of physical quantities in the magnetic sector is qualitatively different in the low- and the high-temperature region. By these, we mean, for the sake of definiteness, the region $T < 2T_c$ and the region $T > 2T_c$, respectively. The factor of "2" in the expression $2T_c$ should not be treated as an exact value, but, of course, it cannot be set to unity or, say, ten. (This issue is discussed below in Subsection 4.3.) In the low-temperature region, we will obtain analytic expressions for the magnetic correlation function, the magnetic condensate, and the spacelike string tension. We also calculate the relative contribution of nonzero Matsubara modes to $\sigma_s(T)$ and study the temperature interval where there occurs a transition to the description of the system in terms of $3d$ Yang–Mills theory. The temperature properties of the magnetic sector are studied at high temperatures ($T > 2T_c$).

### *4.1. Chromomagnetic Correlation Function and Spacelike String Tension*

Gauge-invariant nonlocal expectation values of the field strength (field correlation functions) play an important role in understanding nonperturbative QCD dynamics (see the review article of Di Giacomo *et al.* [33] and references therein). The two-point correlation function in the Yang–Mills vacuum is determined by the gauge-invariant expression [31, 32]

$$D_{\mu\nu\sigma\lambda}(x) = \langle g^2 \mathrm{tr} G_{\mu\nu}(x)\Phi(x,0)G_{\sigma\lambda}(0)\Phi(0,x)\rangle, \tag{68}$$

where $G_{\mu\nu} = t^a G^a_{\mu\nu}$ is the strength tensor of the gluon field, $t^a$ are the generators of the $SU(N)$ gauge group in the fundamental representation, $\mathrm{tr}\, t^a t^b = \delta^{ab}/2$, and

$$\Phi(x,y) = P\exp\left\{ig\int\limits_x^y dx_\mu A_\mu(x)\right\}, \qquad A_\mu = t^a A^a_\mu, \tag{69}$$

is the parallel transporter involving integration from $x$ to $y$ along a straight line. The phase factors $\Phi$ are introduced in order to ensure the gauge invariance of the correlation function. In general, the bilocal correlation function (68) is expressed in terms of two independent functions of $x^2$, $D(x^2)$ and $D_1(x^2)$ [31, 32].

At finite temperature, the Euclidean spacetime $O(4)$ symmetry is broken to the spatial $O(3)$ symmetry, and the bilocal correlation function can be described in terms of independent electric and magnetic correlation functions. Let us consider the magnetic component of the correlation function. In general, we have

$$D^H_{ik}(x) = \langle g^2 \mathrm{tr} H_i(x)\Phi(x,0)H_k(0)\Phi(0,x)\rangle \quad (70)$$
$$= \delta_{ik}(D^H + D^H_1) + \left(\delta_{ik} - \frac{x_i x_k}{\mathbf{x}^2}\right)\mathbf{x}^2\frac{\partial D^H_1}{\partial \mathbf{x}^2},$$

where $H_i = \varepsilon_{ijk}G_{jk}/2$ is the chromomagnetic-field operator.

By using the well-known method that was proposed in [31] and which is similar to the method for considering timelike Wilson loops, one can obtain the area law for a spacelike $W$ loop of dimension $L \gg \xi_m$ ($\xi_m$ is the magnetic correlation length); that is,

$$\langle W(C)\rangle_{\mathrm{spat}} \sim \exp\{-\sigma_s S_{\min}\}. \quad (71)$$

The spacelike string tension $\sigma_s$ can be expressed in terms of the magnetic correlation function as [19]

$$\sigma_s = \frac{1}{2}\int d^2x \langle g^2 \mathrm{tr} H_n(x)\Phi(x,0)H_n(0)\Phi(0,x)\rangle, \quad (72)$$

where $H_n$ is the magnetic-field component orthogonal to the spacelike plane of the contour.

Substituting (70) into (72) and considering that, taken together, the terms involving $D^H_1$ form a total derivative, we recast $\sigma_s$ into the form

$$\sigma_s = \frac{1}{2}\int d^2x D^H(x). \quad (73)$$

We note that only the term $D^H$ appears in this expression for $\sigma_s$. At $T = 0$, the spacelike string tension obviously coincides with the physical (timelike) one owing to $O(4)$ Euclidean invariance—that is, $\sigma_s = \sigma$, where $\sigma = (420\ \mathrm{MeV})^2$ is the standard value of the string tension in QCD.

The magnetic correlation function $D^H_{ik}$ can be written in a form including the matrix $U^{ab}_{\mathrm{adj}}$ in the adjoint representation. Since integration in (69) is performed along the straight line connecting the points $x$ and $y$, we have

$$D^H_{ik}(x,y) = \langle g^2 H^a_i(x)U^{ab}_{\mathrm{adj}}(x,y)H^b_k(y)\rangle, \quad (74)$$
$$U^{ab}_{\mathrm{adj}}(x,y) = 2\,\mathrm{tr}(t^a\Phi(x,y)t^b\Phi(y,x)).$$

This representation is used in lattice calculations of the correlation length (gluelump mass).

The functions $D^H$ and $D^H_1$ include both perturbative ($\propto 1/x^4$) and nonperturbative ($\propto \exp\{-|x|/\xi_m\}$) contributions, and only the nonperturbative components of the correlation functions contribute to the string tension (see a detailed discussion of these issues in [33]). The magnetic gluon condensate is also determined in terms of the nonperturbative contributions to the correlation function and is expressed in terms of the functions $D^H$ and $D^H_1$ at $x^2 = 0$. For the magnetic condensate, we have

$$D^H_{ii}(0) = \langle g^2 \mathrm{tr} H^2_i\rangle \equiv \langle H^2\rangle/2 = 3(D^H(0) + D^H_1(0)), \quad (75)$$

where $\langle H^2\rangle \equiv \langle (gH^a_i)^2\rangle$.

Lattice data obtained in [70] show that, at least in the region of measured temperatures, $T \le 1.5T_c$, the nonperturbative functions $D^H = B_0\exp\{-|x|/\xi_m\}$ and $D^H_1 = B_1\exp\{-|x|/\xi_m\}$ are virtually independent of temperature. The correlation length $\xi_m$ does not change and is equal to its value at $T = 0$, the contribution $D_1$ being strongly suppressed, $B_1 \approx 0.05B_0 \ll B_0$. Only $B_0$ grows slowly with $T$, and this leads to a slight increase in the condensate $\langle H^2\rangle(T)$ at low temperatures.

### 4.2. Low Temperatures

Let us consider the way in which one can derive analytic expressions for the magnetic correlation function and, accordingly, for the spacelike string tension and magnetic gluon condensate as functions of $T$. It is well known that the introduction of temperature $T$ for a quantum-field system in thermodynamic equilibrium is equivalent to the compactification of the Euclidean time coordinate $x_4$ with radius $\beta = 1/T$ and to imposing periodic boundary conditions (PBC) on boson fields (antiperiodic for fermions).

The temperature vacuum expectation values are defined in a standard way as

$$\langle\ldots\rangle_\beta = \frac{1}{Z_\beta}\int\limits_{\mathrm{PBC}}[DA]\ldots e^{-S_\beta[A]}, \quad (76)$$

where the partition function is

$$Z_\beta = \int\limits_{\mathrm{PBC}}[DA]e^{-S_\beta[A]}, \quad S_\beta = \int\limits_0^\beta dx_4\int d^3x L_{\mathrm{YM}}. \quad (77)$$

It was indicated above that, in the low-temperature region, the gluelump mass $M$ (inverse magnetic correlation length $1/\xi_m$) is independent of $T$. Thus, we can use the zero-temperature expression $D^H_{ik}(\mathbf{x},x_4)$ with periodic boundary conditions to obtain the finite-temperature correlation function $\mathcal{D}^H_{ik} = \langle g^2 H^a_i U^{ab}_{\mathrm{adj}} H^b_k\rangle_\beta$ constructed from the gauge fields

$H_i(\mathbf{x}, x_4)$ and the phase factor $U_{\rm adj}[A(\mathbf{x}, x_4)]$, which are taken at a specific time point $x_4$. We have

$$\mathcal{D}^H_{ik}(\mathbf{x}, \beta) = \sum_{n=-\infty}^{+\infty} D^H_{ik}(\mathbf{x}, n\beta). \qquad (78)$$

This relation is generally used in the case of noninteracting fields. For example, the finite-temperature Green's function for a free scalar field is defined as the sum over Matsubara frequencies of the propagator at $T = 0$. At the same time, interaction changes the particle mass, rendering it dependent on $T$. In the case under consideration, we take into account only a kinematical variation of the correlation function in the low-$T$ region and, accordingly, consider it as a periodized zero-temperature correlation function. It is physically clear that expression (78) is valid in the region $\beta > 2\xi_m$—that is, in the case where the two nearest exponential functions do not overlap. Further, we use the value of $1/\xi_m = 1.5$ GeV; therefore, the condition $T < \xi_m/2 = 750$ MeV must hold. In the low-temperature region, $T < 2T_c \sim 600$ MeV, which is considered in this section, this condition is always satisfied. Moreover, a comparison of the results of the ensuing calculations for physical quantities with the corresponding lattice data corroborates the validity of formula (78) for $T < 2T_c$.

If use is made of the Poisson summation formula

$$\frac{1}{\beta}\sum_{n=-\infty}^{+\infty} e^{i\omega_n x} = \sum_{n=-\infty}^{+\infty} \delta(x - n\beta), \qquad (79)$$
$$\omega_n = 2\pi n/\beta,$$

one can see that formula (78) represents the expansion of the magnetic correlation function $D$ in a Fourier series in Matsubara frequencies; that is,

$$\sum_n D^H_{ik}(\mathbf{x}, n\beta) = \frac{1}{\beta}\sum_n \tilde{D}^H_{ik}(\mathbf{x}, \omega_n), \qquad (80)$$

where $\tilde{D}(\omega) = \int d\tau \exp(i\omega\tau) D(\tau)$ is the Fourier transform of $D$.

Let us consider the function

$$f(\mathbf{x}, \beta) = \sum_{n=-\infty}^{+\infty} \exp\left(-M\sqrt{\mathbf{x}^2 + n^2\beta^2}\right), \qquad (81)$$
$$M \equiv 1/\xi_m.$$

In order to go over in (81) to summation over the Matsubara frequencies $\omega_n = 2\pi nT$, we use the integral representation for the exponential function and recast (81) into the form

$$f(\mathbf{x}, \beta) = \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{ds}{\sqrt{s}} \exp\left(-s - M^2\mathbf{x}^2/4s\right) \qquad (82)$$
$$\times \sum_{n=-\infty}^{+\infty} \exp\left(-M^2\beta^2 n^2/4s\right).$$

By using the summation property

$$\sum_{n=-\infty}^{+\infty} \exp\left(-\frac{b^2 n^2}{4s}\right) \qquad (83)$$
$$= \frac{2\sqrt{\pi s}}{b}\sum_{n=-\infty}^{+\infty} \exp\left(-\frac{4\pi^2 n^2}{b^2}s\right),$$

we represent (82) in the form

$$f(\mathbf{x}, \beta) = \frac{2}{M\beta}\sum_{n=-\infty}^{+\infty}\int_0^\infty ds e^{-as - b/s}, \qquad (84)$$

where $a = 1 + 4\pi^2 n^2/M^2\beta^2$ and $b = M^2\mathbf{x}^2/4$.

The integral on the right-hand side of (84) reduces to the modified Bessel function $K_1$. Finally, we have

$$f(\mathbf{x}, T) = 2MT|\mathbf{x}| \qquad (85)$$
$$\times \sum_{n=-\infty}^{+\infty} \frac{1}{\sqrt{M^2 + \omega_n^2}} K_1(|\mathbf{x}|\sqrt{M^2 + \omega_n^2}).$$

By considering the asymptotic behavior of the function $f$ at zero, we derive the temperature dependence of the chromomagnetic condensate: $\langle H^2\rangle(T) = \langle H^2\rangle f(|\mathbf{x}| \to 0, T)$. Considering that $K_1(x \to 0) \to 1/x$, we have

$$\frac{\langle H^2\rangle(T)}{\langle H^2\rangle} = 2MT\sum_{n=-\infty}^{+\infty} \frac{1}{M^2 + \omega_n^2} = \coth\left(\frac{M}{2T}\right). \qquad (86)$$

For $T \ll M$, this reduces to

$$\langle H^2\rangle(T)/\langle H^2\rangle = 1 + 2e^{-M/T} + O(e^{-2M/T}). \qquad (87)$$

This is in perfect agreement with lattice data from [70] on a weak growth of the chromomagnetic condensate at low temperatures, $T \le 1.5T_c$.

From expression (73) for $\sigma_s$ and formula (85) for the spacelike string tension, we obtain

$$\frac{\sigma_s(T)}{\sigma} = \frac{4T}{M}\sum_{n=-\infty}^{+\infty} \frac{1}{(1 + \omega_n^2/M^2)^2}, \qquad (88)$$

where we have used the normalization condition $\sigma_s(0) = \sigma$ and considered that

$$\int_0^\infty x^2 dx K_1(cx) = 2/c^3.$$

Upon calculating the sum on the right-hand side of (88), we obtain an expression for $\sigma_s(T)$ at low temperatures:

$$\frac{\sigma_s(T)}{\sigma} = \frac{\sinh(M/T) + M/T}{\cosh(M/T) - 1}. \qquad (89)$$

We have $\sigma_s(T \ll M)/\sigma = 1 + 2(M/T + 1)e^{-M/T} + O(e^{-2M/T})$.

From (88), the relative contribution $\Delta_n(T) \equiv \sigma_s^{n\neq 0}(T)/\sigma_s(T)$ from nonzero Matsubara frequencies to $\sigma_s(T)$ can be obtained as a function of $T$. Isolating the $n = 0$ term in (88), we obtain

$$\Delta_n(T) = 1 - \frac{4T}{M}\frac{\cosh(M/T) - 1}{\sinh(M/T) + M/T}. \qquad (90)$$

It is obvious that $\sigma_s^{n=0}(T) + \sigma_s^{n\neq 0}(T) = \sigma_s(T)$ and $\Delta_n(0) = 1$.

By using formula (88), one can obtain the contribution of the $n$th individual nonzero Matsubara mode in the form

$$\frac{\sigma_s^n(T)}{\sigma_s^{n=0}(T)} = \frac{M^4}{\omega_n^4}\frac{1}{(1 + M^2/\omega_n^2)^2}. \qquad (91)$$

Thus, we can see that, for $T > M/2\pi$, the contribution of the $n$th nonzero mode to $\sigma_s(T)$ dies out fast with increasing $T$ against the contribution of the zero mode, $n = 0$:

$$\frac{\sigma_s^n(T)}{\sigma_s^{n=0}(T)} = \left(\frac{M}{2\pi T}\right)^4 \frac{1}{n^4} - 2\left(\frac{M}{2\pi T}\right)^6 \frac{1}{n^6} + \ldots. \qquad (92)$$

### *4.3. High Temperatures*

By definition, $M = 1/\xi_m$ is the inverse magnetic correlation length. The correlation lengths at $T = 0$ were determined in various lattice calculations [82, 96, 97]. At the same time, $M$ is the mass of the lowest $1^{+-}$ magnetic gluelump. Gluelumps [98] are not physical objects, and their spectrum cannot be measured experimentally. However, they play an important role in nonperturbative QCD since their masses determine field correlation functions. In particular, the string tension at $T = 0$ is related to the gluelump masses. They were calculated analytically within QCD sum rules [99] and the QCD string model [100] and were numerically determined in various lattice calculations [101]. By using the entire body of these results, the correlation length $\xi_m$ was determined to fall between 0.1 and 0.2 fm.

We will use the value of $M = 1.5$ GeV for the inverse magnetic correlation length at $T = 0$. Figure 8 shows the function $\Delta_n(T)$ versus $T/T_c$ ($T_c = 270$ MeV for $SU(3)$ gauge theory). It can be seen that

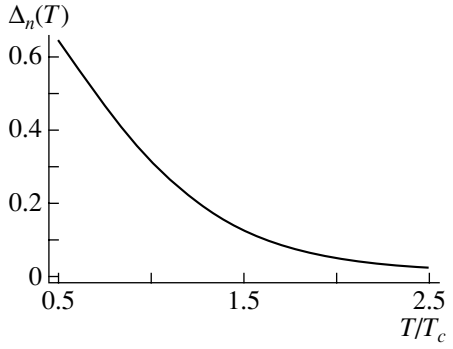


**Fig. 8.** Relative contribution $\Delta_n(T) \equiv \sigma_s^{n\neq 0}(T)/\sigma_s(T)$ from nonzero Matsubara frequencies to the spacelike string tension $\sigma_s(T)$ as a function of $T/T_c$ for $SU(3)$ gauge theory.

$\Delta_n(2T_c) \simeq 0.05$—that is, the contribution of nonstatic (nonzero) Matsubara frequencies to $\sigma_s$ is approximately 5%, which is less than the error of lattice calculations of the spacelike string tension [71–73]. Therefore, the main contribution to the dynamics of the magnetic sector of gauge theory at temperatures $T > 2T_c$ comes from the zero (static) Matsubara mode.

Further, it should be noted that, at any temperature $T$, $SU(N)$ gauge theory is in the magnetic-confinement phase; therefore, the magnetic correlation function at high temperatures does not change its exponential form. At the same time, it is physically clear that the two-point correlation function is determined by the amplitude and the correlation length, which depend on temperature. Therefore, the magnetic correlation function at high temperatures can be written in the form

$$\mathcal{D}_{ik}^H(\mathbf{x}, T) = \frac{1}{6}\delta_{ik}\langle H^2\rangle(T) \qquad (93)$$
$$\times \exp\left(-|\mathbf{x}|/\xi_m(T)\right) + O\left(\mathbf{x}^2\frac{\partial D_1^H}{\partial \mathbf{x}^2}\right).$$

It was mentioned above that the $O(\ldots)$ term does not contribute to the chromomagnetic condensate or the spacelike string tension. From (72) and (93), we then obtain

$$\sigma_s(T) = \frac{\pi}{6}\langle H^2\rangle(T)\xi_m^2(T). \qquad (94)$$

It was shown in [71–73] that, at $T > 2T_c$, the temperature dependence of the spacelike string tension, $\sigma_s(T)$, agrees, to a high accuracy, with the behavior of the string tension $\sigma_3$ in $3d$ Yang–Mills theory. Three-dimensional Yang–Mills theory is a superrenormalized theory, and all physical quantities are determined by a single dimensional coupling constant $g_3^2 = g^2T$. In [71–73], it was found that

$$\sqrt{\sigma_s(T)} = c_\sigma g^2(T)T, \qquad (95)$$

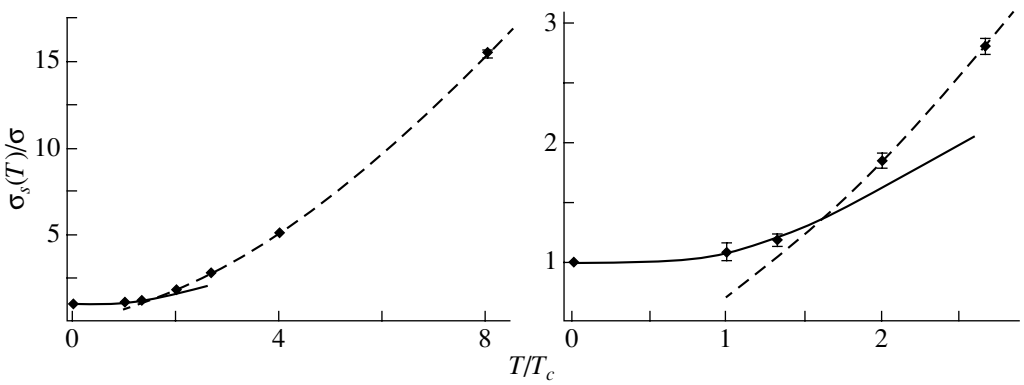


**Fig. 9.** Spacelike string tension $\sigma_s(T)/\sigma$ for $SU(2)$ gauge theory as a function of $T/T_c$: (solid curve) results of calculations by formula (89) and (dashed curve) results of calculations by formulas (95) and (96). The displayed lattice data (points) were borrowed from [71].

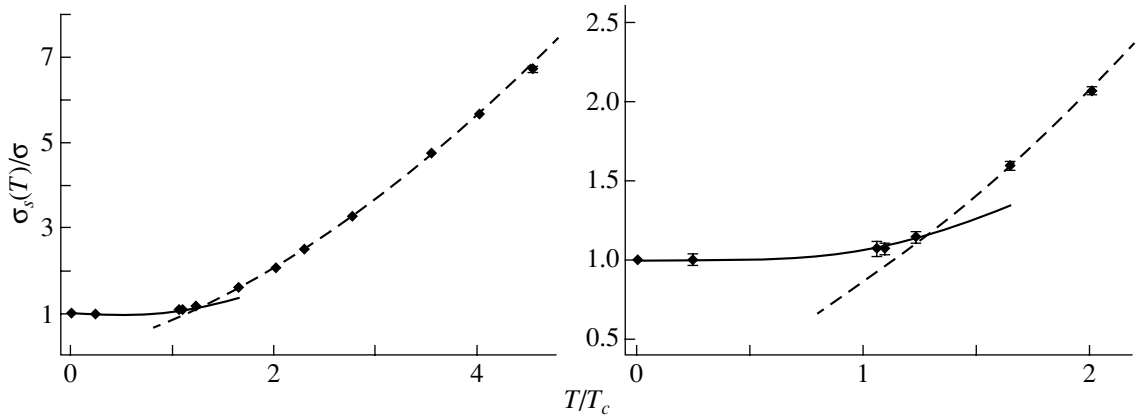


**Fig. 10.** Spacelike string tension $\sigma_s(T)/\sigma$ for $SU(3)$ gauge theory as a function of $T/T_c$: (solid curve) results of calculations by formula (89) and (dashed curve) results of calculations by formulas (95) and (96). The displayed lattice data (points) were borrowed from [73].

where, for $g^2$, use was made of the two-loop expression

$$g^{-2}(T) = 2b_0 \ln \frac{T}{\Lambda_\sigma} + \frac{b_1}{b_0} \ln \left( 2 \ln \frac{T}{\Lambda_\sigma} \right), \qquad (96)$$

$$b_0 = \frac{11N}{48\pi^2}, \quad b_1 = \frac{34}{3} \left( \frac{N}{16\pi^2} \right)^2.$$

The constants $c_\sigma$ and $\Lambda_\sigma$ were evaluated on the basis of a comparison with lattice data on $\sigma_s(T)$ by using a two-parameter fit. We have $c_\sigma = 0.369 \pm 0.014$ and $\Lambda_\sigma = (0.076 \pm 0.013)T_c$ for $SU(2)$ gauge theory [71] and $c_\sigma = 0.566 \pm 0.013$ and $\Lambda_\sigma = (0.104 \pm 0.009)T_c$ for $SU(3)$ gauge theory [73]. Figures 9 and 10 show the graphs of $\sigma_s(T)/\sigma$ for, respectively, $SU(2)$ and $SU(3)$ gauge theories versus $T/T_c$ ($T_c^{SU(2)} = 290$ MeV, $T_c^{SU(3)} = 270$ MeV). The solid and dashed curves represent the results of the calculations by, respectively, formula (89) and formulas (95) and (96). The displayed lattice data were borrowed from [71, 73].

From Eq. (94), one can determine $\langle H^2 \rangle(T)$. At high temperatures, the inverse magnetic correlation length $1/\xi_m$ behaves as

$$1/\xi_m(T) = c_m g^2(T) T, \qquad (97)$$

where $c_m$ is a constant. By using expressions (94), (95), and (97), we find for the temperature dependence of the nonperturbative chromomagnetic condensate that

$$\langle H^2 \rangle(T) = c_H g^8(T) T^4, \qquad (98)$$

where

$$c_H = \frac{6}{\pi} c_\sigma^2 c_m^2. \qquad (99)$$

One can determine the constant $c_H$ from expression (99). For example, we have $c_\sigma \approx 0.37$ [71] and $c_m \approx 0.92$ [102] in $SU(2)$ Yang–Mills theory; therefore, the result is $c_H \approx 0.22$. It should be noted, however, that the quantity $c_H$ must be obtained from a numerical lattice calculation of the thermal magnetic correlation function.

The result presented in (98) is natural from the point of view of $3d$ gauge theory. It was indicated above that, at high temperatures, the thermal behavior of physical quantities within $4d$ $SU(N)$ gauge theory coincides with their behavior in $3d$ $SU(N)$ gauge theory. Therefore, the nonperturbative chromomagnetic condensate is determined by the dimensional coupling constant $g_3$, and we have $\langle H^2\rangle(T) = \text{const} \times g_3^8$. Strictly speaking, physical expectation values in superrenormalized $3d$ Yang–Mills theory either are zero or are determined by the corresponding power of the dimensional coupling constant $g_3$. The vanishing of the magnetic condensate at high temperatures leads to $\sigma_s = 0$, this contradicting lattice data on magnetic confinement. It follows that, at high temperatures (in the scaling region for the spacelike string tension), the chromomagnetic correlation function has the form

$$\mathcal{D}_{ik}^H(\mathbf{x}, T) = \frac{1}{6}\delta_{ik} c_H g^8(T) T^4 \exp(-c_m g^2(T) T|\mathbf{x}|). \tag{100}$$

In the high-temperature region, its amplitude grows in proportion to $T^4$ with increasing temperature, while the correlation length decreases in proportion to $1/T$.

By using the above results, one can estimate the region of temperatures at which there occur a change in the regime in the magnetic sector of $SU(N)$ gauge theory and the transition to the description of dynamics in terms of static correlation functions—that is, to $3d$ gauge theory. Physically, it is clear that, if the compactification radius $\beta$ (inverse temperature $1/T$) becomes much less than the characteristic dimension $l$ of the system (it is determined, for example, by the magnetic-string thickness or by the characteristic dimension of a magnetic gluelump, $l = 1/\sqrt{\sigma_s(T)}$), the correlation functions cease to feel the time coordinate $x_4$, becoming static ones.

Thus, we can estimate the temperature at which the regime changes by employing the condition $T = \sqrt{\sigma_s(T)}$. For the sake of definiteness, we consider $SU(3)$ gauge theory. In the region of low temperatures, we find $T_- = 1.42T_c$ from the numerical solution to the equation $T_- = \sqrt{\sigma_s(T_-)}$ with $\sigma_s(T)$ determined by expression (89). At the same time, one can see that, at high temperatures, the solution to the equation $T_+ = \sqrt{\sigma_s(T_+)}$ where $\sigma_s(T)$ is given by relations (95) and (96) yields $T_+ = 2.74T_c$. Thus, we can state that, in the region $T_- < T < T_+$, there occurs a change in the regime and a transition to reduced $3d$ Yang–Mills theory. Therefore, the commonly accepted value of $2T_c$ is quite justified from the physical point of view (see above).

The growth of the magnetic condensate with increasing temperature, $\langle H^2\rangle(T) \propto T^4$, is thermodynamically advantageous for a quantum-field system. The vacuum energy density is related through the anomaly in the trace of the energy–momentum tensor to the magnetic condensate by the equation $\varepsilon_{\text{vac}} = \langle\theta_{00}\rangle = -(b/64\pi^2)\langle H^2\rangle$, where $b = 11N/3$, and an increase in $\langle H^2\rangle$ reduces the vacuum energy density.

## 5. BILOCAL CORRELATION FUNCTION FOR GLUON-FIELD STRENGTHS IN THE INSTANTON VACUUM AT FINITE TEMPERATURE

In this section, we study the contribution of instantons to the bilocal correlation function at finite temperature [103, 104]. We show that the correlation length for the caloron bilocal correlation function (that is, the bilocal correlation function for instantons at finite temperature) depends quite strongly on temperature in the lowest order in the caloron density. This is not compatible with data from lattice calculations—it follows from these data that the correlation length for the chromomagnetic-field bilocal correlation function is independent, within the errors, of temperature over the entire temperature range from zero to $T_c$ [70]. This discrepancy between the results makes it possible to set an upper bound on the possible instanton density in QCD.

We have derived above the analytic expression for the temperature dependence of the chromomagnetic bilocal correlation function. It was found that the chromomagnetic condensate grows slowly with increasing temperature up to $2T_c$, while the correlation length $\xi_m(T)$ is independent of $T$ in this temperature range. At high temperatures, $T > 2T_c$, the condensate behaves as $\langle H^2\rangle(T) \propto g^8(T)T^4$, and the magnetic correlation length decreases with increasing temperature according to the law $\xi_m(T) \propto 1/(g^2(T)T)$. There exists a simple physical explanation for the temperature independence of the correlation length at $T \le T_c$. As was mentioned in Section 4, the inverse magnetic correlation length $M = 1/\xi_m$ is the mass of the lightest $1^{+-}$ magnetic gluelump, $M \approx 1.5$ GeV. Since the phase-transition temperature is much smaller than the gluelump mass, $T_c \ll M$, it is physically clear that the temperature dependence of the gluelump mass and, hence, of the correlation length must be very weak over the low-temperature region ($T < T_c \simeq 300$ MeV), which is being considered at the moment. This is corroborated by analytic [22] and lattice [70] calculations. Concurrently, it should be noted that the magnetic correlation length does not change its behavior at the phase-transition point. It is well known that the vacuum expectation value of the Polyakov loop

is the order parameter for the deconfining phase transition; therefore, it is the correlation function for the Polyakov loops, $\langle L(\mathbf{x})L^\dagger(0)\rangle$, that changes its behavior significantly at the temperature $T = T_c$.

At finite temperature, $O(4)$ Euclidean spacetime symmetry is broken to $O(3)$ symmetry, and the bilocal correlation function is described by independent electric and magnetic correlation functions, which can be expressed in terms of four independent functions $D^E(x^2)$, $D_1^E(x^2)$, $D^H(x^2)$, and $D_1^H(x^2)$ as

$$\left\langle g^2\mathrm{tr}\left(E_i(x)\Phi(x,y)E_j(y)\Phi^\dagger(x,y)\right)\right\rangle \tag{101}$$
$$= \delta_{ij}\left(D^E + D_1^E + z_4^2\frac{\partial D_1^E}{\partial z^2}\right) + z_i z_j\frac{\partial D_1^E}{\partial z^2},$$
$$\left\langle g^2\mathrm{tr}\left(H_i(x)\Phi(x,y)H_j(y)\Phi^\dagger(x,y)\right)\right\rangle$$
$$= \delta_{ij}\left(D^H + D_1^H + z^2\frac{\partial D_1^H}{\partial z^2}\right) - z_i z_j\frac{\partial D_1^H}{\partial z^2},$$

where $H_i = \varepsilon_{ijk}F_{jk}/2$ is the chromomagnetic-field operator, $E_i = F_{4i}$ is the chromoelectric-field operator, $F_{\mu\nu} = F^a_{\mu\nu}t^a$ is the gluon-field-strength tensor, and

$$\Phi(x,y) = P\exp\left(igz^\mu\int\limits_0^1 dsA_\mu(y+sz)\right) \tag{102}$$

is the parallel transporter along the straight line connecting the points $x$ and $y$; here, $z = x - y$. In the case where these correlation functions are calculated for a self-dual field, it is obvious that $D_1^E = D_1^H = 0$ and $D^E = D^H$ (see [105]). Therefore, it is sufficient to consider only the chromomagnetic correlation function.

The very fact that, for instantons, we have $D^E = D^H$ implies that the nonperturbative vacuum cannot be a simple superposition of instanton–antiinstanton field configurations because, as is known from lattice calculations, the chromomagnetic and chromoelectric condensates manifest different temperature dependences. In particular, the chromoelectric condensate decreases sharply at the temperature of the deconfining phase transition, while the chromomagnetic condensate undergoes virtually no changes [70].

The one-instanton solution at finite temperature, the so-called caloron, can be represented in the well-known form [106]

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\varepsilon^{abc}A^b_\mu A^c_\nu,$$
$$gH^a_i = -\delta_{ai}g_1(x) + \varepsilon_{aij}\frac{x_j}{r}g_2(x) + \frac{x_a x_i}{r^2}g_3(x),$$
$$E^a_i = H^a_i,$$

$$g_1(x) = \frac{1}{r}\frac{1}{\Pi}\frac{\partial\Pi}{\partial r} + \frac{1}{\Pi}\frac{\partial^2\Pi}{\partial\tau^2} \tag{103}$$
$$-\frac{1}{\Pi^2}\left(\frac{\partial\Pi}{\partial\tau}\right)^2 + \frac{1}{\Pi^2}\left(\frac{\partial\Pi}{\partial r}\right)^2,$$
$$g_2(x) = -\frac{2}{\Pi^2}\frac{\partial\Pi}{\partial r}\frac{\partial\Pi}{\partial\tau} + \frac{1}{\Pi}\frac{\partial^2\Pi}{\partial r\partial\tau},$$
$$g_3(x) = \frac{2}{\Pi^2}\left(\frac{\partial\Pi}{\partial r}\right)^2 - \frac{1}{\Pi}\frac{\partial^2\Pi}{\partial r^2} + \frac{1}{r}\frac{1}{\Pi}\frac{\partial\Pi}{\partial r}.$$

The chromoelectric field is equal to the chromomagnetic field by virtue of the self-duality of the caloron field, $F_{\mu\nu} = \tilde{F}_{\mu\nu}$.

Further, we must calculate the parallel transporter for the caloron. The result is

$$\Phi(x,y) = P\exp\left(igz^\mu\int\limits_0^1 dsA_\mu(y+sz)\right) \tag{104}$$
$$= P\exp\left(-iz^\mu\bar\eta^a_{\mu\nu}t^a\int\limits_0^1 ds\partial_\nu\ln\Pi(y+sz)\right)$$
$$= P\exp\left(-it^a\int\limits_0^1 ds\left[\bar\eta^a_{ij}z_i\partial_j\ln\Pi + \bar\eta^a_{i4}z_i\partial_4\ln\Pi\right.\right.$$
$$\left.\left. + \bar\eta^a_{4j}z_4\partial_j\ln\Pi\right]\right),$$
$$z = x - y.$$

With an eye to a comparison with data from lattice calculations, we consider the case of $z_4 = 0$. We then have

$$\Phi(x,y) \tag{105}$$
$$= P\exp\left(-it^a\int\limits_0^1 ds\left[\bar\eta^a_{ij}z_i\partial_j\ln\Pi + \bar\eta^a_{i4}z_i\partial_4\ln\Pi\right]\right).$$

In general, we must evaluate the path-ordered integral. However, there exist special cases where the operators in the exponent at different values of $s$ commute, so that the evaluation of the integral reduces to the calculation of an ordinary exponential. It is obvious that this occurs at $\tau = 0$, $\beta/2$, and $\beta$ since $\partial_4\ln\Pi(r,\tau = 0;\beta/2;\beta) = 0$. The single-instanton field is yet another case where this is so since $\partial_4\ln\Pi^{\mathrm{inst}} = \tau(1/r)(\partial\ln\Pi^{\mathrm{inst}}/\partial r)$.

In order to calculate the contribution of a dilute caloron gas to the two-point correlation function in the lowest order in the density, it is necessary to calculate the contribution of a single caloron and to average the result over its position or, which is the same, to average it over $y$ at a fixed value of $z = x - y$.

Averaging over $y$ reduces to evaluating the integral

$$\int d^4y = \int d^3y \int_0^\beta dy_4.$$

Averaging over the three-dimensional vector $\mathbf{y}$ does not involve any difficulties. As to the integral with respect to $y_4$, it cannot be calculated since, by virtue of the aforesaid, the correlation function can be calculated numerically only at $\tau = 0$, $\beta/2$, and $\beta$. Nevertheless, we will see from a comparison with the single-instanton case that, at low temperatures, it is sufficient to perform calculations at $\tau = 0$ and that, at high temperatures, the correlation length at $\tau = 0$ coincides with that at $\tau = \beta/2$, so that averaging over $\tau$ is trivial.

Let us determine the functions $d$ and $d_1$. We have

$$\left\langle g^2 \mathrm{tr}\left(H_i(x)\Phi(x,y)H_j(y)\Phi^\dagger(x,y)\right)\right\rangle_{\mathbf{y}} \quad (106)$$
$$= \delta_{ij}\left(d + d_1 + z^2\frac{\partial d_1}{\partial z^2}\right) - z_i z_j \frac{\partial d_1}{\partial z^2},$$

where $z_4 = 0$ and the functions $d$ and $d_1$ depend on $z^2$ and $y_4 \equiv \tau$: $d = d(\tau, z^2)$ and $d_1 = d_1(\tau, z^2)$. Averaging over $y_4$, we have

$$\int_0^\beta d\tau d(\tau, z^2) = D^H(z^2), \quad (107)$$

$$\int_0^\beta d\tau d_1(\tau, z^2) = D_1^H(z^2) = 0,$$

$$D^H(z^2 = 0) = 4\pi^2/3.$$

The last equality follows from the relation

$$g^2 \int d^3y \int_0^\beta dy_4 \left(F^a_{\mu\nu}\right)^2 = 32\pi^2$$

(by virtue of the self-duality of the field, this integral is proportional to the topological charge $\sim \int d^4y F\tilde{F}$).

Let us define the correlation length $\xi_\tau$ as the coefficient in the exponent of the exponential function that provides the best fit to the function $d(\tau, z^2)$; that is,

$$d(\tau, z^2) \simeq e^{-|z|/\xi_\tau}. \quad (108)$$

As was discussed above, we will calculate only $d(\tau = 0, z^2)$ and $d(\tau = \beta/2, z^2)$. However, this information is sufficient for determining the correlation length for the function $D^H(z^2)$. Indeed, averaging over $\tau$ shows that the functions $d(\tau = 0, z^2)/d(\tau = 0, z^2 = 0)$ and $D^H(z^2)/D^H(z^2 = 0)$, which are normalized to unity, are nearly coincident. The point is that an instanton is a well-localized field configuration, its field decreasing fast away from its center. This is illustrated in Fig. 11, where the function $d^{\mathrm{inst}}(\tau, z^2)$ calculated for a single instanton is depicted at several values of $\tau$. Since the amplitude of the function $d(\tau, z^2)$ decreases fast with increasing $\tau$, the profile of the function $D^H(z^2)$ is actually determined by the function $d(\tau = 0, z^2)$. This is also true for a caloron in the region $\beta \gg \rho$, because, in this case, instantons in a chain along the time axis are well separated. At smaller values of $\beta$—that is, at $\beta \leq 2\rho$—the functions $d(\tau = 0, z^2)$ and $d(\tau = \beta/2, z^2)$ have virtually the same correlation length (see Fig. 12). Thus, av-

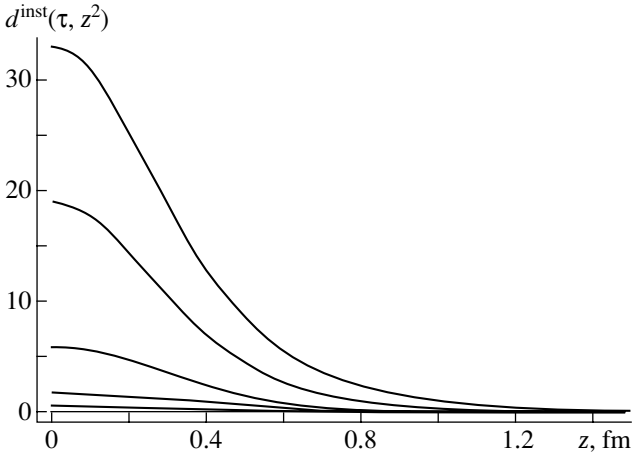


**Fig. 11.** Function $d^{\mathrm{inst}}(\tau, z^2)$ at $\tau = 0$, 0.15, 0.3, 0.45, 0.6 fm for $\rho = 0.3$ fm (the upper curve corresponds to the value of $\tau = 0$).

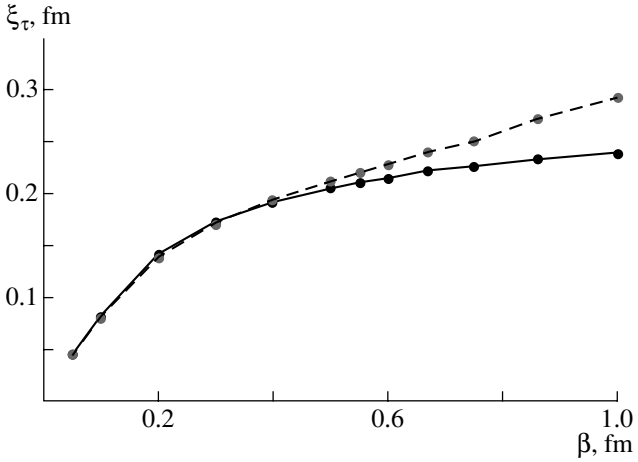


**Fig. 12.** Correlation length $\xi_\tau$ for the functions $d(\tau = 0, z^2)$ and $d(\tau = \beta/2, z^2)$: (solid curve) $\xi_{\tau=0}(\beta)$ and (dashed curve) $\xi_{\tau=\beta/2}(\beta)$; $\rho = 0.3$ fm.

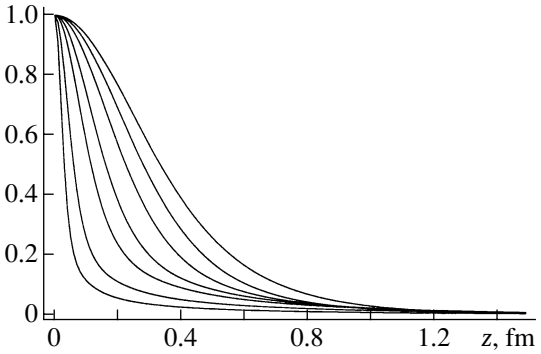


**Fig. 13.** Ratio $d(\tau=0,z^2)/d(\tau=0,z^2=0)$ at $\beta=0.05, 0.1, 0.2, 0.3, 0.5, 0.86, \infty$ fm$^{-1}$ for $\rho=0.3$ fm. The maximum correlation length (upper curve) corresponds to zero temperature ($\beta=\infty$)—that is, to the instanton contribution.

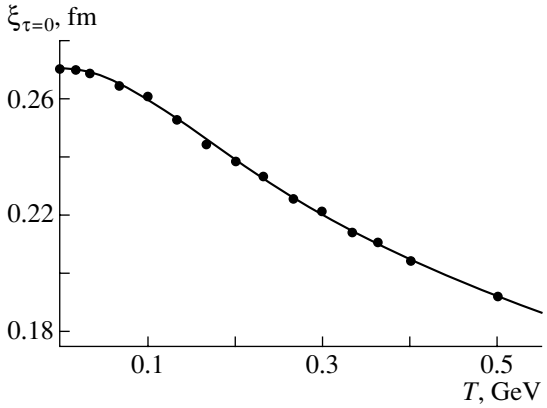


**Fig. 14.** Correlation length $\xi_{\tau=0}(T)$ as a function of temperature for $\rho=0.3$ fm.

eraging over $\tau$ is trivial and merely leads to a change in the amplitude, which is determined by the condition $D^H(z^2=0)=4\pi^2/3$.

Taking into account the aforesaid, we will consider only the function $d(\tau=0,z^2)$, whose correlation length is close to the correlation length for the function $D^H(z^2)$ at all values of $\beta$. At $\tau=0, \beta/2$, and $\beta$, the chromomagnetic correlation function has the form

$$\left\langle g^2 \mathrm{tr}\left(H_i(x)\Phi(x,y)H_j(y)\Phi^\dagger(x,y)\right)\right\rangle_{\mathbf{y}} \qquad (109)$$
$$= g^2\int d^3\mathbf{y}\left(H_i^a(x)U^{ab}(x,y)H_j^b(y)\right),$$
$$U^{ab}(x,y)=\mathrm{tr}\left(t^a\Phi(x,y)t^b\Phi^\dagger(x,y)\right)$$
$$=\frac{1}{2}\delta^{ab}\cos(2\phi)-\frac{1}{2}\varepsilon^{abc}n^c\sin(2\phi)+n^an^b\sin^2(\phi),$$
$$\phi=\frac{1}{2}n^a\bar{\eta}^a_{ij}z_iy_j\int\limits_0^1 ds\left(\frac{1}{r}\frac{\partial\ln\Pi(y+sz)}{\partial r}\right).$$

The correlation function calculated by formula (109) at various values of $\beta$ is shown in Fig. 13. The correlation length as a function of temperature is given in Fig. 14.

The bilocal correlation function at finite temperature was calculated on a lattice in [70], where it was found that the bilocal correlation function for chromomagnetic fields is virtually independent of temperature over the entire range from zero to $T_c$ (its temperature dependence is weak above the critical temperature as well). At the same time, one can see from Fig. 14 that, in the dilute caloron gas, the correlation length changes by about 20% as the temperature increases from zero to 300 MeV. Thus, one can conclude that an instanton gas does not correctly describe the true vacuum of gluodynamics. There exist a few possible explanations of lattice data.

(i) The instanton density is much less than 1 fm$^{-4}$, and the main contribution to the correlation function comes from other nonperturbative fields, which ensure confinement. This pattern is in accord with the stochastic vacuum model.

(ii) The instanton density is so high that higher order corrections in density change the result qualitatively, leading to a temperature-independent correlation length.

(iii) The configuration of the instanton ensemble is rearranged in such a way that instantons and anti-instantons form molecules, and the contribution of these instanton–anti-instanton molecules to the bilocal correlation function differs significantly from the contribution of the instanton gas. This binding into pairs can have some bearing on the deconfining phase transition. For example, this occurs in the three-dimensional model involving a Higgs field in the adjoint representation, for which it was shown that, in the deconfining phase, instantons and anti-instantons are bound into molecules [107–109]. The formation of instanton–anti-instanton molecules in the QCD vacuum and its relation to the phase transition restoring chiral symmetry are discussed in the review article of Schafer and Shuryak [110].

(iv) In recent years, a totally different scenario has been discussed for the caloron vacuum in the confining phase ($T<T_c$). There exist field configurations of nontrivial holonomy—the so-called KvBLL solution [111]. This scenario suggests [112] that a caloron may break up into a dion–antidion pair. Lattice data confirming the existence of these field configurations for the $SU(2)$ and the $SU(3)$ group are presented in [113, 114]. It is obvious that, in this scenario, the behavior of the correlation length in the nonperturbative vacuum is more complicated than that in a pure caloron gas studied above.

In principle, one can consider yet another possibility where the instanton-size distribution also depends on temperature, and this dependence is such that the resulting correlation length is independent of $T$. However, there are lattice data [115] according to which the instanton-size distribution does not change within the errors over the entire temperature range from zero to $T_c$.

## 6. CONCLUSION

We have considered a number of the most characteristic nonperturbative phenomena in finite-temperature QCD. We have not touched upon problems such as chiral and deconfining phase transitions, possible bound states in the high-temperature phase, and the behavior of condensates of higher dimensions (in particular, a mixed quark–gluon condensate). These and other problems require a dedicated detailed consideration. In conclusion, we will briefly formulate the basic results of this study.

On the basis of an analysis of the QCD partition function and its renormalization-group properties, we have obtained an equation that relates the anomalous contribution to the trace of the energy–momentum tensor to the QCD thermodynamic pressure. We have also derived analytic expressions for the temperature dependences of the quark and gluon condensates at low temperatures, $T < M_\pi$. A low-temperature relation has been found for QCD—that is, it has been shown that, in the low-temperature region ($T < M_\pi$), the temperature derivatives of the anomalous and normal (quark-mass term) contributions to the trace of the energy–momentum tensor in QCD are equal to each other. The leading corrections to this relation, which are associated with the excitation of massive $K$ and $\eta$ mesons, have been obtained. An analytic expression has been derived for the temperature dependence of the strange quark condensate $\langle\bar{s}s\rangle$, and it has been shown that, with increasing temperature, it decreases much more slowly than the light-quark condensate $\langle\bar{u}u\rangle = \langle\bar{d}d\rangle$.

The nonperturbative QCD vacuum at finite temperature has been explored within the model of a hadron resonance gas. The temperature dependences of the quark and gluon condensates have been obtained in the confining phase, and it has been shown that the quark condensate and half (chromoelectric component) of the gluon condensate evaporate at the same temperature corresponding to the quark–hadron phase transition. It has been shown that, upon the phase transition, the chromoelectric condensate vanishes, while the chromomagnetic condensate undergoes virtually no changes. This fact corroborates the magnetic-confinement pattern. At the phase-transition temperature, the energy density of the hadron resonance gas is $\varepsilon_h(T_c) \sim 1.5$ GeV/fm$^3$. Upon taking into account the chromoelectric shift of the hadron masses, the critical temperature becomes $T_c \simeq 190$ MeV.

The magnetic-confinement phenomenon has been studied in $4d$ $SU(N)$ Yang–Mills theory at finite temperature. The temperature dependence of the gauge-invariant bilocal correlation function for the chromomagnetic-field strength and the spacelike string tension $\sigma_s(T)$ has been derived in an analytic form. It has been shown that, in the region $T < 2T_c$, the chromomagnetic condensate grows slowly with temperature, $\langle H^2\rangle(T) = \langle H^2\rangle \coth(M/2T)$, where $M = 1/\xi_m \simeq 1.5$ GeV is the inverse magnetic correlation length, which is independent of temperature in this region. At temperatures $T > 2T_c$, the amplitude of the magnetic correlation function grows with increasing temperature according to the law $\langle H^2\rangle(T) \propto g^8(T)T^4$, while the correlation length decreases, $\xi_m(T) \propto 1/(g^2(T)T)$. This behavior of the magnetic correlation function explains the magnetic-confinement phenomenon in $4d$ $SU(N)$ Yang–Mills theory at finite temperature. The resulting temperature dependence of the spacelike string tension is in perfect agreement with lattice data over the entire temperature range.

The contribution of the instanton gas to the bilocal correlation function has been calculated at finite temperature. A comparison of the temperature dependence of the correlation length with the corresponding results of lattice calculations enables one to conclude that, if the instanton density is not overly high, so that corrections to the dilute-gas approximation are insignificant, the instanton density must be much less than $n = 1$ fm$^{-4}$, a value that is adopted in the instanton-liquid model.

## ACKNOWLEDGMENTS

I am grateful to Yu. A. Simonov for numerous discussions on the problems considered here and to A.V. Borisov, B.L. Ioffe, A.B. Kaidalov, V.A. Novikov, L.B. Okun, V.A. Rubakov, and S.M. Fedorov for stimulating discussions and comments.

This work was supported in part by the Council of the President of the Russian Federation for Support of Young Russian Scientists and Leading Scientific Schools (project no. 1774.2003.2) and by the federal program of the Ministry of Industry, Science, and Technology of the Russian Federation (project no. 40.052.1.1.1112).

## REFERENCES


1. J. C. Collins and M. J. Perry, Phys. Rev. Lett. **34**, 1353 (1975).
2. N. Cabibbo and G. Parisi, Phys. Lett. B **59B**, 67 (1975).
3. K. Rajagopal and F. Wilczek, hep-ph/0011333.
4. G. Nardulli, Riv. Nuovo Cimento **25** (3), 1 (2002).
5. D. H. Rischke, Prog. Part. Nucl. Phys. **52**, 197 (2004).
6. D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).
7. H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
8. A. M. Polyakov, Phys. Lett. B **59B**, 79 (1975).
9. A. A. Belavin, A. M. Polyakov, A. S. Shvarts, and Yu. S. Tyupkin, Phys. Lett. B **59B**, 85 (1975).
10. G. 't Hooft, Phys. Rev. D **14**, 3432 (1976).
11. C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. B **63B**, 334 (1976).
12. C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978); **19**, 1826 (1979).
13. J. P. Blaizot and E. Iancu, Phys. Rep. **359**, 355 (2002).
14. U. Kraemmer and A. Rebhan, Rep. Prog. Phys. **67**, 351 (2004).
15. F. Karsch, Nucl. Phys. A **698**, 199 (2002).
16. F. Karsch, Lect. Notes Phys. **583**, 209 (2002).
17. E. Laermann and O. Philipsen, hep-ph/0303042.
18. A. D. Linde, Phys. Lett. B **96B**, 289 (1980).
19. Yu. A. Simonov, hep-ph/9509404.
20. E. L. Gubankova and Yu. A. Simonov, Phys. Lett. B **360**, 93 (1995).
21. N. O. Agasyan, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 65 (2000) [JETP Lett. **71**, 43 (2000)].
22. N. O. Agasian, Phys. Lett. B **562**, 257 (2003).
23. Yu. A. Simonov, Pis'ma Zh. Éksp. Teor. Fiz. **54**, 256 (1991) [JETP Lett. **54**, 249 (1991)].
24. Yu. A. Simonov, Pis'ma Zh. Éksp. Teor. Fiz. **55**, 605 (1992) [JETP Lett. **55**, 627 (1992)].
25. Yu. A. Simonov, Yad. Fiz. **58**, 357 (1995) [Phys. At. Nucl. **58**, 309 (1995)].
26. F. E. Low, Phys. Rev. **110**, 974 (1958).
27. V. A. Novikov, M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **191**, 301 (1981); Fiz. Elem. Chastits At. Yadra **13**, 542 (1982) [Sov. J. Part. Nucl. **13**, 224 (1982)]; A. A. Migdal and M. A. Shifman, Phys. Lett. B **114B**, 445 (1982).
28. E. V. Shuryak, Nucl. Phys. B **203**, 93 (1982).
29. D. Diakonov and V. Y. Petrov, Nucl. Phys. B **245**, 259 (1984).
30. H. G. Dosch, Phys. Lett. B **190**, 177 (1987).
31. H. G. Dosch and Yu. A. Simonov, Phys. Lett. B **205**, 339 (1988).
32. Yu. A. Simonov, Nucl. Phys. B **307**, 512 (1988); **324**, 67 (1989).
33. A. Di Giacomo, H. G. Dosch, V. I. Shevchenko, and Yu. A. Simonov, Phys. Rep. **372**, 319 (2002).
34. M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **163**, 46 (1980).
35. A. B. Migdal, N. O. Agasyan, and S. B. Khokhlachev, Pis'ma Zh. Éksp. Teor. Fiz. **41**, 405 (1985) [JETP Lett. **41**, 497 (1985)]; N. O. Agasyan and S. B. Khokhlachev, Yad. Fiz. **55**, 1116, 1126 (1992) [Sov. J. Nucl. Phys. **55**, 628, 633 (1992)].
36. N. O. Agasian and Yu. A. Simonov, Mod. Phys. Lett. A **10**, 1755 (1995); N. O. Agasyan, Yad. Fiz. **59**, 317 (1996) [Phys. At. Nucl. **59**, 297 (1996)]; hep-ph/9803252; hep-ph/9904227.
37. N. O. Agasian and S. M. Fedorov, J. High Energy Phys. **0112**, 019 (2001); hep-ph/0111305; hep-ph/0211139; Yad. Fiz. **67**, 394 (2004) [Phys. At. Nucl. **67**, 376 (2004)].
38. G. S. Bali, Phys. Rev. D **62**, 114503 (2000).
39. V. I. Shevchenko and Yu. A. Simonov, Phys. Rev. Lett. **85**, 1811 (2000).
40. S. Weinberg, Physica A **96**, 327 (1979).
41. J. Gasser and H. Leutwyler, Ann. Phys. (N.Y.) **158**, 142 (1984).
42. J. Gasser and H. Leutwyler, Nucl. Phys. B **250**, 465 (1985).
43. P. Binétrui and M. K. Gaillard, Phys. Rev. D **32**, 931 (1985).
44. J. Gasser and H. Leutwyler, Phys. Lett. B **184**, 83 (1987).
45. J. Gasser and H. Leutwyler, Phys. Lett. B **188**, 477 (1987).
46. H. Neuberger, Phys. Rev. Lett. **60**, 889 (1988).
47. H. Leutwyler, Nucl. Phys. B (Proc. Suppl.) **4**, 248 (1988).
48. P. Gerber and H. Leutwyler, Nucl. Phys. B **321**, 387 (1989).
49. F. Karsch, E. Laermann, and A. Peikert, Nucl. Phys. B **605**, 579 (2001).
50. N. O. Agasian, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 200 (1993) [JETP Lett. **57**, 208 (1993)].
51. H. Leutwyler, *Restoration of Chiral Symmetry, Lecture Given at Workshop on Effective Field Theories, Dobogoko, Hungary, 1991*, Bern Univ.-BUTP-91-43.
52. N. O. Agasian, Phys. Lett. B **488**, 39 (2000).
53. N. O. Agasian, Yad. Fiz. **64**, 608 (2001) [Phys. At. Nucl. **64**, 554 (2001)].
54. N. O. Agasian and I. A. Shushpanov, J. High Energy Phys. **0110**, 006 (2001).
55. N. O. Agasian and I. A. Shushpanov, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 711 (1999) [JETP Lett. **70**, 717 (1999)].
56. N. O. Agasian and I. A. Shushpanov, Phys. Lett. B **472**, 143 (2000).
57. N. O. Agasian, Phys. Lett. B **519**, 71 (2001).
58. N. O. Agasian, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 387 (2001) [JETP Lett. **74**, 353 (2001)].
59. T. Muta, *Foundations of Chromodynamics, World Scientific Lecture Notes in Physics 57* (World Sci., Singapore, 1998).
60. P. J. Ellis, J. I. Kapusta, and H. B. Tang, Phys. Lett. B **443**, 63 (1998).

61. M. Gell-Mann, R. Oakes, and B. Renner, Phys. Rev. **175**, 2195 (1968).
62. E. V. Shuryak, Phys. Lett. B **207**, 345 (1988).
63. H. Leutwyler and A. Smilga, Nucl. Phys. B **342**, 302 (1990).
64. J. L. Goity and H. Leutwyler, Phys. Lett. B **228**, 517 (1989).
65. H. Bebie, P. Gerber, J. L. Goity, and H. Leutwyler, Nucl. Phys. B **378**, 95 (1992).
66. M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **147**, 385 (1979).
67. N. O. Agasian, D. Ebert, and E.-M. Ilgenfritz, Nucl. Phys. A **637**, 135 (1998).
68. N. O. Agasian, Yad. Fiz. **67**, 409 (2004) [Phys. At. Nucl. **67**, 391 (2004)].
69. Y. Nakamura, V. Bornyakov, and M. N. Chernodub, hep-lat/0309144.
70. M. D'Elia, A. Di Giacomo, and E. Meggiolaro, Phys. Rev. D **67**, 114504 (2003).
71. G. S. Bali, J. Fingberg, U. M. Heller, *et al.*, Phys. Rev. Lett. **71**, 3059 (1993).
72. F. Karsch, E. Laermann, and M. Lutgemeier, Phys. Lett. B **346**, 94 (1995).
73. G. Boyd, J. Engels, F. Karsch, *et al.*, Nucl. Phys. B **469**, 419 (1996).
74. N. O. Agasian and S. M. Fedorov, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 1099 (2003) [JETP Lett. **78**, 607 (2003)].
75. R. Hagedorn, Nuovo Cimento Suppl. **3**, 147 (1965).
76. P. Braun-Munzinger, D. Magestro, K. Redlich, and J. Stachel, Phys. Lett. B **518**, 41 (2001).
77. J. Cleymans and K. Redlich, Phys. Rev. Lett. **81**, 5284 (1998).
78. F. Becattini, J. Cleymans, A. Keranen, *et al.*, Phys. Rev. C **64**, 024901 (2001).
79. F. Karsch, K. Redlich, and A. Tawfik, Eur. Phys. J. C **29**, 549 (2003).
80. N. O. Agasian, B. O. Kerbikov, and V. I. Shevchenko, Phys. Rep. **320**, 131 (1999).
81. F. Karsch, E. Laermann, and A. Peikert, Phys. Lett. B **478**, 447 (2000).
82. M. Campostrini, A. Di Giacomo, and G. Mussardo, Z. Phys. C **25**, 173 (1984).
83. J. Engels, F. Karsch, and K. Redlich, Nucl. Phys. B **435**, 295 (1995).
84. C. Borgs, Nucl. Phys. B **261**, 455 (1985); E. Manousakis and J. Polonyi, Phys. Rev. Lett. **58**, 847 (1987).
85. D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. **53**, 43 (1981).
86. P. Ginsparg, Nucl. Phys. B **170**, 388 (1980).
87. T. Appelquist and R. D. Pisarski, Phys. Rev. D **23**, 2305 (1981).
88. S. Nadkarni, Phys. Rev. D **27**, 917 (1983).
89. N. P. Landsman, Nucl. Phys. B **322**, 498 (1989).
90. K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, Nucl. Phys. B **458**, 90 (1996).
91. K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, Nucl. Phys. B **503**, 357 (1997).
92. M. Laine and O. Philipsen, Phys. Lett. B **459**, 259 (1999).
93. A. Hart and O. Philipsen, Nucl. Phys. B **572**, 243 (2000).
94. F. Karsch, M. Oevers, and P. Petreczky, Phys. Lett. B **442**, 291 (1998); A. Cucchieri, F. Karsch, and P. Petreczky, Phys. Lett. B **497**, 80 (2001).
95. M. Teper, Phys. Lett. B **311**, 223 (1993).
96. A. Di Giacomo and H. Panagopoulos, Phys. Lett. B **285**, 133 (1992).
97. A. Di Giacomo, E. Meggiolaro, and H. Panagopoulos, Nucl. Phys. B **483**, 371 (1997).
98. N. A. Campbell, I. H. Jorysz, and C. Michael, Phys. Lett. B **167B**, 91 (1986); I. H. Jorysz and C. Michael, Nucl. Phys. B **302**, 448 (1988).
99. H. G. Dosch, M. Eidemuller, and M. Jamin, Phys. Lett. B **452**, 379 (1999); M. Eidemuller, H. G. Dosch, and M. Jamin, Nucl. Phys. B (Proc. Suppl.) **86**, 421 (2000).
100. Yu. A. Simonov, Nucl. Phys. B **592**, 350 (2001).
101. O. Philipsen, Nucl. Phys. B **628**, 167 (2002); M. Laine and O. Philipsen, Nucl. Phys. B **523**, 267 (1998).
102. P. Petreczky, hep-ph/9907247.
103. N. O. Agasian and S. M. Fedorov, Pis'ma Zh. Éksp. Teor. Fiz. **80**, 86 (2004) [JETP Lett. **80**, 78 (2004)].
104. N. O. Agasian and S. M. Fedorov, J. High Energy Phys. **0407**, 007 (2004).
105. E.-M. Ilgenfritz, B. V. Martemyanov, S. V. Molodtsov, *et al.*, Phys. Rev. D **58**, 114508 (1998).
106. B. J. Harrington and H. K. Shepard, Phys. Rev. D **17**, 2122 (1978).
107. N. O. Agasian and K. Zarembo, Phys. Rev. D **57**, 2475 (1998).
108. N. O. Agasian and D. Antonov, J. High Energy Phys. **0106**, 058 (2001).
109. N. O. Agasian and D. Antonov, Phys. Lett. B **530**, 153 (2002).
110. T. Schafer and E. V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998).
111. T. C. Kraan and P. van Baal, Phys. Lett. B **435**, 389 (1998); K. M. Lee and C. h. Lu, Phys. Rev. D **58**, 025011 (1998).
112. E. M. Ilgenfritz, B. V. Martemyanov, A. I. Veselov, *et al.*, Phys. Rev. D **66**, 074503 (2002).
113. E. M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, *et al.*, Nucl. Phys. B (Proc. Suppl.) **119**, 754 (2003).
114. C. Gattringer, E. M. Ilgenfritz, B. V. Martemyanov, *et al.*, Nucl. Phys. B (Proc. Suppl.) **129–130**, 653 (2004).
115. B. Lucini, M. Teper, and U. Wenger, hep-lat/0401028.


*Translated by A. Isaakyan*

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 748–757. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 780–789.*


# Pentaquarks: Facts and Puzzles*

**I. M. Narodetskii**

*Institute of Theoretical and Experimental Physics,*
*Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received May 31, 2004; in final form, September 29, 2004

**Abstract**—On the occasion of the celebration of the 70th birthday of Prof. Yu.A. Simonov, we contribute a brief review of the status of exotic baryons, which are most likely pentaquark states. We summarize the experimental status of exotic baryons, discuss the baryon antidecuplet to which exotic baryons possibly belong, and review theoretical expectations for the masses and widths of recently discovered $\Theta$ and $\Xi_{3/2}$ baryons which have come from studies of QCD sum rules, lattice QCD, and quark models. We also pay special attention to the dynamical calculation of pentaquark masses in a framework of the QCD string approach originally elaborated by Simonov for baryons and using the Jaffe–Wilczek $[ud]^2\bar{q}$ approximation for the pentaquark states. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Recently, the LEPS [1] and DIANA [2] Collaborations reported the observation of a very narrow peak in the $K^+n$ and $K^0p$ invariant mass distribution, whose existence has been confirmed by several experimental groups in various reaction channels [3–11]. These experimental results were motivated by a pioneering paper on the chiral soliton model ($\chi$SM) [12]. The reported mass determinations for $\Theta$ are very consistent, falling in the range $1540 \pm 10$ MeV, with the width smaller than the experimental resolution of 20 MeV for the photon- and neutrino-induced reactions and of 9 MeV for the ITEP $K^+\mathrm{Xe} \to K^0p\mathrm{Xe}'$ experiment. A recent analysis of $K^+d$ total cross section data [13] found a structure corresponding to the resonance at the mass $1.559 \pm 0.003$ GeV with the width $0.9 \pm 0.3$ MeV.

From the soliton point of view, $\Theta$ is nothing exotic compared with other baryons—it appears naturally in the rigid rotator approach to the three-flavor Skyrme model.[1)] However, in the sense of the quark model, the $\Theta^+(1540)$ baryon with positive amount of strangeness is manifestly exotic—its minimal configuration cannot be satisfied by three quarks. The positive strangeness requires an $\bar{s}$ and $qqqq$ (where $q$ refers to the lightest quarks $u, d$) are required for the net baryon number, thus making a pentaquark $uudd\bar{s}$ state as the minimal "valence" configuration. The experimental evidence for the manifestly exotic baryon states provides an opportunity to refine our quantitative understanding of nonperturbative QCD at low energy.

In October 2003, the NA49 Collaboration at CERN SPS [14] announced evidence for an additional narrow $\Xi^-\pi^-$ resonance with $I = 3/2$, a mass of $1862 \pm 2$ MeV, and a width below the detector resolution of about 18 MeV. NA49 also reported evidence for a $\Xi^0(1860)$ decaying into $\Xi(1320)\pi$. Recently, a narrow resonance $\Theta_c(3099)$ in $D^{*-}p$ and $D^{*+}\bar{p}$ invariant mass combinations was observed in inelastic electron–proton collisions at center-of-mass energies of 300 and 320 GeV by the H1 Collaboration at HERA [15] with a mass of $3099 \pm 6$ MeV and decay width of $12 \pm 3$ MeV. It can be interpreted as a pentaquark baryon with minimal quark content $(udud\bar{c})$, thus being the first exotic baryon with anticharm quark, implying the existence of other exotic baryons with heavy quarks. A partial summary of experimental results is given in Table 1. The complete list of both positive and negative experimental results existing as of May 2004 is given, e.g., in [16]. While a dozen experiments have reported evidence for the phenomenon, another dozen have not seen the states. In particular, the $e^+e^-$-collider data (Belle, BaBar, ALEPH, DELPHI) give null results. The Fermilab high-energy proton experiments do not observe $\Xi(1860)$, $\Theta^+$, or $\Theta_c$ (CDF); do not confirm $\Theta_c$ in either $D^{*-}p$ or $D^-p$ channels (FOCUS); and do not observe $\Theta \to pK^0_S$ (HyperCP, E690) [17].

The results from $KN$ scattering known prior 2003 and referring to the phenomenology of $Z^*$ (the old name of $KN$ resonances with $S = 1$) can be summarized as follows. The 1982 version of the

*This article was submitted by the author in English.

[1)] In this framework, the lowest lying exotic baryon is an $S = +1$ member of the antidecuplet $\overline{\mathbf{10}}_F$, with other exotic representations, such as the $\mathbf{27}_F$, $\mathbf{35}_F$, and $\overline{\mathbf{35}}_F$, being heavier.

**Table 1.** Summary of experimental data on $\Theta$ and $\Xi_{3/2}$ hyperons

| Collaboration | Reaction | Mass, MeV | $\Gamma$, MeV | Significance, $\sigma$ |
|---|---|---|---|---|
| LEPS [1] | $\gamma n \to K^+K^-n\,(^{12}\mathrm{C})$ | $1540 \pm 10$ | $\leq 25$ | 4.6 |
| DIANA [2] | $K^+\mathrm{Xe} \to K^0 p\mathrm{Xe}'$ | $1539 \pm 2$ | $\leq 9$ | 4.5 |
| CLAS [3] | $\gamma d \to K^+K^-pn$ | $1642 \pm 5$ | $\leq 21$ | 5.3 |
| SAPHIR [5] | $\gamma p \to K^+K_S^0 n$ | $1640 \pm 4$ | $\leq 20$ | 4.8 |
| FNAL [6] | Bubble chambers | $1533 \pm 5$ | $\leq 20$ | 6.7 |
| HERMES [7] | $\gamma d, \Theta^+ \to pK_S \to p\pi\pi$ | $1528 \pm 3$ | $\geq 4$–$6$ | $\sim 6$ |
| NA49 [14] | $pp \to \Xi^-\pi^- X$ | $1862 \pm 2$ | $\leq 18$ | 4.6 |
| H1 [15] | $e^+e^-$ | $3099 \pm 6$ | $12 \pm 3$ | |

"Review of Particle Properties" lists several $Z_0$ and $Z_1$ structures in the energy region 1800–2100 MeV: $Z_0(1780)(P01)$, $Z_0(1865)(D03)$, $Z_1(1900)(P13)$,[2] $Z_1(2150)$, and $Z_1(2500)$[3] with widths of 100 MeV or more. At least some of these have been understood as "pseudoresonance" structures [18] arising from the coupling to inelastic $K^*N$ and $K\Delta$ channels. An earlier search [19] for a $Z^*$ in the reaction $\pi^-p \to K^-Z^*$ at $\pi^-$ momenta of 6 and 8 GeV/$c$ found enhancements in the missing mass of recoiling $K^-$ at 1.5 and 1.9 GeV/$c^2$ being explained as kinematical reflection of the $f_2(1275), a_2(1320)$, and $\rho_3(1690)$ resonances. The absence of fresh news has left this interpretation in limbo. A similar possibility [20] for the enhancement observed in the $K^+n$ effective mass distribution at the mass of $\Theta$ seems to be unlikely, but should be carefully checked in the forthcoming experiments.

Existence of a narrow $\Theta^+$ state around 1540 MeV should also be seen in available $K^+d$ data. The lack of a prominent $\Theta^+$ signature in the total $K^+d$ cross section seems to exclude the width of $\Theta$ beyond the few-MeV level [21]. Similar but even more restrictive conclusions were drawn from the $K^+N$ phase-shift analysis [22].

In this paper, we explore the phenomenology of, and models for, the exotic baryons, the lowest of which is $\Theta^+(1540)$. The next section includes a pedagogical discussion of the baryon antidecuplet to which exotic baryons possibly belong. Then, in Section 3, we review the theoretical expectations for the masses and widths of exotic baryons. Finally, we provide an overview of the effective Hamiltonian approach to QCD pioneered by Yu.A. Simonov and pay special attention to the dynamical calculation of pentaquark masses in the framework of this approach.

## 2. BARYON ANTIDECUPLET

The possibility of low-lying $q^4\bar{s}$ states in the $P$ wave fitting in the antidecuplet flavor representation $\overline{\mathbf{10}}_F$ of the quark model was mentioned long ago by Golovich [23]. Later on, it was discovered that the $\overline{\mathbf{10}}_F$ multiplet appears as the third (after the octet with spin 1/2 and decuplet with spin 3/2) rotational state in the $SU(3)$ version of the $\chi$SM [24]. To get a feeling for the antidecuplet (to which the newly discovered $\Theta^+$ and $\Xi^-_{3/2}$ presumably belong), we first recall the Gell-Mann–Okubo mass formula for a $SU(3)_F$ multiplet

$$M = M_0 + \alpha Y + \beta D_3^3, \tag{1}$$

where $M_0$ is the center of gravity for a given multiplet,

$$D_3^3 = I(I+1) - Y^2/4 - C/6, \tag{2}$$

and

$$C = 2(p+q) + \frac{2}{3}(p^2 + pq + q^2) \tag{3}$$

is expressed in terms of the labels $(p, q)$ which denote the $SU(3)_F$ representation. Schematically, $q$ is the number of boxes in the lower line of the Young tableau depicting an $SU(3)_F$ representation and $p$ is the number of extra boxes in the upper line. For example, $(p, q) = (1, 1), (3, 0)$, and $(0, 3)$ for $J^P = 1/2^+$ octet, $J^P = 3/2^+$ decuplet, and $J^P = 1/2^+$ antidecuplet, respectively. In Eq. (1), $\alpha$ and $\beta$ are constants

[2] We use the standard notation $L_{I2J}$.

[3] The last two have no spin or parity assignment.

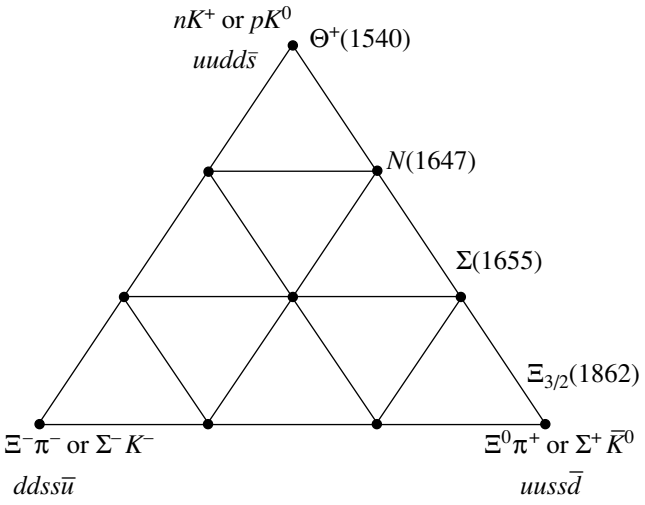


**Fig. 1.** The $(I_3, Y)$ diagram for the baryonic antidecuplet. The corners of diagram are manifestly exotic. The $\Theta$ and $\Xi_{3/2}$ masses are used as input; the $N_{\overline{\mathbf{10}}}$ and $\Sigma_{\overline{\mathbf{10}}}$ masses are predicted using Eq. (1).

generally different for different multiplets.[4)] Splitting of otherwise degenerate baryon masses inside each $SU(3)_F$ multiplet is due to nonzero strange quark mass.

It follows from Eq. (1) that, for the antidecuplet, there are not three but only two independent parameters: $M_{\overline{\mathbf{10}}}$ and $\delta m_{\overline{\mathbf{10}}} = \alpha_{\overline{\mathbf{10}}} - \frac{3}{2}\beta_{\overline{\mathbf{10}}}$.[5)] Therefore, if $\mathbf{10}_F$ does not mix with other multiplets, one gets the equidistant masses of antidecuplet baryons:

$$M_\Theta = M_{\overline{\mathbf{10}}} + 2\delta m_{\overline{\mathbf{10}}}, \quad (4)$$
$$M_{N_{\mathbf{10}}} = M_{\overline{\mathbf{10}}} + \delta m_{\overline{\mathbf{10}}},$$
$$M_{\Sigma_{\mathbf{10}}} = M_{\overline{\mathbf{10}}}, M_{\Xi_{\mathbf{10}}} = M_{\overline{\mathbf{10}}} - \delta m_{\overline{\mathbf{10}}},$$

where $\delta m_{\overline{\mathbf{10}}} = \alpha_{\overline{\mathbf{10}}} - \frac{3}{2}\beta_{\overline{\mathbf{10}}}$. This suggests that, if any two of the masses in the antidecuplet are known, the other masses can be predicted by relations (4). For example, using $M_\Theta = 1540$ MeV and identifying the recent measurement of $M_{\Xi_{3/2}} = 1862$ MeV by the NA49 experiment as $M_{\Xi_{\mathbf{10}}}$, one obtains

$$M_{N_{\mathbf{10}}} = 1647 \text{ MeV}, \quad M_{\Sigma_{\mathbf{10}}} = 1755 \text{ MeV}. \quad (5)$$

[4)] $\alpha$ and $\beta$ are related to the constants $x, y, z'$, and $z$ in Eqs. (1), (2) of [25] by

$$\alpha_8 = -y - \frac{5}{4}x, \quad \beta_8 = x,$$
$$\alpha_{\mathbf{10}} - \frac{3}{2}\beta_{\mathbf{10}} = z', \quad \alpha_{\overline{\mathbf{10}}} - \frac{3}{2}\beta_{\overline{\mathbf{10}}} = z.$$

[5)] This is also true for the decuplet with $\delta m_{\mathbf{10}} = \alpha_{\mathbf{10}} + (3/2)\beta_{\mathbf{10}}$. However, the mass parameters are quite different: $\delta m_{\overline{\mathbf{10}}} \sim -107$ MeV for the antidecuplet and $\delta m_{\mathbf{10}} \sim -150$ MeV for the decuplet.

The $(Y, I_3)$ diagram for the antidecuplet of baryons is shown in Fig. 1. Also shown is the quark content of each state together with the predicted masses for $\mathrm{N}_{\overline{\mathbf{10}}}$ and $\Sigma_{3/2}$. The corners of this diagram $[ud]^2\bar{s}$, $[ds]^2\bar{u}$, and $[us]^2\bar{d}$ are manifestly exotic.

The nonzero strange quark mass also leads to the mixing of octet and antidecuplet states. The exotic members of the antidecuplet $\Theta^+$ and $\Xi_{3/2}$ cannot mix with octet baryons, though in principle they can mix with higher exotic multiplets like $\overline{\mathbf{27}}_F$ with $J = 1/2$ and $3/2$. Two other particles common to the octet and the antidecuplet are $N$ and $\Sigma$, and only they can mix,

$$|N_{\overline{\mathbf{10}}}\rangle = |N, \overline{\mathbf{10}}\rangle - c_{\overline{\mathbf{10}}}|N, \mathbf{8}\rangle, \quad (6)$$
$$|\Sigma_{\overline{\mathbf{10}}}\rangle = |\Sigma, \overline{\mathbf{10}}\rangle - c_{\overline{\mathbf{10}}}|\Sigma, \mathbf{8}\rangle. \quad (7)$$

Estimation in the $\chi$SM yields for $c_{\overline{\mathbf{10}}}$ a not negligible quantity: $c_{\overline{\mathbf{10}}} \sim 0.1$ [12]. The general consideration of octet–antidecuplet mixing has been given in [25] with the conclusion that a good octet candidate for mixing with a $\overline{\mathbf{10}}_F$ may consist of $N(1410)$, $\Lambda(1600)$, $\Sigma(1660)$, and $\Xi(1690)$, although in any case the mixing cannot be too large.

## 3. THE PENTAQUARK WIDTH

The most striking experimental result is a very small width of $\Theta^+$. In the $\chi$SM, the key element in the narrow $\Gamma_{\Theta^+}$ prediction is the strong cancellation between the coupling constants $G_0$ and $G_1$ which enter the decay operator:

$$\Gamma_\Delta \sim \left(G_0 + G_1 - \frac{1}{2}G_2\right)^2, \quad (8)$$
$$\Gamma_\Theta \sim \left(G_0 - \frac{1}{2}G_1 - \frac{1}{2}G_2\right)^2.$$

This cancellation seems, at first sight, unnatural because the constants $G_{1,2}$ are nonleading as far as $N_c$ counting is concerned. At leading order in the $1/N_c$ expansion, all the couplings of the $\mathbf{8}_F, \mathbf{10}_F$, and $\overline{\mathbf{10}}_F$ baryons to pseudoscalar mesons are proportional to the dimensionless constant $G_0$, in which case the width of $\Theta^+$ would be of the same order as $\Gamma_\Delta$. However, in antidecuplet decays, the $G_1$ contribution gets an additional $N_c$ enhancement from the $SU(3)$-flavor Clebsch–Gordan coefficients calculated in large $N_c$. A dedicated discussion of this problem can be found in [26]. The nonrelativistic limit implying $G_1/G_0 = 4/5, G_2/G_0 = 2/5$ would predict a strong suppression of $\Gamma_{\Theta^+}$ and $\Gamma_{\Xi^{--}}$. Reliable predictions for the decay widths depend on modeling and

educated guesses, and hence are subject to additional uncertainties. According to [27],

$$\Gamma_{\Theta^+} = \frac{3G^2_{\overline{\mathbf{10}}}}{8\pi M_{\Theta^+} M_N} \frac{1}{5} p_K^3 = 20.6 \text{ MeV}. \qquad (9)$$

Although this number is relatively small, it is considerably larger than recent experimental estimates given in Table 1.

Note that, if the $\Theta$ and $\Xi_{3/2}$ have the same spin–parity, then their widths can be related by assuming $SU(3)_F$ symmetry for the matrix elements and correcting for phase space. This should be reliable at the level of typical $SU(3)_F$ symmetry violation, i.e., $\sim$30%. This estimate yields $\Xi_{3/2}$ width of order 3.5 times the width of the $\Theta$.

From the hadronic point of view, the $\Theta^+(1540)$ lies about 100 MeV above the $K^+n$ threshold at the center-of-mass momentum 270 MeV/$c$. There are no other two-body hadronic channels coupled to $K^+n$ below the $K\Delta$ threshold at 1725 MeV. Since there are no $q\bar{q}$ annihilation graphs in $K^+n$ scattering, one expects that low-energy $K^+n$ scattering is dominated by the ordinary exchange forces (see [28] and references therein). The $S$-wave phase shifts are known to be repulsive at low energies, which implies that there are no $1/2^-$ resonances. In the $P$-wave sector, the $P01$ and $P13$ waves are attractive, while the $P03$ and $P11$ waves are repulsive. In the $P03$ and $P11$ states, attractive forces may generate a resonance through the interplay between the attraction and repulsive angular-momentum barrier. A simple estimation yields the width of a $P$-wave resonance 100 MeV above threshold of order 100–200 MeV depending on the range of attractive potential. Therefore, a narrow $P$-wave state cannot be explained without suppression beyond $KN$ physics. Additional smallness may arise from recoupling color and flavor–spin to overlap the $uudd\bar{s}$ onto color singlets $uud$ and $d\bar{s}$. According to [29], the color recoupling costs $1/\sqrt{3}$ and color–spin recoupling costs a further 1/4. Given typical widths for baryons in the $\Theta^+$ mass range of 200 MeV, we would estimate naively that the $\Theta^+$ width is about a few tens of MeV. A further reduction may be due to differences in spatial overlaps of the coordinate wave functions [30]. Hence, a width of $O$ (a few MeV) does not have to be unrealistic.

## 4. THEORETICAL PREDICTIONS FOR THE $\Theta^+$ MASS

As to the theoretical predictions, we are faced with a somewhat ambiguous situation, in which exotic baryons *may* have been discovered, but there are important controversies with theoretical predictions for masses of pentaquark states. The experimental results triggered vigorous theoretical activity and put a renewed urge in the need to understand how baryon properties are obtained from QCD.

The $\Theta$ hyperon has hypercharge $Y = 2$ and third component of isospin $I_3 = 0$. The apparent absence of the $I_3 = +1$, $\Theta^{++}$ in $K^+p$ argues against $I = 1$;[6] therefore, the $\Theta$ is usually assumed to be an isosinglet, although other suggestions have been made in the literature [31]. The other quantum numbers are not established yet.

All attempts at theoretical estimations of the pentaquark masses can be subdivided into the following four categories: (i) dynamical calculations using the $SU(3)$ $\chi$SM, (ii) dynamical calculations using the sum rules or lattice QCD, (iii) phenomenological analyses of the $SU(3)_F$ mass relations, and (iv) phenomenological analysis of the hyperfine splitting in the Sakharov–Zeldovich quark model.

The success of the prediction of [12] for the $\Theta^+$ mass was somewhat fortuitous. Reference [12] identified the $N_{\overline{\mathbf{10}}}$ with $N(1710)$[7] and assumed a value of the $\pi$–nucleon $\sigma$ term $\Sigma = 45$ MeV. This latter value was responsible for unsuccessful prediction of a very heavy mass $\sim$2070 MeV for the $\Xi^{--}$. More recently Ellis *et al.* [27] critically discussed the calculation of the $\Theta^+$ and $\Xi^{--}$ masses and widths in the $\chi$SM and explored the unbiased theoretical and phenomenological uncertainties. Overall, they found the ranges

$$1432 < M_{\Theta^+} < 1657 \text{ MeV}, \qquad (10)$$
$$1786 < M_{\Xi^{--}} < 1970 \text{ MeV}.$$

These ranges certainly include the observed masses $M_{\Theta^+} = 1539 \pm 2$ MeV and $M_{\Xi^{--}} = 1862 \pm 2$ MeV, but more precise predictions cannot be made without introducing more assumptions.

The QCD sum rules predict a negative parity $\Theta^+$ of mass $\simeq$1.5 GeV, while no positive-parity state was found [32]. The lattice QCD calculations have claimed a pentaquark signal of either negative parity [33, 34] or positive parity [35] in the vicinity

[6] On general grounds, one may suspect that, if the isosinglet pentaquark has been found, the $uduu\bar{s}$ and $uuuu\bar{s}$ may be discovered as well. In $\chi$SM, the existence of $I = 1$ and $I = 2$ exotics is tied to the existence of the $I = 0$ one. In particular, in the $SU(3)$ rigid-rotator model, they belong to $\mathbf{27}_F$ and $\mathbf{35}_F$, respectively. In the Jaffe–Wilczek (JW) diquark model with scalar–isoscalar diquarks (see below), such states are absent, but appear if one assumes the presence of spin-1 diquark symmetric in flavor. The $I = 1$ resonances are typically predicted in the range of 1600–1680 MeV and should be observable through the decay into $pK^+$. There is no evidence for such a resonance in the available $pK^+$ data and recent photoproduction experiments. SAPHIR and CLAS Collaborations rule out $I = 2$; HERMES makes $I = 1$ also highly unlikely.

[7] This identification was abandoned in [25].

**Table 2.** Flavor wave functions in the Jaffe–Wilczek model

| $(Y, I, I_3)$ | $\overline{\mathbf{10}}_F$ | $(Y, I, I_3)$ | $\mathbf{8}_F$ |
|---|---|---|---|
| $(2, 0, 0)$ | $[ud]^2\bar{s}$ | – | – |
| $\left(1, \frac{1}{2}, \frac{1}{2}\right)$ | $\sqrt{\frac{2}{3}}[ud][us]_+\bar{s} + \sqrt{\frac{1}{3}}[ud]^2\bar{d}$ | $\left(1, \frac{1}{2}, \frac{1}{2}\right)$ | $\sqrt{\frac{1}{3}}[ud][us]_+\bar{s} - \sqrt{\frac{2}{3}}[ud]^2\bar{d}$ |
| $\left(1, \frac{1}{2}, -\frac{1}{2}\right)$ | $\sqrt{\frac{2}{3}}[ud][ds]_+\bar{s} + \sqrt{\frac{1}{3}}[ud]^2\bar{u}$ | $\left(1, \frac{1}{2}, -\frac{1}{2}\right)$ | $\sqrt{\frac{1}{3}}[ud][ds]_+\bar{s} - \sqrt{\frac{2}{3}}[ud]^2\bar{u}$ |
| $(0, 1, 1)$ | $\sqrt{\frac{2}{3}}[ud][us]_+\bar{d} + \sqrt{\frac{1}{3}}[us]^2\bar{s}$ | $(0, 1, 1)$ | $\sqrt{\frac{1}{3}}[ud][us]_+\bar{d} - \sqrt{\frac{2}{3}}[us]^2\bar{s}$ |
| $(0, 1, 0)$ | $\sqrt{\frac{1}{3}}([ud][ds]_+\bar{d} + [ud][us]_+\bar{u} + [us][ds]_+\bar{s})$ | $(0, 1, 0)$ | $\sqrt{\frac{1}{6}}([ud][ds]_+\bar{d} + [ud][us]_+\bar{u}) - \sqrt{\frac{2}{3}}[us][ds]_+\bar{s}$ |
| $(0, 1, -1)$ | $\sqrt{\frac{2}{3}}[ud][ds]_+\bar{u} + \sqrt{\frac{1}{3}}[ds]^2\bar{s}$ | $(0, 1, -1)$ | $\sqrt{\frac{1}{3}}[ud][ds]_+\bar{u} - \sqrt{\frac{2}{3}}[ds]^2\bar{s}$ |
| $\left(-1, \frac{3}{2}, \frac{3}{2}\right)$ | $[us]^2\bar{d}$ | – | – |
| $\left(-1, \frac{3}{2}, \frac{1}{2}\right)$ | $\sqrt{\frac{2}{3}}[us][ds]_+\bar{d} + \sqrt{\frac{1}{3}}[us]^2\bar{u}$ | $\left(-1, \frac{1}{2}, \frac{1}{2}\right)$ | $\sqrt{\frac{1}{3}}[us][ds]_+\bar{d} - \sqrt{\frac{2}{3}}[us]^2\bar{u}$ |
| $\left(-1, \frac{3}{2}, -\frac{1}{2}\right)$ | $\sqrt{\frac{2}{3}}[ds][us]_+\bar{u} + \sqrt{\frac{1}{3}}[ds]^2\bar{d}$ | $\left(-1, \frac{1}{2}, -\frac{1}{2}\right)$ | $\sqrt{\frac{1}{3}}[ds][us]_+\bar{u} - \sqrt{\frac{2}{3}}[ds]^2\bar{d}$ |
| $\left(-1, \frac{3}{2}, -\frac{3}{2}\right)$ | $[ds]^2\bar{u}$ | – | – |
| – | – | $(0, 0, 0)$ | $\sqrt{\frac{1}{2}}([ud][ds]_+\bar{d} - [ud][us]_+\bar{u})$ |

of 1540 MeV. However, the most recent quenched calculation [36] seems to reveal no evidence for a pentaquark state with quantum numbers $I(J^P) = 0(1/2^{\pm})$ near this mass region.

The constituent quark models (CQM), in which all constituents are in a relative $S$ wave, naturally predict the ground-state energy of a $J^P = 1/2^-$ pentaquark to be lower than that of a $J^P = 1/2^+$ one. Using the arguments based on both Goldstone boson exchange between constituent quarks and color-magnetic exchange, it was mentioned that the increase in hyperfine energy in going from negative- to positive-parity states can be quite enough to compensate the orbital excitation energy ~200 MeV [37]. However, existing dynamical calculations of pentaquark masses in CQM (see, e.g., [38, 39]) are subject to significant uncertainties and cannot be considered as conclusive. Note that, if $\Theta$ is indeed $ududs\bar{s}$ with $J^P = 1/2^+$, then the spin–orbit QCD forces necessarily imply that there has to be the $J^P = 3/2^+$ partner of $\Theta^*$, which is probably only a few MeV heavier [40].

## 5. JAFFE–WILCZEK MODEL

Pentaquark baryons are unexpectedly light. Indeed, a naive quark model with quark mass ~350 MeV predicts $\Theta^+$ at about $350 \times 5 = 1750$ MeV, plus ~150 MeV for strangeness, plus ~200 MeV for the $P$-wave excitation. A natural remedy would be to decrease the number of constituents. This leads one to consider dynamical clustering into subsystems of diquarks like $[ud]^2\bar{s}$ and/or triquarks like $[ud][ud\bar{s}]$ which amplify the attractive color-magnetic forces. There are two routes that emerge naturally. One is that of [41], based on the Sakharov–Zeldovich CQM [42], in which the hadron mass is just the sum of the effective masses $m_i$ of its constituents plus hyperfine interaction

$$M = \sum_i m_i + \sum_{i>j} \frac{\sigma_i \sigma_j}{m_i m_j} v_{ij}^{\mathrm{hyp}}. \qquad (11)$$

The other is the JW model of [43] (see also [44]), where it has been proposed that the systematics of exotic baryons can be explained by diquark correlations.

In the last scenario in the $\Theta^+(1540)$ and other $q^4\bar{q}$ baryons, the four quarks are bound into two scalar, singlet isospin diquarks. Diquarks must couple to $\mathbf{3}_c$ to join in a color singlet hadron. In total, there are six flavor symmetric diquark pairs $[ud]^2$, $[ud][ds]_+$, $[ds]^2$, $[ds][su]_+$, $[su]^2$, and $[su][ud]_+$ combining with the remaining antiquark, which give 18 pentaquark states in $\mathbf{8}_F$ plus $\overline{\mathbf{10}}_F$. This is a general result: in the quark model one cannot get a $\overline{\mathbf{10}}_F$ without an $\mathbf{8}_F$. All these states are degenerate in the $SU(3)_F$ limit. The flavor $\overline{\mathbf{10}}_F$ and $\mathbf{8}_F$ wave functions of pentaquarks in the JW model are presented in Table 2.

The diquarks are postulated as building blocks for multiquark states on equal footing with constituent quarks. Note that the notion of diquarks is almost as old as that of quarks. In fact, diquarks were suggested by Gell-Mann in his pioneering paper about quarks [45]. Later, it was recognized that the strong chromomagnetic attraction and the attraction due to pion exchange in the $ud$ color triplet state with net spin 0 results in forces which may keep $u$ and $d$ quarks close together, thus forming an almost pointlike diquark. Such an idea has a long history and many phenomenological applications (for a review, see, e.g., [46]). Diquark correlations have played an important supporting role in QCD. A similar idea is a basis of color superconductivity in dense quark matter [47].

A crucial point of the JW approach is that, if both diquarks are identical scalars, Bose statistics would demand total symmetry under their exchange, while the color wave function is antisymmetric. Since diquarks have no spin, the only possibility is to make the spatial wave function antisymmetric by putting one diquark into a $P$-wave state, thus making the total parity positive.

## 6. EFFECTIVE HAMILTONIAN APPROACH TO QCD

Neither the Sakharov–Zeldovich model nor the quark constituent model has yet been derived from QCD. Therefore, it is tempting to consider the effective Hamiltonian (EH) approach in QCD, which, on one hand, can be derived from QCD and, on the other hand, leads to results for the $\bar{q}q$ mesons and $3q$ baryons which are equivalent to the quark model ones with some important modifications. This approach has been suggested by Simonov (see, e.g., the review paper [48] and references therein).

The EH approach contains the minimal number of input parameters, current (or pole) quark masses, the string tension $\sigma$, and the strong coupling constant $\alpha_s$, and does not contain fitting parameters such as the total subtraction constant in the Hamiltonian. It should be useful and attractive to consider expanding this approach to include diquark degrees of freedom with appropriate interactions. This program, based on assumption that chiral forces are responsible for the formation of $ud$ diquarks in $\Theta$, while the strings are mainly responsible for binding constituents in $\Theta$, was partially done in [49]. In this and subsequent sections, we briefly review application of the EH approach to the JW model of pentaquarks.

The EH for the three constituents has the form

$$H = \sum_{i=1}^{3}\left(\frac{m_i^2}{2\mu_i} + \frac{\mu_i}{2}\right) + H_0 + V, \qquad (12)$$

**Table 3.** Comparison of the EH approach predictions for spin-averaged masses (in GeV) of nucleons, hyperons, and pentaquark with lattice results and experimental data (assuming $\Theta^+$ is an $I^P = 0^+$ state)

| | $(N, \Delta)$ | $(N^-, \Delta^-)$ | $(\Lambda, \Sigma, \Sigma^*)$ | $\Theta^+$ $(I^P = 0^+)$ |
|---|---|---|---|---|
| Lattice | 1.07 [53] | 1.76 [54] | 1.21 [53] | 2.80 [33] |
| EH | 1.14 [56] | 1.63 [55] | 1.24 [56] | >2.07 [49] |
| Experiment | 1.08 | 1.62 | 1.27 | 1.54 |

where $H_0$ is the kinetic energy operator and $V$ is the sum of the perturbative one-gluon-exchange potentials and the string potential, which is proportional to the total length of the strings, i.e., to the sum of the distances of antiquark or diquarks from the string-junction point. The dynamical masses $\mu_i$ (analogs of the constituent ones) are expressed in terms of the current quark masses $m_i$ from the condition of the minimum of the hadron mass $M_H^{(0)}$ as a function of $\mu_i$[8]:

$$\frac{\partial M_H^{(0)}(m_i, \mu_i)}{\partial \mu_i} = 0, \qquad (13)$$

$$M_H^{(0)} = \sum_{i=1}^{3}\left(\frac{m_i^2}{2\mu_i} + \frac{\mu_i}{2}\right) + E_0(\mu_i),$$

$E_0(\mu_i)$ being eigenvalue of the operator $H_0 + V$. Quarks acquire constituent masses $\mu_i \sim \sqrt{\sigma}$ due to the string interaction in (12). The EH in the form of (12) does not include chiral-symmetry-breaking effects. A possible interplay with these effects should be carefully clarified in the future.

The physical mass $M_H$ of a hadron is

$$M_H = M_H^{(0)} + \sum_i C_i. \qquad (14)$$

The (negative) constants $C_i$ have the meaning of the constituent self-energies and are explicitly expressed in terms of string tension $\sigma$ [52]:

$$C_i = -\frac{2\sigma}{\pi\mu_i}\eta_i, \qquad (15)$$

where $\eta_q = 1$ and $\eta_s = 0.88$ is the correction factor due to nonvanishing current mass of the strange

[8] Technically, this is done using the auxiliary field (AF) approach to get rid of the square-root term in the Lagrangian [50, 51]. Applied to the QCD Lagrangian, this technique yields the EH for hadrons (mesons, baryons, pentaquarks) depending on auxiliary fields $\mu_i$. In practice, these fields are finally treated as $c$ numbers determined from (13).

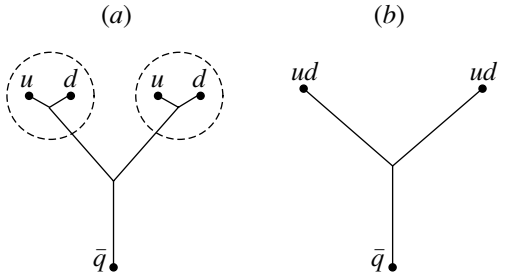


**Fig. 2.** The Jaffe–Wilczek reduction of the $uudd\bar{q}$ pentaquark ($a$) to the effective $[ud]^2\bar{q}$ problem ($b$).

quark. The self-energy corrections are due to constituent spin interaction with the vacuum background fields and equal zero for any scalar constituent.

Accuracy of the method is illustrated in Table 3, in which the results of the approach for baryon masses are compared with the lattice ones and experiment. One observes that the accuracy of the EH method for the three-quark systems is $\sim 100$ MeV or better. One can expect the same accuracy for the diquark–diquark–antiquark system. We shall comment on this point later.

## 7. PENTAQUARKS IN THE EFFECTIVE HAMILTONIAN METHOD

In the quark model, five quarks are connected by seven strings (Fig. 2$a$). In the diquark approximation, the short legs on this figure shrink to points and the five-quark system effectively reduces to the three-body one (Fig. 2$b$), studied within the EH approach in [55, 57].

Consider a pentaquark consisting of two identical diquarks with current mass $m_{[ud]}$ and antiquark with current mass $m_{\bar{q}}$ ($q = d, s$). In the hyperspherical formalism [58], the wave function $\psi(\boldsymbol{\rho}, \boldsymbol{\lambda})$ expressed in terms of the Jacobi coordinates $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$ can be written in symbolic shorthand as

$$\psi(\boldsymbol{\rho}, \boldsymbol{\lambda}) = R^{-5/2} \sum_K \chi_K(R) Y_{[K]}(\Omega), \qquad (16)$$

where $Y_{[K]}$ are eigenfunctions (hyperspherical harmonics) [59] of the angular momentum operator $\hat{K}(\Omega)$ on the six-dimensional sphere: $\hat{K}^2(\Omega) Y_{[K]} = -K(K+4) Y_{[K]}$, with $K$ being the grand orbital momentum. For identical diquarks, like $[ud]^2$, the lightest state must have a wave function antisymmetric under diquark space exchange. There are two possible pentaquark wave functions antisymmetric under diquark exchange, one (with lower energy) corresponding to the total orbital momentum and parity $L^P = 1^+$ and the second one (with higher energy) corresponding to $L^P = 0^-$. For a state with $L^P = 1^+$, $l_\rho = 1$, $l_\lambda = 0$, the wave function in the lowest hyperspherical approximation $K = 1$ reads

$$\psi = R^{-5/2} \chi_1(R) u_1(\Omega), \qquad (17)$$
$$u_1(\Omega) = \sqrt{\frac{8}{\pi^2}} \sin\theta \cdot Y_{1m}(\hat{\boldsymbol{\rho}}),$$

where $R^2 = \boldsymbol{\rho}^2 + \boldsymbol{\lambda}^2$. Here, one unit of orbital momentum between the diquarks is with respect to the $\rho$ variable, whereas the $\lambda$ variable is in an $S$ state. For a state with $L^P = 0^+$, $l_\rho = 1, l_\lambda = 1$, the wave function in the lowest hyperspherical approximation $K = 2$ is

$$\psi = R^{-5/2} \chi_2(R) u_2(\Omega), \qquad (18)$$
$$u_2(\Omega) = \frac{4}{\sqrt{\pi^3}} \sin\theta \cos\theta (\mathbf{n}_\rho \cdot \mathbf{n}_\lambda).$$

The functions $\chi_K(R)$, $K = 1, 2$, satisfy the Schrödinger equations. Written in terms of the variable $x = \sqrt{\mu} R$, where $\mu$ is an arbitrary scale of mass dimension which drops off in the final expressions, these equations read

$$\frac{d^2\chi_K(x)}{dx^2} + 2\left[E_0 + \frac{a_K}{x} - b_K x \right. \qquad (19)$$
$$\left. - \frac{(K+3/2)(K+5/2)}{2x^2}\right]\chi_K(x) = 0,$$

with the boundary condition $\chi_K(x) \sim O(x^{5/2+K})$ as $x \to 0$ and the asymptotic behavior $\chi_K(x) \sim \mathrm{Ai}((2b_K)^{1/3} x)$ as $x \to \infty$. The constants $a_K$ and $b_K$ in (19) are given by

$$a_K = R\sqrt{\mu} \int V_{\mathrm{C}}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) u_K^2 d\Omega, \qquad (20)$$
$$b_K = \frac{1}{R\sqrt{\mu}} \int V_{\mathrm{string}}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) u_K^2 d\Omega,$$

where

$$V_{\mathrm{C}}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = -\frac{2}{3}\alpha_s \sum_{i<j} \frac{1}{r_{ij}}. \qquad (21)$$

In the $Y$ shape in Fig. 2$b$, the strings meet at 120° in order to insure the minimum energy. This shape moves continuously to a two-leg configuration, where the legs meet at an angle larger than 120°. An explicit expression of $V_{\mathrm{string}}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ in terms of Jacobi variables is given in [56].

The mass of the $\Theta^+$ obviously depends on $m_{[ud]}$ and $m_s$. The current masses of the light quarks are relatively well known: $m_{u,d} \approx 0$, $m_s \approx 170$ MeV. The only other parameter of strong interactions is the effective mass of the diquark $m_{[ud]}$. In principle, this mass could be computed dynamically. Instead, one

can tune $m_{[ud]}$ to give the nucleon mass $M_N$ (in the quark–diquark approximation). Neglecting for simplicity Coulomb-like interaction, one obtains

$$M_N = M_N^{(0)} - \frac{2\sigma}{\pi\mu_d}, \tag{22}$$

where

$$M_N^{(0)} = \frac{m_{[ud]}^2}{2\mu_{[ud]}} + \frac{\mu_{[ud]}+\mu_d}{2} + \gamma\left(\frac{\sigma^2}{2}\frac{\mu_{[ud]}+\mu_d}{\mu_{[ud]}\mu_d}\right)^{1/3}. \tag{23}$$

In Eq. (23), $\mu_{[ud]}$ and $\mu_d$ are defined from the minimum condition (13),

$$\frac{\partial M_N^{(0)}}{\partial \mu_{[ud]}} = \frac{\partial M_N^{(0)}}{\partial \mu_d} = 0, \tag{24}$$

where $\gamma = 2.338$ is the first zero of the Airy function: $\mathrm{Ai}(-\gamma) = 0$. Equating $M_N$ to the experimental value of the $N-\Delta$ center of gravity ($M_N = 1.085$ GeV), we find that $m_{[ud]}$ varies in the interval $0 \le m_{[ud]} \le 300$ MeV, when $\sigma$ varies in the interval $0.15 \le \sigma \le 0.17$ GeV$^2$. The last value of $\sigma$ is preferred by meson spectroscopy, while the former one follows from the lattice calculations of [60]. In what follows, we use $\sigma = 0.15$ GeV$^2$ and explicitly include the Coulomb-like interaction between quark and diquarks.

For the pedagogy, let us first assume $m_{[ud]} = 0$. This assumption leads to the lowest $uudd\bar{d}$ and $uudd\bar{s}$ pentaquarks. If the current diquark masses vanish, then the $[ud]^2\bar{d}$ pentaquark is dynamically analogous to the $J^P = 1/2^-$ nucleon resonance and the $[ud]^2\bar{s}$ pentaquark is an analog of the $J^P = 1/2^-$ $\Lambda$ hyperon, with one important exception. The masses of $P$-wave baryons calculated using the EH method acquire the (negative) contribution $3C_q$ for $J^P = 1/2^-$ nucleons and $2C_q + C_s$ for the $J^P = 1/2^-$ $\Lambda$ hyperons. These contributions are due to the interaction of constituent spins with the vacuum chromomagnetic field.

However, the above discussion shows that the self-energies $C_{[ud]}$ equal zero for the scalar diquarks. This means that introducing any scalar constituent increases the pentaquark energy (relative to the $N$ and $\Lambda$ $P$-wave resonances). The numerical calculation for $m_{[ud]} = 0$ yields the mass of $[ud]^2\bar{s}$ pentaquark $\sim$2100 MeV (see Table 4 and subsequent discussion). Similar calculations yield the mass of $[ud]^2\bar{c}$ pentaquark $\sim$3250 MeV (for $m_c = 1.4$ GeV) and the mass of $[ud]^2\bar{b}$ pentaquark $\sim$6509 MeV (for $m_b = 4.8$ GeV)[61]. For illustration of accuracy of the AF formalism, in Table 4 are also shown the masses

**Table 4.** The pentaquark masses in the quark–diquark–diquark approximations (shown are the masses of $[ud]^2\bar{q}$ states ($q = d, s$) for $J^P = 1/2^+$ pentaquarks)

| | | $\mu_{[ud]}$ | $\mu_q$ | $M$ |
|---|---|---|---|---|
| $[ud]^2\bar{s}\ \frac{1}{2}^+$ | AF | 0.482 | 0.458 | 2.171 |
| | SSE | 0.463 | 0.468 | 2.070 |
| $[ud]^2\bar{d}\ \frac{1}{2}^+$ | AF | 0.476 | 0.415 | 2.091 |
| | SSE | 0.469 | 0.379 | 1.934 |

of $[ud]^2\bar{s}$ and $[ud]^2\bar{d}$ pentaquarks calculated using the spinless Salpeter equation (SSE):

$$H_{\mathrm{S}} = \sum_{i=1}^{3}\sqrt{\mathbf{p}_i^2 + m_i^2} + V,$$

$$M = M_0 - \frac{2\sigma}{\pi}\sum_{i=1}^{3}\frac{\eta_i}{\left\langle\sqrt{\mathbf{p}_i^2 + m_i^2}\right\rangle},$$

where $V$ is the same as in Eq. (12), $M_0$ is the eigenvalue of $H_{\mathrm{S}}$, $\langle(\mathbf{p}_{[ud]}^2 + m_{[ud]}^2)^{1/2}\rangle$, $\langle(\mathbf{p}_{\bar{q}}^2 + m_{\bar{q}}^2)^{1/2}\rangle$ are the average kinetic energies of diquarks and an antiquark, and $\eta_i$ are the same correction factors as in (15).[9] In Table 4 the quantities $\mu_{[ud]}$ and $\mu_q$ denote either the constituent masses calculated in the AF formalism using Eq. (15) or $\langle(\mathbf{p}_{[ud]}^2 + m_{[ud]}^2)^{1/2}\rangle$, $\langle(\mathbf{p}_{\bar{q}}^2 + m_{\bar{q}}^2)^{1/2}\rangle$ found from the solution of SSE. It is seen from Table 4 that these quantities agree with accuracy better than 5%. The diquark masses calculated by the two methods differ by 100 MeV for $([ud]^2\bar{s})$ and by 160 MeV for $[ud]^2\bar{d}$. The approximation of $V_{\mathrm{string}}$ mentioned above introduces a correction to the energy eigenvalues of $\le$30 MeV, so we conclude that the results obtained using the AF formalism and the SSE agree within $\sim$5%, i.e., the accuracy of the AF results for pentaquarks is the same as for the $q\bar{q}$ system (see, e.g., [64]).

When the $[ud]$ diquark mass is fitted on the baryon spectra, the pentaquark mass is found around 2.2 GeV in the JW picture [65]. Increasing $\alpha_s$ up to 0.6 (the value used in the Capstick–Isgur model [66]) decreases the $[ud]^2\bar{s}$ mass by $\sim$100 MeV. The hyperfine interaction due to $\sigma$ exchange between diquarks

[9] The numerical algorithm to solve the three-body SSE is based on an expansion of the wave function in terms of harmonic oscillator functions with different sizes [62]. In fact, to apply this technique to the three-body SSE, we need to use an approximation of the three-body potential $V_{\mathrm{string}}$ by a sum of the two- and one-body potentials (see [63]). This approximation, however, introduces a marginal correction to the energy eigenvalues.

and a strange antiquark can lower the $\Theta^+$ energy by $\sim$100 MeV for $g_\sigma^2/4\pi \sim 1$. Therefore, the lower bound for the pentaquark mass in the JW model is $\sim$1900 MeV, still 350 MeV higher than the observed $\Theta^+(1540)$ state.

We therefore conclude that the string dynamics alone in its simplified form predicts overly high masses of pentaquarks. A drastic modification of the present results requires a completely new dynamics, either that of the chiral soliton type [12] or introducing other multiquark clusters, like a triquark $ud\bar{s}$ plus a diquark $ud$ in a relative $P$ wave [41]. In the last model, pentaquark masses were computed using the Sakharov–Zeldovich mass formulas only. We believe that firmer conclusions must be obtained with dynamical calculations.

## 8. CONCLUSIONS

The existence of the pentaquark is, at the present time, an *experimental* issue. In particular, determining the $J^P$ quantum numbers of $\Theta$ and $\Xi_{3/2}$ pentaquarks would be a very useful discriminant for theoretical interpretations. Beyond pinning down the character of $\Theta^+(1540)$, the outstanding theoretical challenge is to understand possible pentaquark production mechanisms that give unambiguous signatures of pentaquarks. New experiments are required with higher statistics, varying the incident-beam momentum and establishing the spin and parity of pentaquarks before claiming solid evidence of the $S = 1$ baryon resonances. New results are expected in the near future: CLAS high-statistics proton data (just done), HERMES with double statistics (now running), COSY-TOF (five times more statistics by the end of 2004), KEK $\pi^+ p \to K\Theta^+$ experiment E359 (will run in May 2005), and some others.

Beyond spectroscopy, one can look forward to new insights into the production of pentaquark states in baryonic $B$ decays [67, 68]. Although $B$ decays into a baryon and antibaryon typically have branching ratios $\lesssim 10^{-4}$, the large number of events accumulated at the $B$ factories offer good opportunities for pentaquark detection that are distinct from photoproduction or kaon scattering that have been used so far.

From the theoretical point of view, the anomalous lightness of the $\Theta$ seems to unambiguously indicate a large role of the chiral-symmetry-breaking effects in pentaquarks. Roughly speaking, pentaquarks may contain a lot of almost massless pions—an effect that may considerably decrease their energies. The CQM and $\chi$SM are to a large degree complementary. Each of them reproduces certain aspects of hadronic physics and incorporates some features of QCD that are missing in the other. An approach of $\chi$SM totally neglects the confinement effects and concentrates on the pure chiral properties of baryons. The string dynamics alone in its simplified form seems to predict overly high masses of pentaquarks. Therefore, the existence of $\Theta$, if confirmed, provides a unique possibility to clarify the interplay between the quark and chiral degrees of freedom in light baryons.

## ACKNOWLEDGMENTS

I am very grateful to Yu.A. Simonov for numerous discussions concerning the topic of this paper.

This work was supported by the Russian Foundation for Basic Research, project nos. 03-02-17345, 04-02-17263, and the grant for Leading Scientific Schools no. 1774.2003.2.

## REFERENCES

1. T. Nakano (LEPS Collab.), in *PANIC 2002, Osaka, 2002*; T. Nakano *et al.*, Phys. Rev. Lett. **91**, 012002 (2003); hep-ex/0301020.
2. V. A. Shebanov (DIANA Collab.), *Talk at the Session of the Nuclear Physics Division of the Russian Academy of Sciences, Moscow, 2002*; V. V. Barmin *et al.* (DIANA Collab.), Yad. Fiz. **66**, 1763 (2003) [Phys. At. Nucl. **66**, 1715 (2003)]; hep-ex/0304040.
3. S. Stepanyan (CLAS Collab.), Phys. Rev. Lett. **91**, 252001 (2003); hep-ex/0307018.
4. V. Kubarovsky and S. Stepanyan (CLAS Collab.), *Talk presented at CIPANP 2003, New York, 2003*; hep-ex/0307088.
5. J. Barth *et al.* (SAPHIR Collab.), Phys. Lett. B **572**, 127 (2003); hep-ex/0307083.
6. A. E. Asratyan, A. G. Dolgolenko, and M. A. Kubantsev, Yad. Fiz. **67**, 704 (2004) [Phys. At. Nucl. **67**, 682 (2004)]; hep-ex/0309042.
7. A. Airapetian *et al.* (HERMES Collab.), Phys. Lett. B **585**, 213 (2004); hep-ex/0312044.
8. A. Aleev *et al.* (SVD Collab.), hep-ex/0401024.
9. M. Abdel-Bary *et al.* (COSY-TOF Collab.), Phys. Lett. B **595**, 127 (2004); hep-ex/0403011.
10. P. Zh. Aslanyan, V. N. Emelyanenko, and G. G. Rikhkvitzkaya, hep-ex/0403044.
11. S. Chekanov (ZEUS Collab.), in *YITP Workshop "Multiquark Hadrons; Four, Five, and More?", Kyoto, Japan, 2004*; hep-ex/0404007.
12. D. Diakonov, V. Petrov, and M. V. Polyakov, Z. Phys. A **359**, 305 (1997); M. Praszalowicz, Phys. Lett. B **575**, 234 (2003); hep-ph/0308114.
13. W. R. Gibbs, Phys. Rev. C **70**, 045208 (2004); nucl-th/0405024.
14. C. Alt *et al.* (NA49 Collab.), Phys. Rev. Lett. **92**, 042003 (2004); hep-ex/0310014.
15. A. Aktas *et al.* (H1 Collab.), Phys. Lett. B **588**, 17 (2004).

16. M. Kubantsev, *Talk at the International Workshop on Hadron Structure and QCD: From Low to High Energies, Repino, St. Petersburg, Russia, May 2004*.
17. D. Litvintsev, *Talk at BEACH 2004 Conference, Chicago, June–July 2004*.
18. I. M. Narodetskii and Yu. A. Simonov, Yad. Fiz. **28**, 1356 (1978) [Sov. J. Nucl. Phys. **28**, 698 (1978)].
19. E. W. Anderson *et al.*, Phys. Lett. B **29B**, 136 (1969).
20. A. R. Dzierba *et al.*, Phys. Rev. D **69**, 051901 (2004); hep-ph/0311125.
21. S. Nussinov, hep-ph/0307357.
22. R. A. Arndt, I. I. Strakovsky, and R. L. Workman, Phys. Rev. C **68**, 042201 (2003); nucl-th/0308012.
23. E. Golovich, Phys. Rev. D **4**, 262 (1971).
24. D. Diakonov and V. Petrov, Preprint No. 957, LPNI (1984).
25. D. Diakonov and V. Petrov, Phys. Rev. D **69**, 094011 (2004); hep-ph/0310212.
26. M. Praszalowicz, Phys. Lett. B **583**, 96 (2004); hep-ph/0311230.
27. J. Ellis, M. Karliner, and M. Praszalowicz, J. High Energy Phys. **0405**, 002 (2004); hep-ph/0401127.
28. J. Heidenbauer and J. Krein, Phys. Rev. C **68**, 052201 (2003); hep-ph/0309243.
29. F. Close, hep-ph/0311087.
30. D. Melikhov, S. Simula, and B. Stech, Phys. Lett. B **594**, 265 (2004); hep-ph/0405037.
31. S. Capstick, Ph. R. Page, and W. Roberts, Phys. Lett. B **570**, 185 (2003); hep-ph/0307019.
32. R. Matheus *et al.*, Phys. Lett. B **578**, 323 (2004); hep-ph/0309001; J. Sugiyama, T. Doi, and M. Oka, Phys. Lett. B **581**, 167 (2004); hep-ph/0309271.
33. F. Csikor *et al.*, J. High Energy Phys. **0311**, 070 (2003); hep-lat/0309090.
34. S. Sasaki, hep-lat/0310014.
35. T. W. Chiu and T. H. Hsieh, hep-ph/0403020; hep-ph/0404007.
36. N. Mathur *et al.*, hep-ph/0406196.
37. Fl. Stancu and D. O. Riska, Phys. Lett. B **575**, 242 (2003); hep-ph/0307010.
38. F. Huang *et al.*, Phys. Lett. B **586**, 69 (2004); hep-ph/0310040.
39. Fl. Stancu, Phys. Lett. B **595**, 269 (2004); hep-ph/0402044.
40. J. J. Dudek and F. E. Close, Phys. Lett. B **583**, 278 (2004); hep-ph/0311258.
41. M. Karliner and H. J. Lipkin, Phys. Lett. B **575**, 249 (2003); hep-ph/0307243.
42. Ya. B. Zeldovich and A. D. Sakharov, Yad. Fiz. **4**, 395 (1966) [Sov. J. Nucl. Phys. **4**, 283 (1966)].
43. R. L. Jaffe and F. Wilczek, Phys. Rev. Lett. **91**, 232003 (2003); hep-ph/0307341.
44. E. Shuryak and I. Zahed, hep-ph/0310270.
45. M. Gell-Mann, Phys. Lett. **8**, 214 (1964).
46. M. Anselmino *et al.*, Rev. Mod. Phys. **65**, 1199 (1993).
47. K. Rajagopal and F. Wilczek, hep-ph/0011333.
48. Yu. A. Simonov, in *Proceedings of the XVII Autumn School, Lisboa, Portugal, 1999*, Ed. by L. Ferreira, P. Nogueira, and J. I. Silva-Marco (World Sci., Singapore, 2000), p. 60; hep-ph/9911237.
49. I. M. Narodetskii, Yu. A. Simonov, M. A. Trusov, and A. I. Veselov, Phys. Lett. B **578**, 318 (2004); hep-ph/0310118.
50. A. M. Polyakov, *Gauge Fields and Strings* (Harwood, Chur, 1987).
51. L. Brink, P. Di Vecchia, and P. Howe, Nucl. Phys. B **118**, 76 (1977).
52. Yu. A. Simonov, Phys. Lett. B **515**, 137 (2001).
53. S. Aoki *et al.* (CP-PACS Collab.), Phys. Rev. D **67**, 034503 (2003); hep-lat/0206009.
54. C. M. Maynard (UKQCD Collab.) and D. G. Richard (LHP Collab.), hep-lat/0209165.
55. Yu. A. Simonov, Yad. Fiz. **66**, 363 (2003) [Phys. At. Nucl. **66**, 338 (2003)]; hep-ph/0205334.
56. I. M. Narodetskii, A. N. Plekhanov, and A. I. Veselov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 64 (2003) [JETP Lett. **77**, 58 (2003)].
57. I. M. Narodetskii and M. A. Trusov, Yad. Fiz. **67**, 783 (2004) [Phys. At. Nucl. **67**, 762 (2004)]; hep-ph/0307131.
58. M. Fabre de la Ripelle and Yu. A. Simonov, Ann. Phys. (N.Y.) **212**, 235 (1991).
59. Yu. A. Simonov, Yad. Fiz. **3**, 630, 1032 (1966) [Sov. J. Nucl. Phys. **3**, 461, 1032 (1966)].
60. T. T. Takahashi, H. Matsufuru, Y. Nemoto, and H. Suganuma, Phys. Rev. Lett. **86**, 18 (2001).
61. A. I. Veselov, private communication.
62. P. Nunberg, D. Prosperi, and E. Pace, Nucl. Phys. A **285**, 58 (1977).
63. B. Silvestre-Brac, C. Semay, I. M. Narodetskii, and A. I. Veselov, Eur. Phys. J. C **32**, 385 (2003).
64. V. L. Morgunov, A. V. Nefediev, and Yu. A. Simonov, Phys. Lett. B **459**, 653 (1999).
65. I. M. Narodetskii, C. Semay, B. Silvestre-Brac, *et al.*, Phys. At. Nucl. **68**, 536 (2005).
66. S. Capstick and N. Isgur, Phys. Rev. D **34**, 2809 (1986).
67. J. Rosner, Phys. Rev. D **69**, 094014 (2004); hep-ph/0312269.
68. T. R. Browder, I. R. Klebanov, and D. Marlow, Phys. Lett. B **587**, 62 (2004); hep-ph/0401115.

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 758–770. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 790–803.*


# QCD and Models on Multiplicities in $e^+e^-$ and $p\bar{p}$ Interactions *

**I. M. Dremin****

*Lebedev Institute of Physics, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*

Received April 23, 2004; in final form, September 23, 2004

**Abstract**—A brief survey of theoretical approaches to describing multiplicity distributions in high-energy processes is given. It is argued that the multicomponent nature of these processes leads to some peculiar characteristics observed experimentally. Predictions for LHC energies are presented. It is shown that similarity of the energy dependence of average multiplicities in different reactions is not enough alone to suggest the universal mechanism of particle production in strongly interacting systems. Other characteristics of multiplicity distributions depend on the nature of colliding partners. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Multiplicity distributions are the most general characteristics of any high-energy process of multiparticle production. They depend on the nature of colliding particles and on their energy. Nevertheless, it has been found that their shapes possess some common qualitative features in all reactions studied. At comparatively low energies, these distributions are relatively narrow and have sub-Poissonian shapes. With increasing energy, they widen and fit the Poisson distribution. At even higher energies, the shapes become super-Poissonian; i.e., their widths are larger than that for the Poisson distribution. The width increases with energy and, moreover, some shoulder-like substructures appear.

Their origin is usually ascribed to multicomponent contents of the process. In a QCD description of $e^+e^-$ processes, these could be subjets formed inside quark and gluon jets (for reviews, see, e.g., [1, 2]). In phenomenological approaches, the multiplicity distribution in a single subjet is sometimes approximated by the negative binomial distribution (NBD) first proposed for hadronic reactions in [3]. For hadron-initiated processes, these peculiarities are often ascribed to multiple parton–parton collisions [4–8], which could lead, e.g., to two-, three-, ... ladder formation of the dual parton model (DPM) [9] or quark–gluon string model (QGSM) [10–12] and/or to different (soft, hard) types of interactions [13, 14]. They become increasingly important as collision energy is increased. These subprocesses are related to the matter state during the collision (e.g., there are speculations about nonhomogeneous matter distribution in impact parameters [15], not to speak of quark–gluon plasma [16] behaving as a liquid [17], etc.).

The theoretical description of the subprocesses differs drastically in $e^+e^-$ and $p\bar{p}$ processes. Jet evolution in $e^+e^-$ is well described by perturbative QCD equations with the only adjustable parameter $\Lambda_{\mathrm{QCD}}$. This parameter is approximately known from other characteristics and is, therefore, bounded. The production of perturbative gluons and quark–antiquark pairs can be described in terms of dipole or "antenna" radiation. Color interference plays a crucial role. Many predictions of the perturbative QCD approach have been confirmed by experimental data. Concerning multiplicity distributions, their shape in $e^+e^-$ processes is known in QCD only implicitly from studies of their moments. This is determined by the fact that the QCD evolution equations are formulated in terms of the generating function. They can be rewritten as equations for the moments but not directly for probabilities of $n$-parton emission. Their solutions up to higher order perturbative QCD [18] have predicted some completely new features of the moments. The moments also contain complete information about the distribution. However, the reconstruction of the shape of the distribution from them is not a trivial task. Some attempts to solve the inverse problem [19, 20] and directly get the shape of the multiplicity distribution were successful only in the lowest order perturbative QCD approximations with several additional assumptions.

The situation with hadronic processes is, in some respect, more complicated. The confinement property is essential, and perturbative QCD methods cannot be applied directly. Therefore, some models have been

*This article was submitted by the author in English.

**e-mail: dremin@lpi.ru

developed. Hadron interactions used to be considered as proceeding via collisions of their constituent partons. In preparton times, their role was played by pions, and one-meson-exchange model [21] dominated. Pions were treated as hadron constituents. Their high-energy interaction produced a ladder of one-pion $t$-channel exchanges with blobs of low-energy pion–pion interactions. This is the content of the multiperipheral model. These blobs were first interpreted as $\rho$ mesons [22] and later called fireballs [23], clusters [24], or clans [25] when higher mass objects were considered. Multiperipheral dynamics tells us that the length of a multiperipheral chain fluctuates and the number of these blobs is distributed according to the Poisson law. It was argued that its convolution with the distribution of the number of pions produced in each center can lead to the NBD of created particles. This supposition fits experimental data on multiplicity distributions of $pp$ and $p\bar{p}$ reactions at tens of GeV quite well. However, at higher energies, this fit by a single NBD becomes unsatisfactory. A shoulder appears at high multiplicities. Sums of NBD with different parameters were used [14] to get agreement with experiment. Better fits are achieved at the expense of a larger number of adjustable parameters. These shortcomings can be minimized if one assumes that each high-energy binary parton collision is independent of some others simultaneously proceeding. With this supposition, the whole process is described as a set of independent pair parton interactions (IPPI model, proposed in [26]). The effective energy of a pair of partons does not depend on how many other pairs interact and what these interacting partons (quarks or gluons) are. The number of adjustable parameters does not increase compared to a single NBD if the probabilities of $j$ pairs' interactions and the number of active pairs are known.

Earlier, a somewhat different way to account for multiple parton collisions, which generalized the multiperipheral one-ladder model in the framework of the eikonal approximation, was proposed in DPM and QGSM (the latter one takes into account the Reggeization of exchanged particles). They differ from IPPI by probabilities of processes with different number of active parton pairs and by multiplicity distributions of final particles. Also, the Lund model [27] with its parton cascades and the string hadronization due to a linearly increasing QCD potential has been extremely successful in describing many features of multiparticle production. Several Monte Carlo programs implement this model to provide some hints to experimentalists at present-day experiments and give predictions at even higher energies (like PYTHIA, FRITIOF, etc.). Some assumptions have to be used for the hadronization of partons at the final stage (e.g., these assumptions differ in the PYTHIA and the HERWIG cluster model). More important, the predictions at higher energies (in particular, for multiplicity distributions) also differ in these models, and it is necessary to try various approaches and confront them with experiment when the LHC enters operation. Being interesting in themselves, these fits provide the background of so-called minimum bias events for "triggered" experiments.

## 2. MOMENTS OF MULTIPLICITY DISTRIBUTIONS

The shape of multiplicity distributions is so complicated that it is difficult to get any analytical expression for it from the solution of QCD equations. It has been obtained only in the simplest perturbative approximations [19, 20]. In particular, it has been demonstrated [20] that the recoil has a profound effect on the multiplicity distribution in QCD jets. It becomes much narrower than according to the leading perturbative term, where energy conservation is not taken into account. However, no shoulders appear. On the contrary, the tail of the distribution is more strongly suppressed.

An alternative and, in some sense, more accurate approach is proposed by studies of moments of the distribution. One can get some QCD predictions for these moments [18] up to high orders in the perturbative expansion. The moments of various ranks contain complete information about multiplicity distributions. Hence, the shape and energy evolution of multiplicity distributions can be quantitatively described by the rank dependence and energy behavior of their moments. Moment analysis of multiplicity distributions can also be performed for models of hadronic processes as well as for experimental data. Therefore, this approach is common to all processes and methods of analysis.

To introduce moments on the most general grounds, we write the generating function $G(E,z)$ of the multiplicity distribution $P(n,E)$:

$$G(E,z)=\sum_{n=0}^{\infty}P(n,E)(1+z)^n. \qquad (1)$$

The multiplicity distribution is obtained from the generating function as

$$P(n)=\frac{1}{n!}\frac{d^nG(E,z)}{dz^n}\bigg|_{z=-1}. \qquad (2)$$

In what follows, we will use the so-called unnormalized factorial $\mathcal{F}_q$ and cumulant $\mathcal{K}_q$ moments defined according to the formulas

$$\mathcal{F}_q=\sum_n P(n)n(n-1)\dots(n-q+1) \qquad (3)$$

$$= \left.\frac{d^q G(E,z)}{dz^q}\right|_{z=0},$$

$$\mathcal{K}_q = \left.\frac{d^q \ln G(E,z)}{dz^q}\right|_{z=0}. \quad (4)$$

They determine, respectively, the total and genuine (i.e., irreducible to lower) order correlations among the particles produced (for more details, see [2, 28]). The first factorial moment defines the average multiplicity $\mathcal{F}_1 = \langle n \rangle$, the second one is related to the width (dispersion) of the multiplicity distribution $\mathcal{F}_2 = \langle n(n-1) \rangle$, etc. Factorial and cumulant moments are not independent. They are related by the formula

$$\mathcal{F}_q = \sum_{m=0}^{q-1} C_{q-1}^m \mathcal{K}_{q-m} \mathcal{F}_m, \quad (5)$$

where

$$C_{q-1}^m = \frac{(q-1)!}{m!(q-m-1)!} \quad (6)$$

are the binomial coefficients.

Since both $\mathcal{F}_q$ and $\mathcal{K}_q$ strongly increase with their rank and energy, the ratio

$$H_q = \mathcal{K}_q / \mathcal{F}_q, \quad (7)$$

first introduced in [18], is especially useful due to partial cancellation of these dependences. More important is that some valuable predictions about its behavior can be obtained in perturbative QCD. Also, it will be shown below that $H_q$ moments of the IPPI model depend on a smaller number of its adjustable parameters than factorial and cumulant moments. Thus, even though $\mathcal{F}_q$, $\mathcal{K}_q$, and $H_q$ are interrelated, they can provide knowledge about different facets of the same multiplicity distribution.

It is easy to find the ratio $H_q$ from iterative formulas,

$$H_q = 1 - \sum_{p=1}^{q-1} \frac{\Gamma(q)}{\Gamma(p+1)\Gamma(q-p)} H_{q-p} \frac{\mathcal{F}_p \mathcal{F}_{q-p}}{\mathcal{F}_q}, \quad (8)$$

once the factorial moments have been evaluated.

The factorial moments $\mathcal{F}_q$ are always positive by definition [Eq. (3)]. The cumulant moments $\mathcal{K}_q$ can change sign. They are equal to zero for the Poisson distribution. Consequently, $H_q = 0$ in this case.

Let us emphasize that $H_q$ moments are very sensitive to minute details of multiplicity distributions and can be used to distinguish between different models and experimental data.

## 3. QCD ON MOMENTS IN $e^+e^-$ COLLISIONS

All moments can be calculated from the generating function as explained above. The generating functions for quark and gluon jets satisfy definite equations in perturbative QCD (see [2, 19]). They are

$$G'_G = \int_0^1 dx K_G^G(x) \gamma_0^2 [G_G(y + \ln x) \times G_G(y + \ln(1-x)) - G_G(y)] + n_f \int_0^1 dx K_G^F(x) \gamma_0^2 [G_F(y + \ln x) \times G_F(y + \ln(1-x)) - G_G(y)], \quad (9)$$

$$G'_F = \int_0^1 dx K_F^G(x) \gamma_0^2 [G_G(y + \ln x) \times G_F(y + \ln(1-x)) - G_F(y)], \quad (10)$$

where the labels $G$ and $F$ correspond to gluons and quarks, the energy scale of the process is defined by $y = \ln(Q/Q_0)$, $Q = p\theta$ is a virtuality of a jet, $p \approx \sqrt{s}/2$ is its momentum, $\theta$ is an opening angle, and $Q_0 = \text{const}$. Here, $G'(y) = dG/dy$, $n_f$ is the number of active flavors,

$$\gamma_0^2 = \frac{2N_c \alpha_s}{\pi}. \quad (11)$$

The running coupling constant in the two-loop approximation is

$$\alpha_s(y) = \frac{2\pi}{\beta_0 y}\left(1 - \frac{\beta_1}{\beta_0^2}\frac{\ln 2y}{y}\right) + O(y^{-3}), \quad (12)$$

where

$$\beta_0 = \frac{11N_c - 2n_f}{3}, \quad (13)$$

$$\beta_1 = \frac{17N_c^2 - n_f(5N_c + 3C_F)}{3}.$$

The kernels of the equations are

$$K_G^G(x) = \frac{1}{x} - (1-x)[2 - x(1-x)], \quad (14)$$

$$K_G^F(x) = \frac{1}{4N_c}[x^2 + (1-x)^2], \quad (15)$$

$$K_F^G(x) = \frac{C_F}{N_c}\left[\frac{1}{x} - 1 + \frac{x}{2}\right], \quad (16)$$

$N_c = 3$ is the number of colors, and $C_F = (N_c^2 - 1)/2N_c = 4/3$ in QCD.

Here, one can get equations for any moment of the multiplicity distribution, for both quark and gluon

jets. One should just equate the terms with the same powers of $u = 1 + z$ on both sides of the equations, where expressions (1) are substituted for both quarks and gluons.

In particular, the nontrivial energy dependence of mean multiplicity in quark and gluon jets has been predicted. Within two lowest perturbative QCD approximations, it has a common behavior

$$\langle n_{G,F}\rangle = A_{G,F} y^{-a_1 c^2} \exp(2c\sqrt{y}), \qquad (17)$$

where $A_{G,F} = \text{const}$, $c = (4N_c/\beta_0)^{1/2}$, and $a_1 \approx 0.3$.

The main features of the solutions can be demonstrated in gluodynamics where only the first equation with $n_f = 0$ is considered. At asymptotically high energies, it can be reduced [19] to the differential equation

$$[\ln G(y)]'' = \gamma_0^2[G(y) - 1]. \qquad (18)$$

From the definitions of moments, one can easily guess that this equation determines the asymptotic behavior of $H_q$, because $\ln G$ on the left-hand side gives rise to $\mathcal{K}_q$ and $G$ on the right-hand side to $\mathcal{F}_q$. The second derivative on the left-hand side would result in the factor $q^2$. Thus, it can be shown [18] that asymptotical ($y \to \infty$) values of $H_q$ moments are positive and decrease as $q^{-2}$. At present energies, they become negative at some values of $q$ and reveal a negative minimum at

$$q_{\min} = \frac{1}{h_1\gamma_0} + 0.5 + O(\gamma_0), \qquad (19)$$

where $h_1 = b/8N_c = 11/24$, $b = 11N_c/3 - 2n_f/3$. At $Z^0$ energy $\alpha_s \approx 0.12$, and this minimum is located at about $q \approx 5$. It moves to higher ranks with energy increase, because the coupling strength decreases. Some hints to possible oscillations of $H_q$ vs. $q$ at higher ranks at LEP energies were predicted in [18]. They were obtained with account of recoil effects in higher order perturbative QCD and stressed the importance of both nonsingular terms in kernels (14), (16) of the equations and energy conservation in high-energy processes so often mentioned by Andersson (see [27]). The approximate solution of the gluodynamics equation for the generating function [29] agrees with this conclusion and predicts the oscillating behavior at higher ranks. The same conclusions were obtained from an exact solution to equations for quark and gluon jets in the framework of fixed coupling QCD [30]. A recent exact numerical solution of the gluodynamics equation in a wide energy interval [31] coincides with the qualitative features of multiplicity distributions described above. These oscillations were confirmed by experimental data for $e^+e^-$ collisions and found also for hadron-initiated processes first in [32], later in [33], and most recently in [34]. This will be demonstrated below in Fig. 10. Let us mention that the oscillations shape depends crucially on the particular form of QCD kernels (14)–(16), and their observation confirms the basic principles of QCD.

However, one should be warned that the amplitudes of oscillations strongly depend on a multiplicity distribution cutoff due to limited experimental statistics (or by other reasoning) if it is done at rather low multiplicities [35]. Usually, there are no such cutoffs in analytical expressions for $H_q$. One can control the influence of cutoffs by shifting them appropriately. The qualitative features of $H_q$ behavior persist nevertheless. In what follows, we consider very high energy processes where the cutoff due to experimental statistics is practically insignificant. Numerical estimates of the relation between the maximum multiplicity measured and effective ranks of the moments are given in the Appendix.

## 4. NEGATIVE BINOMIAL DISTRIBUTION AND IPPI MODEL

Independently of progress in perturbative QCD calculations, the NBD fits of multiplicity distributions were attempted for both $e^+e^-$ and $pp(p\bar{p})$ collisions [14, 36, 37]. The single NBD parametrization is

$$P_{\text{NBD}}(n, E) = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} \left(\frac{m}{k}\right)^n \left(1 + \frac{m}{k}\right)^{-n-k}, \qquad (20)$$

where $\Gamma$ denotes the gamma function. This distribution has two adjustable parameters $m(E)$ and $k(E)$, which depend on energy. However, the simple fit by formula (20) is valid until shoulders appear. In that case, this formula is often replaced [14] by the hybrid NBD which simply sums up two or more expressions like (20). Each of them has its own parameters $m_j, k_j$. These distributions are weighted with the energy-dependent probability factors $w_j$ which sum up to 1. Correspondingly, the number of adjustable parameters drastically increases if the distributions are completely unrelated.

It was proposed recently [26] that hadron interactions can be represented by a set of independent pair parton interactions (IPPI model). This means that colliding hadrons are considered as "clouds" of partons which interact pairwise. It is assumed that each binary parton collision is described by the same NBD distribution. The only justification for this supposition is given by previous fits of multiplicity distributions at lower energies. Then the convolution of these distributions, subject to a condition that the sum of

**Table 1.** The values of $w_j$ according to (22) (left-hand side) and (27) (right-hand side)

| $j_{\max}$ | 3 | 4 | 5 | 6 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| $w_1$ | 0.544 | 0.519 | 0.509 | 0.504 | 0.562 | 0.501 | 0.450 | 0.410 |
| $w_2$ | 0.295 | 0.269 | 0.259 | 0.254 | 0.278 | 0.255 | 0.236 | 0.219 |
| $w_3$ | 0.161 | 0.140 | 0.131 | 0.128 | 0.160 | 0.153 | 0.152 | 0.147 |
| $w_4$ | 0 | 0.072 | 0.067 | 0.065 | 0 | 0.091 | 0.100 | 0.104 |
| $w_5$ | 0 | 0 | 0.034 | 0.033 | 0 | 0 | 0.062 | 0.073 |
| $w_6$ | 0 | 0 | 0 | 0.016 | 0 | 0 | 0 | 0.047 |

binary-collision multiplicities is the total multiplicity $n$, leads [26] to a common distribution

$$P(n;m,k)=\sum_{j=1}^{j_{\max}} w_j P_{\mathrm{NBD}}(n;jm,jk). \qquad (21)$$

This is the main equation of the IPPI model. The NBD distribution on its right-hand side with products $jm$ and $jk$ in its arguments is a consequence of the property of the convolution of NBDs to produce new NBD. It can be easily shown if one considers the NBD generating function. One gets a sum of negative binomial distributions with shifted maxima and larger widths for a larger number of collisions. No new adjustable parameters appear in the distribution for $j$ pairs of colliding partons. The probabilities $w_j$ are determined by collision dynamics and, in principle, can be evaluated if some model is adopted (e.g., see [11, 12]). Independence of parton pair interactions implies that, at very high energies, $w_j$ is a product of $j$ probabilities $w_1$ for one pair. Then, from the normalization condition

$$\sum_{j=1}^{j_{\max}} w_j=\sum_{j=1}^{j_{\max}} w_1^j=1, \qquad (22)$$

one can find $w_1$ if $j_{\max}$, which is determined by the maximum number of parton interactions at a given energy, is known. This is the only new parameter. It depends on energy. Thus, three parameters are sufficient to describe multiplicity distributions at any energy. Moreover, asymptotically, $j_{\max}\to\infty$ and $w_1=0.5$ according to (22).

The factorial moments of the distribution (21) are

$$\mathcal{F}_q=\sum_{j=1}^{j_{\max}} w_j\frac{\Gamma(jk+q)}{\Gamma(jk)}\left(\frac{m}{k}\right)^q=f_q(k)\left(\frac{m}{k}\right)^q \qquad (23)$$

with

$$f_q(k)=\sum_{j=1}^{j_{\max}} w_j\frac{\Gamma(jk+q)}{\Gamma(jk)}. \qquad (24)$$

For $H_q$ moments, one gets

$$H_q=1-\sum_{p=1}^{q-1}\frac{\Gamma(q)}{\Gamma(p+1)\Gamma(q-p)}H_{q-p}\frac{f_pf_{q-p}}{f_q}. \qquad (25)$$

Note that, according to Eqs. (24) and (25), $H_q$ are functions of the parameter $k$ only and do not depend on $m$ in the IPPI model. This remarkable property of $H_q$ moments provides an opportunity to fit them with one parameter. It is nontrivial because the oscillating shapes of $H_q$ are quite complicated as shown below.

Once the parameter $k$ is found from fits of $H_q$, it is possible to get another parameter $m$ rewriting Eq. (23) as follows:

$$m=k\left(\frac{\mathcal{F}_q}{f_q(k)}\right)^{1/q}. \qquad (26)$$

This formula is a sensitive test for the whole approach, because it states that the definite ratio of $q$-dependent functions to the power $1/q$ becomes $q$-independent if the model is correct. Moreover, this statement should be valid only for those values of $k$ which are determined from $H_q$ fits. Therefore, it can be considered as a criterion of a proper choice of $k$ and of the model validity, in general. This criterion of constancy of $m$ happens to be extremely sensitive to the choice of $k$ as shown below.

In [12], the energy dependence of the probabilities $w_j$ was estimated according to the multiladder-exchange model [11] (its various modifications are known as DPM or QGSM). They are given by the following normalized expressions:

$$w_j(\xi_j)=\frac{p_j}{\sum\limits_{j=1}^{j_{\max}} p_j} \qquad (27)$$

$$=\frac{1}{jZ_j\left(\sum\limits_{j=1}^{j_{\max}} p_j\right)}\left(1-e^{-Z_j}\sum_{i=0}^{j-1}\frac{Z_j^i}{i!}\right),$$

where

$$\xi_j=\ln(s/s_0j^2),\quad Z_j=\frac{2C\gamma}{R^2+\alpha'_P\xi_j}\left(\frac{s}{s_0j^2}\right)^{\Delta} \qquad (28)$$

with numerical parameters obtained from fits of experimental data on total and elastic scattering cross sections: $\gamma=3.64\ \mathrm{GeV}^{-2}$, $R^2=3.56\ \mathrm{GeV}^{-2}$, $C=1.5$, $\Delta=\alpha_P-1=0.08$, $\alpha'_P=0.25\ \mathrm{GeV}^{-2}$, and $s_0=1\ \mathrm{GeV}^2$. It is seen that each $w_j$ depends on six adjustable parameters in these models.

Below, we will use both possibilities (22) and (27) in our attempts to describe experimental data. The

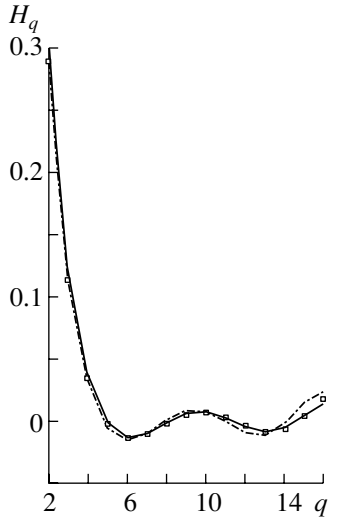


**Fig. 1.** A comparison of $H_q$ moments derived from experimental data at 1.8 TeV (squares) with their values calculated with parameter $k = 4.4$ (dash-dotted curve) and 3.7 (solid curve) [26].

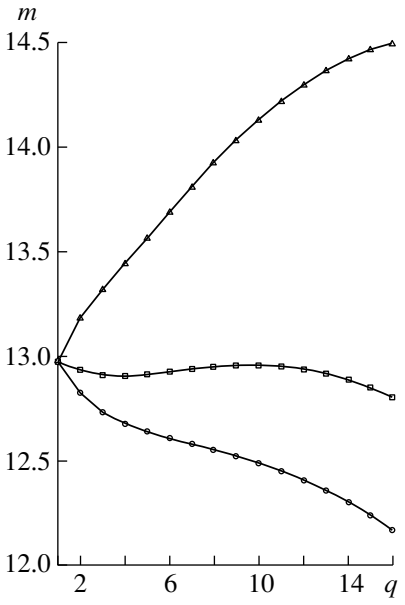


**Fig. 2.** The $q$ dependence of $m$ for $k = 4.4$ (squares), 3.7 (circles), and 7.5 (triangles) [26].

probabilities $w_j$ are different for them (see Table 1). In the IPPI model they decrease exponentially with the number of active partons, while in the ladder models they are inversely proportional to this number with additional suppression at large $j$ due to the term in parentheses in (27). This is the result of the modified eikonal approximation. Let us stress that, when we use expressions (27) for probabilities, this does not directly imply comparison with DPM–QGSM because in our case the NBD is chosen for the multiplicity distribution in a single ladder, while it is a Poisson distribution for resonances created in a ladder with a fluctuating length in DPM–QGSM. Thus, we will call it the modified ladder model.

We show the values $w_j$ for $j_{\max} = 3$–6 pairs calculated according to Eq. (22) on the left-hand side of Table 1 and according to Eq. (27) on its right-hand side. These values of $j_{\max}$ are chosen because previous analysis of experimental data [8] has shown that two pairs become active at an energy of about 120 GeV and the number of binary collisions increases with increasing energy. Thus, precisely these numbers will be used in comparison with experiment at higher energies. In particular, we shall choose $j_{\max} = 3$ at 300 and 546 GeV, 4 at 1000 and 1800 GeV, 5 at 14 TeV, and 6 at 100 TeV (see below).

One can clearly see the difference between the two approaches. The value of $w_1$ is always larger than 0.5 in the IPPI model, while it can become less than 0.5 in the (modified) ladder model [9, 11] at high energies. In the ladder model, $w_j$ depend explicitly on energy (not only on the $j_{\max}$ cutoff). We show their values at 546 and 1800 GeV in the right-hand side columns of $j_{\max} = 3$ and 4. Those at 300 and 1000 GeV are larger for $w_1$ by about 1% and smaller for $w_3$ by about 3%. When energy increases, the processes with a larger number of active pairs play a more important role in the modified ladder approach compared to the IPPI model. Thus, the $j_{\max}$ cutoff is also more essential there.

In principle, one can immediately try a two-parameter fit of experimental multiplicity distributions using Eq. (21) if $w_j$ are known according to Eqs. (22) or (27). However, the use of their moments is preferred.

## 5. COMPARISON WITH EXPERIMENT

Let us begin with hadronic collisions and then compare them with other processes.

We have compared [26] IPPI model conclusions with experimental (but extrapolated [8, 38] to the full phase space) multiplicity distributions of the E735 Collaboration [39] for $p\bar{p}$ collisions at Tevatron energies 300, 546, 1000, and 1800 GeV. The multiplicity of charged particles was divided by 2 to get the multiplicity of particles with the same charge. Then the above formulas for moments were used. Correspondingly, the parameters $m$ and $k$ refer to these distributions. The parameter $j_{\max}$ is chosen according to prescriptions discussed above.

Factorial and $H_q$ moments were obtained from experimental data on $P(n)$ according to Eqs. (3) and (8). Experimental $H_q$ moments were fitted by Eq. (25) to get the parameters $k(E)$. We show in Fig. 1 how

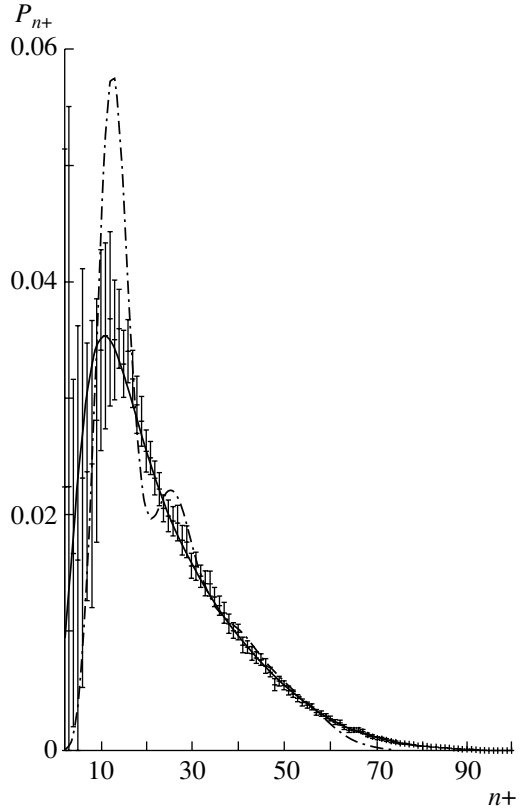


**Fig. 3.** The same-charge multiplicity distribution at 1.8 TeV, its fit at $m = 12.94, k = 4.4$. The dash-dotted curve demonstrates what would happen if the NBD were replaced by a Poisson distribution [26].

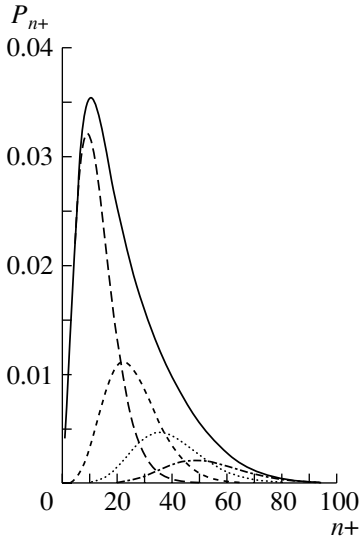


**Fig. 4.** The decomposition of the fit in Fig. 3 to one, two, three, and four parton–parton collisions (corresponding curves are ordered from left to right) [26].

perfect these fits are at 1.8 TeV for $k$ equal to 3.7 (solid curve) and 4.4 (dash-dotted curve). At this energy, we considered four active parton pairs with $w_j$

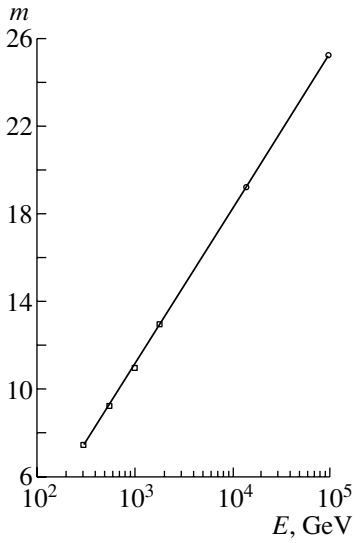


**Fig. 5.** The energy dependence of $m$ (squares) and its linear extrapolation (circles at 14 and 100 TeV) [26].

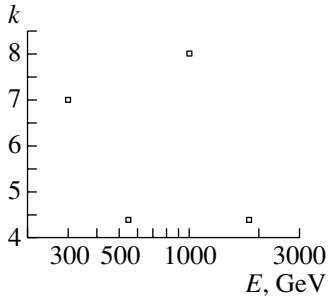


**Fig. 6.** The values of $k$ as calculated with $w_j$ satisfying relation (22) [26].

given by Eq. (22). It is surprising that oscillations of $H_q$ moments are well reproduced with one adjustable parameter $k$. The general tendency of this quite complicated oscillatory dependence is clearly seen.

With these values of the parameter $k$, we have checked whether $m$ is constant as a function of $q$ as required by Eq. (26). The $m(q)$ dependence is shown in Fig. 2 for the same values of $k$ and for the much larger value 7.5. The constancy of $m$ is fulfilled with an accuracy better than 1% for $k = 4.4$ up to $q = 16$. The upper and lower curves in Fig. 2 demonstrate clearly that this condition bounds substantially admissible variations of $k$.

The same-charge multiplicity distribution at 1.8 TeV has been fitted with parameters $m = 12.94$ and $k = 4.4$ as shown in Fig. 3 (solid curve). To estimate the accuracy of the fit, we calculated $\sum_{n=1}^{125}(P_{\text{theor}}(n) - P_{\text{exp}}(n))^2/\Delta^2$ over all 125 experimental points. Here, $P_{\text{theor}}$ and $P_{\text{exp}}$ are theoretical and experimental distributions and $\Delta$ is the total

experimental error. It includes both statistical and systematic errors. Note that the latter ones are large at low multiplicities in E735 data. This sum is equal to 50 for 125 degrees of freedom. No minimization of it was attempted. This is twice as good as the three-parameter fit by the generalized NBD considered in [40]. A Poisson distribution of particles in binary collisions is completely excluded. This is shown in Fig. 3 by the dash-dotted curve.

In Fig. 4, we also show a breakdown of the total distribution into contributions due to a different number of interacting pairs, thus illustrating the mechanism underlying the formation of the tails of the multiplicity distribution.

The same procedure has been applied to data at energies 300, 546, and 1000 GeV. As stated above, we have assumed that three binary parton collisions are active at 300 and 546 GeV and four at 1000 GeV. We plot in Figs. 5 and 6 the energy dependence of parameters $m$ and $k$. The parameter $m$ increases with energy logarithmically (Fig. 5). This is expected because the increase in $M_1 = \sum_{j=1}^{j_{\max}} w_j j$ due to increasing number of active pairs at these energies leads to a somewhat faster than logarithmic increase in average multiplicity in accordance with experimental observations. The energy dependence of $k$ is more complicated and rather irregular (Fig. 6).

We tried to ascribe the latter to the fact that the effective values of $k$, which we actually find from the above-described fits, depend on the effective number of parton interactions, i.e., on $w_j$ variation at a threshold. The threshold effects can be important in this energy region. Then, the simple relation (22) is invalid. This influences the functions $f_q(k)$ (24) and, consequently, $H_q$ calculated from Eq. (25). One can reduce the effective number of active pairs to about 2.5 at 300 GeV and 3.5 at 1000 GeV if chooses the following values of $w_j$: 0.59, 0.34, 0.07 at 300 GeV and 0.54, 0.29, 0.14, 0.03 at 1000 GeV instead of those calculated according to (22) and shown in Table 1. This gives rise to values of $k$ which are not drastically different from previous ones. However, the quality of fits becomes worse. Fits with two active pairs at 300 GeV and three pairs at 1000 GeV fail completely.

Hence, we have to conclude that this effect results from dynamics of hadron interactions which is not yet understood and should be incorporated in the model. The preliminary explanation of this effect could be that, at the thresholds of a new pair formation, the previous active pairs produce more squeezed multiplicity distributions due to smaller phase-space room available for them because of a newcomer. Therefore, the single pair dispersion would decrease and the $k$ values would increase. It would imply that thresholds

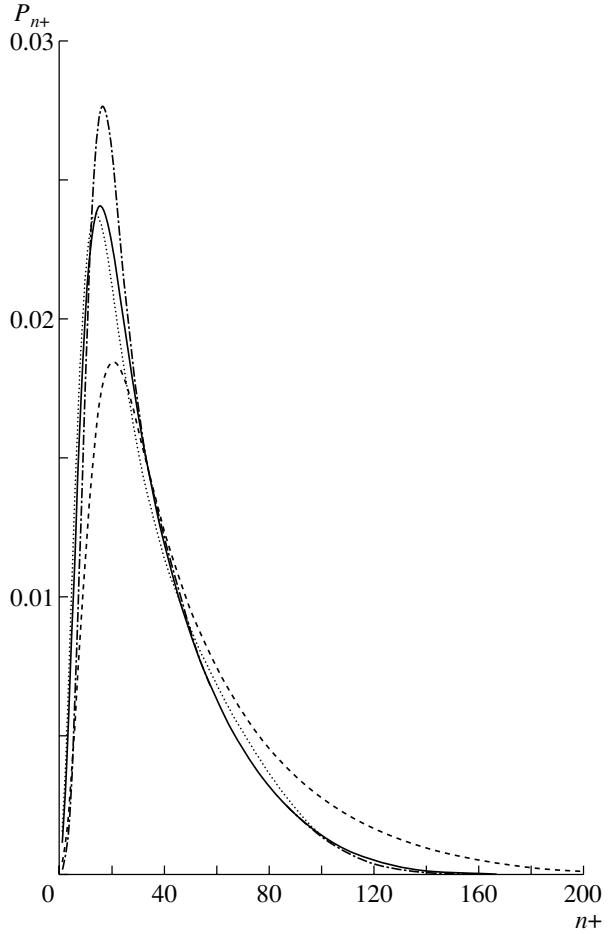


**Fig. 7.** The same-charge multiplicity distributions at 14 and 100 TeV obtained by extrapolation of parameters $m$ and $k$ with five active pairs at 14 TeV and six at 100 TeV [for the IPPI model, (solid curve) 14 TeV, $k = 4.4$; (dash-dotted curve) 14 TeV, $k = 8$; and (dashed curve) 100 TeV, $k = 4.4$; for the modified ladder model, (dotted curve) 14 TeV, $k = 4.4$] [26].

are marked not only by the change in $w_j$ shown in Table 1 but also by the variation of the parameter $k$.

The threshold effects become less important at higher energies. We assume that there are five active pairs at 14 TeV and six at 100 TeV. Then we extrapolate to these energies. The parameter $m$ becomes equal to 19.2 at 14 TeV and 25.2 at 100 TeV if a logarithmic dependence is adopted as shown in Fig. 5 by the straight line. We choose two values of $k$ equal to 4.4 and 8 since we do not know which one is responsible for thresholds. The predicted multiplicity distributions with these parameters are plotted in Fig. 7. The oscillations of $H_q$ still persist at these energies (see Fig. 8). The minima are, however, shifted to $q = 6$ at 14 TeV and 7 at 100 TeV as expected.

The fit at 1.8 TeV with an approximation of $w_j$ according to the modified ladder model (27) with the NBD for a binary parton collision is almost as successful as the fit with values of $w_j$ given by the

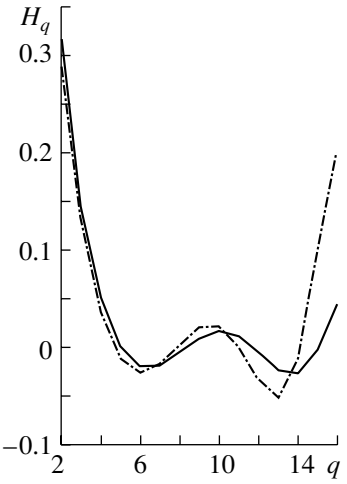


**Fig. 8.** The behavior of $H_q$ predicted at 14 TeV [$k = 4.4$ (solid line); $k = 8$ (dash-dotted line)][26].

IPPI model. However, some difference at 14 TeV (see Fig. 7) and especially at 100 TeV between these models is predicted. To keep the same mean multiplicity in both models at the same energy, we have chosen different values of $m$ according to $\langle n\rangle = m\sum_{j=1}^{j_{\max}} w_j j$ and $w_j$ values shown in Table 1. Namely, their ratios are $m_{\rm IPPI}/m_{\rm lad} = 0.988$, 1.039, 1.145, 1.227 for $j_{\max} = 3$, 4, 5, 6, respectively. This shows that the maximum of the distribution moves to smaller multiplicities and its width becomes larger in the modified ladder model compared to the IPPI model with increasing energy.

Surely, one should not overestimate the success of the IPPI model in its present initial state. It has been applied just to multiplicity distributions. For more detailed properties, say, rapidity distributions, one would need a model for the corresponding features of the one-pair process. Moreover, the screening effect (often described by the triple Pomeron vertex) will probably become more important at higher energies. All these features are somehow implemented in the well-known Monte Carlo programs PYTHIA [41], HERWIG [42], and DPM–QGSM [9, 11]. However, for the latter one, for example, the multiplicity distribution for a single ladder is given by the Poisson distribution of emission centers (resonances) convoluted with their decay properties, and probabilities $w_j$ contain several adjustable parameters. It differs from the IPPI model. The latter approach proposes a more economical way with a smaller number of such parameters. What concerns the further development of the event generator codes, it is tempting to incorporate there the above approach with an NBD of particles created by a single parton pair and confront the results with a wider set of experimental data. We intend to do this later to learn how it influences other characteristics.

## 6. IS HADRONIC PRODUCTION SIMILAR IN VARIOUS PROCESSES?

This question was first raised by the statement of [43] that the average multiplicities in $e^+e^-$ and $pp(p\bar{p})$ processes increase with energy in a similar way. Recently, the PHOBOS Collaboration [44] even claimed that the energy behavior of mean multiplicities in all processes is similar. Therefore, it has been concluded that the dynamics of all hadronic processes is the same. Besides our general belief in QCD, we cannot claim that other characteristics of multiple production processes initiated by different partners coincide. To answer the above question, we compare characteristics of multiplicity distributions for processes initiated by different partners.

**Average multiplicities.** Total yields of charged particles in high-energy $e^+e^-$, $p\bar{p}$, $pp$, and central $AA$ collisions become similar if special rescaling is done. The average charged-particle multiplicity in $pp/p\bar{p}$ is similar [43] to that for $e^+e^-$ collisions if the effective energy $s_{\rm eff}$ is inserted in place of $s$, where $\sqrt{s_{\rm eff}}$ is the $pp/p\bar{p}$ center-of-mass energy minus the energy of the leading particles. In practice, $\sqrt{s_{\rm eff}} = \sqrt{s}/2$ is chosen. This corresponds to the horizontal shift of empty squares to the diamond positions in Fig. 9, borrowed from [44]. Then the diamonds lie very close to the dashed curve, which shows QCD predictions for multiplicities in $e^+e^-$ collisions according to (17). For central nucleus–nucleus collisions, the particle yields have been scaled by the number of participating nucleon pairs $N_{\rm part}/2$. Then the energy dependence of mean multiplicities is very similar for all these processes at energies exceeding 10 GeV up to 1 TeV, and even at lower energies for $e^+e^-$ compared with $pp/p\bar{p}$. This is well demonstrated in Fig. 9.

However, the situation changes at higher energies. In Table 2, we show experimentally measured mean multiplicities at Tevatron energies (first row). They are compared with the results of the IPPI model in the second row, where the IPPI predictions at 14 TeV (LHC) and 100 TeV are also shown. According to the above hypothesis, these values should coincide with QCD predictions (17) at a factor of 2 smaller energy. The latter ones are presented in the third row. The difference between them and experimentally measured values of the first row is demonstrated in the fourth row for Tevatron energies, while at higher energies the difference of QCD and IPPI predictions is shown.

It is seen that both experimental and theoretical values of average multiplicity coincide pretty well in $p\bar{p}$ and $e^+e^-$ up to 1 TeV. However, already at Tevatron energy 1.8 TeV, the rescaled $p\bar{p}$ multiplicity is lower by 3.4 charged particles. This difference between rescaled predictions of the IPPI model for $p\bar{p}$

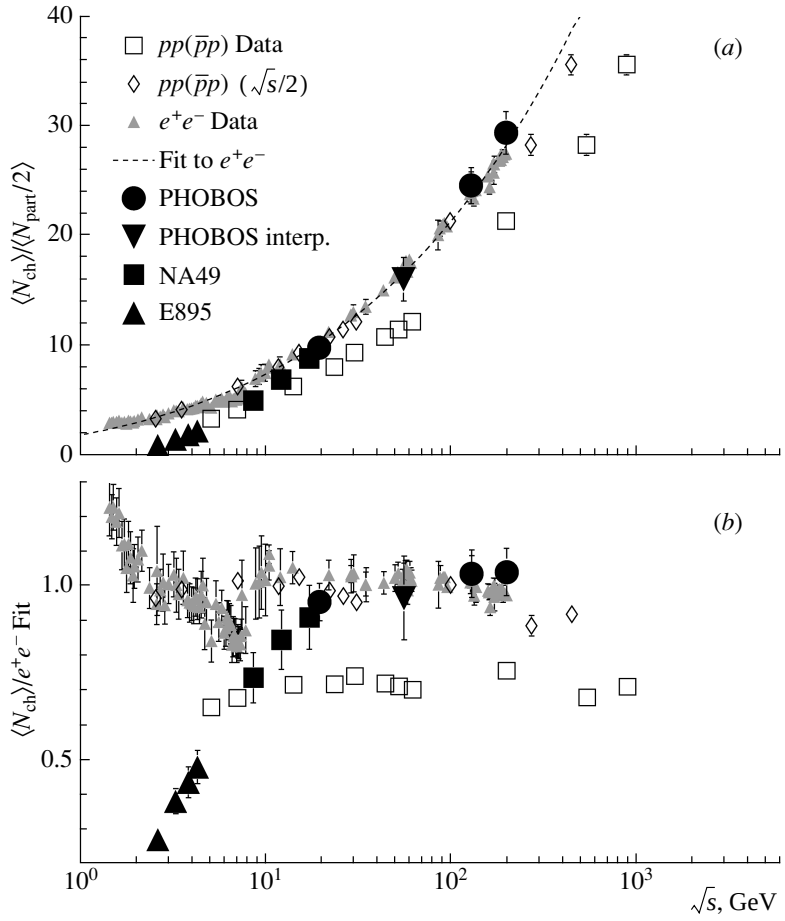


**Fig. 9.** The energy dependence of average multiplicities for various processes ($a$) and their ratio to QCD prediction for $e^+e^-$ collisions ($b$) [44].

and QCD for $e^+e^-$ becomes extremely large at LHC (26.6), and even more so at higher energies.

With these observations, one tends to claim (if at all!) the approximate quantitative similarity of all processes up to 1 TeV and "a universal mechanism of particle production in strongly interacting systems controlled mainly by the amount of energy available for particle production" [44] only at energies below the highest Tevatron values. The situation becomes even worse if we compare other features of multiplicity distributions.

***$H_q$* moments.** First, we have calculated [45] $H_q$ moments for experimental multiplicity distributions in various high-energy processes. They are shown in Fig. 10. These moments weakly depend on energy in the energy regions available to present experiments. Their oscillating shape is typical for all processes. However, one notices that the amplitudes of $H_q$ oscillations in Fig. 10 differ in various reactions. They are larger for processes with participants possessing more complicated internal structure. For example, the amplitudes in $e^+e^-$ are about two orders of magnitude smaller than those in $pp$. Both QCD applied to $e^+e^-$ [2, 31] and models of $pp/p\bar{p}$ [26] can fit these observations.

There is, however, one definite QCD prediction, which allows one to ask the question whether QCD and, e.g., the IPPI model are compatible, in principle. This is the asymptotic behavior of $H_q$ moments in QCD. They should behave [18] as $H_q^{\rm as} = 1/q^2$. One can also determine the asymptotics of $H_q$ moments in the IPPI model and compare both approaches [46]. These values are noticeably larger than QCD predictions of $1/q^2$. Thus, QCD and the IPPI model have different asymptotics. It is an open question whether other asymptotic relations for probabilities of multiparton interactions different from those adopted in the IPPI model can be found which would lead to the same behavior of $H_q$ moments in $p\bar{p}$ and $e^+e^-$ collisions. Only then can one hope to declare an analogy between these processes.

**Fractal properties.** The particle density within the phase space in individual events is much more structured and irregular than in the sample averaged distribution. However, even for the latter ones, fluctuations depend on the phase space of a sample. The

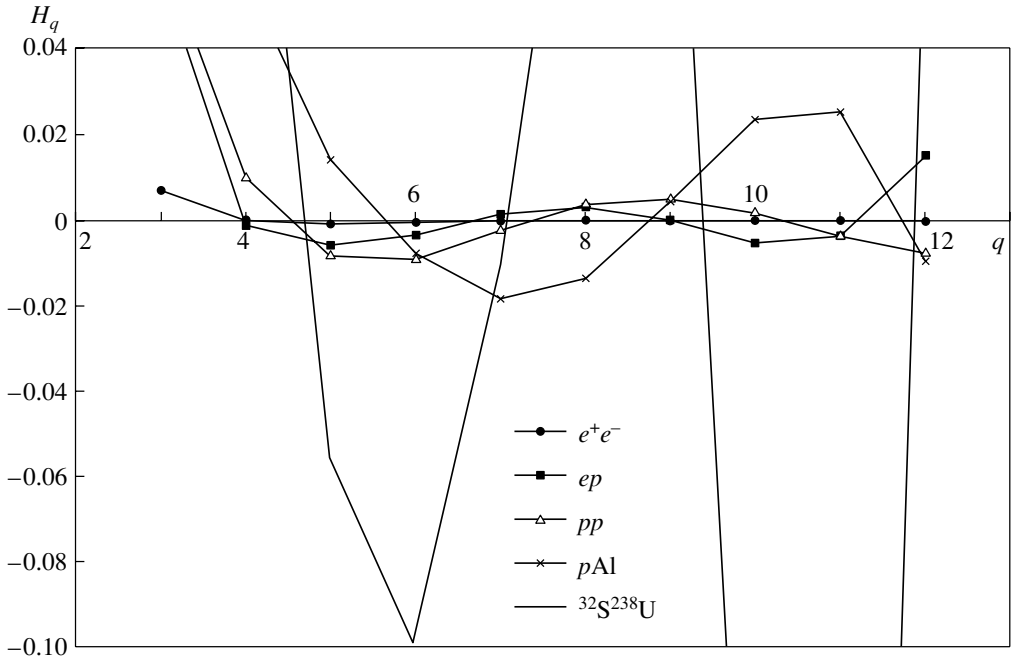


**Fig. 10.** Oscillations of $H_q$ moments for various processes [45].

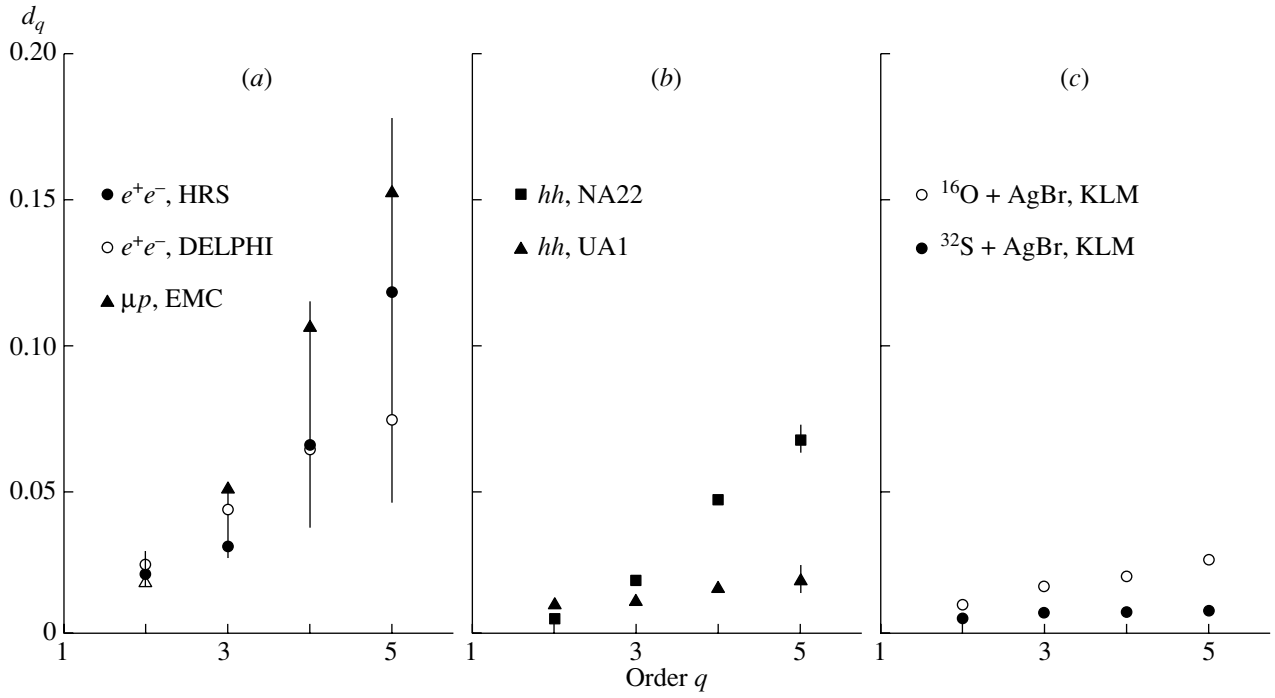


**Fig. 11.** Anomalous fractal dimensions for various processes [28].

smaller the phase space, the larger the fluctuations. This can be described by the behavior of normalized factorial moments $F_q = \mathcal{F}_q/\langle n\rangle^q$ as functions of the amount of phase space available. In the one-dimensional case of the rapidity distributions within the interval $\delta y$, the powerlike behavior $F_q(\delta y) \propto (\delta y)^{-\phi(q)}$ for $\delta y \to 0$ and $\phi(q) > 0$ would correspond to an intermittent phenomenon well known from turbulence. This also shows that the created particles are distributed in the phase space in a fractal manner. The anomalous fractal dimension $d_q$ is connected with the intermittency index $\phi(q)$ by the relation $\phi(q) = (q-1)(d-d_q)$, where $d$ is the ordinary dimension

**Table 2.** The mean multiplicities at Tevatron and higher energies

| $\sqrt{s}$, GeV | 300 | 546 | $10^3$ | $1.8 \times 10^3$ | $1.4 \times 10^4$ | $10^5$ |
|---|---|---|---|---|---|---|
| $n_{p\bar{p},\mathrm{exp}}$ | 25.4 | 30.5 | 39.5 | 45.8 | – | – |
| $n_{p\bar{p},\mathrm{theor}}$ | 24.1 | 30.0 | 39.4 | 45.7 | 71.6 | 97.0 |
| $n_{e^+e^-,\mathrm{theor}}$ | 25.3 | 31.8 | 39.9 | 49.2 | 98.2 | 180 |
| $n_{e^+e^-} - n_{p\bar{p}}$ | −0.1 | 1.3 | 0.4 | 3.4 | 26.6 | 83 |

of a sample studied ($d = 1$ for the one-dimensional rapidity plot). It has been calculated in QCD [47–49] only in the lowest order perturbative approximations. The qualitative features of the behavior of factorial moments as functions of the bin size observed in experiment are well reproduced by QCD.

The anomalous dimensions as functions of the order $q$ derived from experimental data are shown in Fig. 11 for various collisions [28]. They are quite large for $e^+e^-$ and become smaller for $hh$ and even more so for $AA$ collisions. This shows that the more structured the colliding partners, the more strongly smoothed the density fluctuations in the phase space. In some way, this observation correlates with the enlarged amplitudes of $H_q$ oscillations mentioned above.

In any case, this is another indication that it is premature to claim the similarity of $e^+e^-$- and hadron-initiated processes. We have compared just some features of multiplicity distributions. There are many more characteristics of the processes that can be compared, but this is beyond the scope of the present paper.

## 7. CONCLUSIONS

We have briefly described the theoretical approaches to collisions of high-energy particles. They seem to be quite different for various processes. No direct similarity in multiplicity distributions has been observed. According to experimental data, the energy dependence of average multiplicities in different collisions can be similar in the energy region above 10 GeV if some rescaling procedures are used. However, this is just the first moment of multiplicity distributions. To speak about their similarity, one should compare other moments. Again, the qualitative features of moment oscillations are somewhat similar, but quantitatively they differ. The same can be said about the fractal properties of particle densities within the phase space. No deep reasoning for the corresponding rescaling has been promoted. Thus, besides our general belief that the QCD Lagrangian is at the origin of all these processes, we cannot present any serious arguments in favor of similar schemes applicable to the dynamics of the processes. Moreover, QCD and the considered models of hadron interactions predict different asymptotics for some characteristics of $e^+e^-$ and $pp(p\bar{p})$. We have considered here just multiplicity distributions. Other inclusive and exclusive characteristics also seem to be different in these processes.

In conclusion, multiplicity distributions in various high-energy processes possess some common qualitative features but differ quantitatively. Theoretical approaches to their description have very little in common. Thus, it is premature to claim their common dynamical origin independently of our belief in QCD as a theory of strong interactions.

## ACKNOWLEDGMENTS

I am grateful to V.A. Nechitailo for collaboration on many aspects of the problems discussed. Special thanks are to the PHOBOS Collaboration for providing Fig. 9. This work has been supported in part by the Russian Foundation for Basic Research, project nos. 02-02-16779, 04-02-16445-a, and NSH-1936.2003.2.

## *APPENDIX*

The higher the rank of the moment, the higher the multiplicities that contribute to it. Therefore, the high-rank moments are extremely sensitive to the high-multiplicity tail of the distribution. At the same time, the energy-momentum conservation and experimental statistics limitations are important at the tail of the distribution. Therefore, the question about the limits of applicability of the whole approach is quite reasonable.

Let us estimate the range of validity of considering large-$q$ values of $H_q$ moments imposed by some cutoffs (see also [35]). It is well known that experimental cutoffs of multiplicity distributions due to the limited statistics of an experiment can influence the behavior of $H_q$ moments. Consequently, they impose some limits on $q$ values allowed to be considered when a comparison is done. Higher rank moments can be evaluated if larger multiplicities have been measured. To estimate the admissible range of $q$, we use the results obtained in QCD. Characteristic multiplicities

that determine the moment of the rank $q$ can be found. By inverting this relation, one can write the asymptotic expression for the characteristic range of $q$ [20]. This provides the bound $q_{\max} \approx Cn_{\max}/\langle n\rangle$, where $C \approx 2.5527$. However, it underestimates the factorial moments. Moreover, the first moment is not properly normalized (it becomes equal to $2/C$ instead of 1). Strongly overestimated values (however, with a correct normalization of the first moment) are obtained if $C$ is replaced by 2. Hence, one can say that the limiting values of $q$ are given by the inequalities

$$2n_{\max}/\langle n\rangle < q_{\max} \leq Cn_{\max}/\langle n\rangle.$$

The ratio $n_{\max}/\langle n\rangle$ measured by the E735 Collaboration at 1.8 TeV is about 5. Thus, $q_{\max}$ should be in the interval between 10 and 13. The approximate constancy of $m$ and proper fits of $H_q$ demonstrated above persist to even higher ranks.

## REFERENCES


1. V. A. Khoze and W. Ochs, Int. J. Mod. Phys. A **12**, 2949 (1997).
2. I. M. Dremin and J. W. Gary, Phys. Rep. **349**, 301 (2001).
3. A. Giovannini, Nuovo Cimento A **10**, 713 (1972); **15**, 543 (1973).
4. P. V. Landshoff and J. C. Polkinghorne, Phys. Rev. D **18**, 3344 (1978).
5. B. Humpert, Phys. Lett. B **131B**, 461 (1983).
6. K. Goulianos, Phys. Lett. B **193**, 151 (1987).
7. E735 Collab. (T. Alexopoulos *et al.*), Phys. Lett. B **435**, 453 (1998).
8. W. D. Walker, Phys. Rev. D **69**, 034007 (2004).
9. A. Capella, U. Sukhatme, C. I. Tan, and J. Tran Thanh Van, Phys. Lett. B **81B**, 68 (1979).
10. A. B. Kaidalov, Phys. Lett. B **116B**, 459 (1982).
11. A. B. Kaidalov and K. A. Ter-Martirosyan, Phys. Lett. B **117B**, 247 (1982); Yad. Fiz. **40**, 211 (1984) [Sov. J. Nucl. Phys. **40**, 135 (1984)].
12. S. G. Matinyan and W. D. Walker, Phys. Rev. D **59**, 034022 (1999).
13. T. K. Gaisser and F. Halzen, Phys. Rev. Lett. **54**, 1754 (1985).
14. A. Giovannini and R. Ugoccioni, Phys. Rev. D **59**, 094020 (1999).
15. C. Bourelly *et al.*, in *Proceedings of the VI Blois Workshop on Frontiers in Strong Interactions,* Ed. by J. Tran Thanh Van (Frontières, Gif-sur-Yvette, 1995), p. 15.
16. L. McLerran, Rev. Mod. Phys. **58**, 1021 (1986).
17. E. Shuryak, hep-ph/0312227.
18. I. M. Dremin, Phys. Lett. B **313**, 209 (1993).
19. Yu. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troyan, *Basics of Perturbative QCD,* Ed. by J. Tran Thanh Van (Frontières, Gif-sur-Yvette, 1991).
20. Yu. L. Dokshitzer, Phys. Lett. B **305**, 295 (1993).
21. I. M. Dremin and D. S. Chernavsky, Zh. Éksp. Teor. Fiz. **38**, 229 (1960) [Sov. Phys. JETP **11**, 167 (1960)].
22. D. Amati, S. Fubini, S. Stanghellini, and A. Tonin, Nuovo Cimento **26**, 896 (1962).
23. V. N. Akimov, D. S. Chernavsky, I. M. Dremin, and I. I. Royzen, Nucl. Phys. B **14**, 285 (1969).
24. I. M. Dremin and A. M. Dunaevsky, Phys. Rep. **18**, 159 (1975).
25. A. Giovannini and L. Van Hove, Z. Phys. C **30**, 391 (1986).
26. I. M. Dremin and V. A. Nechitailo, hep-ph/0402286.
27. B. Andersson, *The Lund Model*, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. (1997), Vol. 7.
28. E. A. De Wolf, I. M. Dremin, and W. Kittel, Phys. Rep. **270**, 1 (1996).
29. I. M. Dremin and V. A. Nechitailo, Mod. Phys. Lett. A **9**, 1471 (1994); Pis'ma Zh. Éksp. Teor. Fiz. **58**, 945 (1993) [JETP Lett. **58**, 881 (1993)].
30. I. M. Dremin and R. C. Hwa, Phys. Rev. D **49**, 5805 (1994); Phys. Lett. B **324**, 477 (1994).
31. M. A. Buican, C. Förster, and W. Ochs, Eur. Phys. J. C **31**, 57 (2003).
32. I. M. Dremin, V. Arena, G. Boca, *et al.*, Phys. Lett. B **336**, 119 (1994).
33. SLD Collab. (K. Abe *et al.*), Phys. Lett. B **371**, 149 (1996).
34. L3 Collab. (P. Achard *et al.*), Phys. Lett. B **577**, 109 (2003).
35. I. M. Dremin and V. A. Nechitailo, Eur. Phys. J. C **30**, d01, 004 (2003).
36. A. Giovannini and R. Ugoccioni, Phys. Rev. D **68**, 034009 (2003).
37. E. K. G. Sarkisyan, Phys. Lett. B **477**, 1 (2000).
38. UA5 Collab. (G. Arnison *et al.*), Phys. Lett. B **118B**, 167 (1982).
39. E735 Collab. (F. Turkot *et al.*), Nucl. Phys. A **525**, 165 (1991).
40. S. Hegyi, Phys. Lett. B **387**, 642 (1996); **418**, 186 (1998).
41. M. Bertini, L. Lönnblad, and T. Sjöstrand, Comput. Phys. Commun. **134**, 365 (2001); see also http://www.thep.lu.se/~torbjorn/Pythia.html
42. HERWIG 6.5, G. Corcella, I. G. Knowles, G. Marchesini, *et al.*, J. High Energy Phys. **0101**, 010 (2001); see also http://hepwww.rl.ac.uk/theory/seymour/herwig/
43. M. Basile *et al.*, Phys. Lett. B **92B**, 367 (1980); **95**, 311 (1980).
44. PHOBOS Collab. (B. B. Back *et al.*), nucl-ex/0301017.
45. I. M. Dremin, V. A. Nechitailo, M. Biyajima, and N. Suzuki, Phys. Lett. B **403**, 149 (1997).
46. I. M. Dremin, hep-ph/0404092; in *Proceedings of the XXXIX Rencontres de Moriond, 2004*, Ed. by J. Tran Thanh Van (in press).
47. W. Ochs and J. Wosiek, Phys. Lett. B **289**, 159 (1992); **304**, 144 (1993).
48. Yu. L. Dokshitzer and I. M. Dremin, Nucl. Phys. B **402**, 139 (1993).
49. Ph. Brax, J. L. Meunier, and R. Peschanski, Z. Phys. C **62**, 649 (1994).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 771–778. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 804–811.*


# Relative Yield of Charged and Neutral Heavy Meson Pairs in $e^+e^-$ Annihilation near Threshold*

**M. B. Voloshin**

*William I. Fine Theoretical Physics Institute, University of Minnesota, Minneapolis, MN, and Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received May 13, 2004

**Abstract**—The subject of the charged-to-neutral yield ratio for $B\bar{B}$ and $D\bar{D}$ pairs near their respective thresholds in $e^+e^-$ annihilation is revisited. As previously argued for the $B$ mesons, this ratio should exhibit a substantial variation across the $\Upsilon(4S)$ resonance due to interference of the resonance scattering phase with the Coulomb interaction between the charged mesons. A simple alternative derivation of the expression describing this effect is presented here, and the analysis is extended to include the $D$-meson production in the region of the $\psi(3770)$ resonance. The available data on kaon production at the $\phi(1020)$ resonance are also discussed in connection with the expected variation of the charged-to-neutral yield ratio.


## 1. INTRODUCTION

The near-threshold resonances $\phi(1020)$, $\psi(3770)$, and $\Upsilon(4S)$, decaying, respectively, to pairs of pseudoscalar mesons $K\bar{K}$, $D\bar{D}$, and $B\bar{B}$, are well-known "factories" for production of these mesons in $e^+e^-$ annihilation. In a number of experimental studies, it is important to know the relative yield of the pairs of charged and neutral mesons in the corresponding resonance region, i.e., the ratio

$$R^{c/n} = \frac{\sigma(e^+e^- \to P^+P^-)}{\sigma(e^+e^- \to P^0\bar{P}^0)}, \qquad (1)$$

where $P$ stands for the pseudoscalar meson, i.e., $K$, $D$, or $B$. The knowledge of this ratio is of particular importance for the studies of the $B$ mesons at the $B$ factories, and dedicated measurements have been done at the $\Upsilon(4S)$ resonance by CLEO [1, 2], BABAR [3, 4], and Belle [5]. The central values of the ratio $R^{c/n}$ found in these measurements typically range from 1.01 to 1.10 with the latest result [4] being $1.006 \pm 0.036(\text{stat.}) \pm 0.031(\text{syst.})$. The same ratio for the production of $D$-meson pairs is likely to be studied in detail at the $\psi(3770)$ resonance by the imminent CLEO-c experiment [6]. For the kaon pair production, the most detailed available scan of production of charged and neutral mesons in the $\phi$-resonance region has been done by the SND Collaboration [7].

The theoretical treatment of the ratio $R^{c/n}$ at these three thresholds near the corresponding resonances bears considerable similarity, although specific factors contributing to the ratio in each case are slightly different. In a way, the simplest case is presented by the threshold $B$-pair production, where the deviation of $R^{c/n}$ from one is essentially entirely determined by the Coulomb interaction between the charged $B$ mesons, since the mass difference between the $B$ mesons is very small: $m_{B^0} - m_{B^+} = 0.33 \pm 0.28$ MeV [8]. For the $D\bar{D}$ pair production at the $\psi(3770)$ peak, the difference in the $P$-wave kinematical factor $p^3$ due to the substantial mass difference between the charged and neutral $D$ mesons is the main source of deviation of $R^{c/n}$ from one, and the Coulomb interaction effect is somewhat smaller, but is still of a measurable $O(10\%)$ magnitude. Finally, in the $\phi(1020)$-resonance region in addition to the mass difference and the Coulomb interaction effects, the amplitudes of production of $K^+K^-$ and $K_LK_S$ also differ due to a nonresonant isovector, $I = 1$, contribution coming from the electromagnetic current of the $u$ and $d$ quarks. Another technical difference between the heavy mesons and the kaons is that the former can safely be considered nonrelativistic at energies in the region of the corresponding resonance (the velocities of the mesons at the corresponding resonances are $v(B)/c \approx 0.06$ and $v(D^+)/c \approx 0.13$), while the velocity of the charged kaons at the $\phi$ resonance is $v(K^+)/c \approx 0.25$, and the $O(v^2/c^2)$ relativistic effects can show up at some level of intended accuracy in $R^{c/n}$.

The kinematical effect of the mass difference, and of the nonresonant isovector amplitude for the case

*This article was submitted by the author in English.

1063-7788/05/6805-0771$26.00 

of kaon production, leads to factors in the $R^{c/n}$, which are rather smoothly varying within the width of the corresponding resonance. The behavior of the Coulomb interaction effect is, however, quite different. Namely, as recently pointed out [9], with a proper treatment of the strong resonant interaction between the mesons, there is a phase interference between the Coulomb interaction and the resonance (Breit–Wigner) scattering phase. Since the latter phase changes by $\pi$ across the resonance, the effect of the Coulomb interaction in $R^{c/n}$ should exhibit a substantial variation with energy within the width of the resonance. Thus, in particular, a comparison between different measurements of the charged-to-neutral ratio at the $\Upsilon(4S)$ peak can only be meaningful with proper account of the differences in the specific experimental setups, such as the beam energy spread, stability, and calibration. Needless to say, the best approach would be a dedicated scan of the $\Upsilon(4S)$ peak. The details of this variation depend [9] on the specifics of the transition from the strong interaction region at short distances to the long-range Coulomb interaction and also on such details as the nonresonant scattering phase for the mesons. A dedicated scan of the charged-to-neutral ratio thus could possibly provide information on rather fine properties of strong interaction between the heavy mesons, which are not likely to be accessible by other means. It may well be that a more experimentally feasible object for such study is provided by the $\psi(3770)$ resonance, and possibly can be explored in detail in the forthcoming CLEO-c experiment. The main motivation of the present paper is to extend the analysis of [9] to the case of the $D\bar{D}$ production in the region of this resonance, and to estimate the magnitude of the expected effects, to the extent that it can be done within the present (rather approximate at best) understanding of the details of the strong interaction between heavy mesons.

The typical amplitude of the varying part of the ratio $R^{c/n}$ due to the Coulomb interaction is set by the Coulomb parameter $\alpha/v$ for the charged mesons, which tells us that, for the $D$ mesons at $\psi(3770)$, the amplitude of the variation should be approximately two times smaller than for the $B$ mesons at $\Upsilon(4S)$ (and by another factor of two smaller for the kaons at $\phi$). In either case, the Coulomb parameter is sufficiently small to justify limiting the consideration to only the linear term in the Coulomb interaction treated as a perturbation. A simple derivation of this Coulomb interaction correction to production of a charged meson pair with the exact treatment of the strong interaction scattering is presented in Section 2. This derivation is slightly different from that given in [9] and, hopefully, more transparently leads to the expression for the $O(\alpha/v)$ term found there. In Section 3 are presented specific estimates of the behavior of the ratio $R^{c/n}$ across the $\psi(3770)$ resonance, which take into account both the mass difference between the $D$ mesons and the effect of the Coulomb interaction.

It can be noticed that the very fact of the existence of the variation of $R^{c/n}$ within the resonance width is essentially model independent and is a direct consequence of a somewhat standard application of the quantum-mechanical scattering theory. Nevertheless, there has been some skepticism expressed regarding the existence of the predicted [9] rapid variation of $R^{c/n}$ at the resonance, most notably in the introductory part of the recent experimental paper [4]. Although the measurement reported in [4] has been done at just one point in the energy at the peak of the $\Upsilon(4S)$ resonance, the paper implicitly puts the prediction of the variation in doubt by arguing that "such rapid variation in the charged-to-neutral ratio has not been observed in scans across the $\phi(1020)$ resonance" (with a reference to the scans by the SND Collaboration [7]). Thus, it appears to be useful to point out the difference between the predictions for the $B$ production at $\Upsilon(4S)$ and the kaon production at $\phi(1020)$ following from the discussed approach. The main difference is that the amplitude of the variation at the $\phi$ is expected to be suppressed by a factor of approximately 0.25 due to about four times larger velocity of the kaons at the relevant energy, and additional differences in the details arise through different parameters (e.g., widths) of the $\phi$ and $\Upsilon(4S)$ resonances. A comparison of a "representative" theoretical curve with the behavior of $R^{c/n}$ across the $\phi$ resonance extracted from the scan data [7] is presented in Section 4. Given the existing theoretical and experimental uncertainties, a meaningful comparison can only be done at a mostly qualitative level, and it is left entirely up to the reader to assess from the presented comparison whether the scan data indeed exclude a variation with expected amplitude of the charged-to-neutral ratio across the $\phi$ resonance.

## 2. COULOMB INTERACTION CORRECTION TO PRODUCTION OF CHARGED MESONS

If the mesons were point particles devoid of strong interaction and produced by a point source, the Coulomb interaction between charged mesons would enhance their production cross section by the textbook factor

$$F_c = 1 + \frac{\pi\alpha}{2v} + O(\alpha^2/v^2), \qquad (2)$$

where $v$ is the velocity of each of the mesons in the center-of-mass frame, which was the early prediction [10] for the ratio $R^{c/n}$ at the $\Upsilon(4S)$ resonance. It

was, however, then realized that the effects of the meson electromagnetic form factor [11] and of the form factor in the decay vertex $\Upsilon(4S) \to B\bar{B}$ [12, 13] tend to reduce the enhancement. In all these approaches, the Coulomb rescattering of the produced mesons is considered, assuming that the propagation of the meson pair between the production vertex and the rescattering as well as after the Coulomb rescattering is described by a free propagator, and thus ignoring the fact that the strong interaction introduces the scattering phase factors $\exp(i\delta)$ in these propagators. This would be a reasonable approximation if the strong phase $\delta$ were small. In the region of a strong resonance, however, this assumption is definitely invalid, since the phase changes by $\pi$ within the width of the resonance. The calculation described in [9] allows one to completely take into account the strong scattering phase as well as the Coulomb interaction between the mesons. In the first order in the Coulomb rescattering, the expression [9] for the Coulomb correction factor has a simple form

$$F_c = 1 + \frac{1}{v}\,\mathrm{Im}\left[e^{2i\delta}\int\limits_a^\infty e^{2ipr}\left(1+\frac{i}{pr}\right)^2 V(r)dr\right], \tag{3}$$

where $p$ is the meson momentum and $V(r)$ is the potential for the rescattering interaction. Clearly, up to a possible form factor, $V(r) = -\alpha/r$ for the Coulomb interaction. The short-distance cutoff parameter $a$ in Eq. (3) accounts for the fact that, in the region of a strong interaction at short distances, the mesons spatially overlap and in fact the system is a mixture of heavy and light quarks and gluons rather than a state of two individual mesons. Therefore, at such distances, there is no separation between the states with charged and neutral mesons and thus any difference in their Coulomb interaction disappears. If one introduces form factors in the Coulomb interaction in order to account for the structure of the mesons, the interaction potential $V(r)$ would automatically vanish at short distances and the ultraviolet cutoff in the integral in Eq. (3) would be provided by the form factors. If, as in the earlier analyses, one sets the strong scattering phase $\delta$ to zero, the expression in Eq. (3) is finite even if an unmodified Coulomb potential is taken down to $a = 0$, in the limit of which it reproduces Eq. (2). However, at any finite $\delta$, there is an essential dependence on the ultraviolet cutoff parameter $a$ which cannot be removed.

In this section, we present a derivation of Eq. (3) slightly different from that in [9] by considering the following problem. Two mesons, each with mass $m$ interact strongly at distances $r < a$. The strong interaction in the $P$ wave at the total kinetic energy $E$ produces the scattering phase $\delta$. At longer distances, $r > a$, the mesons interact via the potential $V(r)$. The meson pair is produced by a source localized at distances shorter than $a$ (this obviously corresponds to the production in the $e^+e^-$ annihilation). The problem is to find the correction of the first order in $V$ to the production rate.

In order to solve the formulated problem, we consider the radial part of the wave function of a stationary $P$-wave scattering state, written in the form $R(r) = \chi(r)/r$. According to the general scattering theory [14], the asymptotic form of $\chi(r)$ at $r \to \infty$ in the absence of the long-range potential $V(r)$ is

$$\chi(r) = 2\cos(pr + \delta) = e^{i\delta}e^{ipr} + e^{-i\delta}e^{-ipr}. \tag{4}$$

Furthermore, in the absence of the potential $V$, the function $\chi(r)$ satisfies the Schrödinger equation for free motion in the $P$ wave,

$$\chi''(r) + \left(p^2 - \frac{2}{r^2}\right)\chi(r) = 0, \tag{5}$$

at all distances $r$ outside the region of strong interaction, i.e., at $r > a$. Thus, the solution valid at all such distances and having the asymptotic form (4) can be written explicitly,

$$\chi(r) = e^{i\delta}f(pr) + e^{-i\delta}f^*(pr), \tag{6}$$

in terms of the function

$$f(pr) = \left(1 + \frac{i}{pr}\right)e^{ipr} \tag{7}$$

and of its complex conjugate $f^*$. At distances $r < a$, the dynamics is unknown to the extent that continuing the description of the system in terms of the meson pair wave function in the region of strong interaction does not make much sense, since most likely such description should involve entirely different degrees of freedom. Nevertheless, however complicated the "inner" dynamics of the system may be, the wave function in Eq. (5) at $r = a$ provides the boundary condition for the "inner" problem. In particular, it determines the normalization of the state for the inner problem. Thus, the rate of production of the system, and consequently of the meson pair, by a source localized at $r \ll a$ is proportional to $|\chi(a)|^2$.

When the long-range potential $V(r)$ is turned on and considered as a perturbation, the wave function at $r > a$ changes to $\chi(r) + \delta\chi(r)$. Once the corrected solution is normalized at $r \to \infty$ in the standard way, i.e., as the asymptotic wave in Eq. (4), the normalization at $r = a$ generally changes and hence also changes the production rate. In the first order in $V$, the correction factor for the production rate is obviously given by

$$F_c = 1 + 2\frac{\delta\chi(a)}{\chi(a)}. \tag{8}$$

The perturbation $\delta\chi$ of the scattering state wave function can be found in the standard way by writing it as a sum of outgoing and incoming waves, $\delta\chi(r) = \delta\chi_+(r) + \delta\chi_-(r)$, similarly to Eq. (6), and where $\delta\chi_+(r)$ contains at $r \to \infty$ only an outgoing wave ($\exp(ipr)$), while $\delta\chi_- = \delta\chi_+^*$ contains at asymptotic $r$ only an incoming wave. The function $\delta\chi_+(r)$ is a perturbation of the outgoing wave part of the function $\chi(r)$ in Eq. (6) and is thus determined by the equation

$$\delta\chi_+''(r) + \left(p^2 - \frac{2}{r^2}\right)\delta\chi_+(r) = mV(r)e^{i\delta}f(pr). \tag{9}$$

The solution to this equation is found by the standard method using the Green's function $G_+(r,r')$ satisfying the equation

$$\left(\frac{\partial^2}{\partial r^2} + p^2 - \frac{2}{r^2}\right)G_+(r,r') = \delta(r-r') \tag{10}$$

and the condition that $G_+(r,r')$ contains only an outgoing wave when either of its arguments goes to infinity. The Green's function is constructed from two solutions of the homogeneous equation, i.e., from the functions $f(pr)$ and $f^*(pr)$, as

$$G_+(r,r') = \frac{1}{2ip}[f(pr)f^*(pr')\theta(r-r') + f(pr')f^*(pr)\theta(r'-r)], \tag{11}$$

where $\theta$ is the standard unit step function. The solution to Eq. (9) is then found as

$$\delta\chi_+(r) = me^{i\delta}\int_a^\infty G_+(r,r')V(r')f(pr')dr'. \tag{12}$$

It can be noted that at no point in this consideration is knowledge of the dynamics at $r < a$ required. In particular, the integral in Eq. (12) runs from the lower limit at $a$, since by our assumptions the perturbation potential has support only at $r > a$.

It is further important that adding the found perturbation $\delta\chi$ to the wave function does not change the normalization of the wave function at $r \to \infty$. Indeed, in this limit one has $r > r'$ in the Green's function in Eq. (12), so that the asymptotic behavior of $\delta\chi_+(r)$ should be derived from the expression

$$\delta\chi_+(r)\,|_{r\to\infty} = -\frac{i}{2v}e^{i\delta}f(pr)\int_a^\infty V(r')|f(pr')|^2dr', \tag{13}$$

which gives the complex phase of $\delta\chi_+(r)$ at the asymptotic distances manifestly orthogonal to that of the outgoing-wave part of $\chi(r)$, since the integral is explicitly real.[1)] In other words, at $r \to \infty$, the correction changes only the scattering phase by adding the "Coulomb" phase to $\delta$.

At $r = a$, the perturbation $\delta\chi_+(a)$ is found from the same expression (12). In this case, one has $r < r'$ in the entire integration region, so that

$$\delta\chi_+(a) = -\frac{i}{2v}e^{i\delta}f^*(pa)\int_a^\infty V(r')\left[f(pr')\right]^2dr'. \tag{14}$$

Using this expression, one readily finds the change in the normalization of the wave function at $r = a$ due to the potential $V$ and thus arrives at expression (3).

As is clear from the presented derivation, the result in Eq. (3) is valid at arbitrary $P$-wave scattering phase $\delta$ and also generally does not assume any condition for the value of the product $pa$. At energy $E$ near a strong resonance dominating the dynamics of the meson pair, the scattering phase is determined by the Breit–Wigner formula

$$\delta = \delta_{\rm BW} + \delta_1, \quad e^{2i\delta_{\rm BW}} = \frac{\Delta - i\gamma}{\Delta + i\gamma}, \tag{15}$$

where $\Delta = E - E_0$ is the distance in energy to the nominal position $E_0$ of the resonance, and $\gamma$ is generally a function of energy such that its value at $E_0$ is related to the nominal width $\Gamma$ of the resonance as $\gamma(E_0) = \Gamma/2$. Finally, $\delta_1$ in Eq. (15) is the nonresonant $P$-wave scattering phase.

The properties of the discussed resonances, $\Upsilon(4S)$ and $\psi(3770)$, are determined by the two meson-pair channels: $P^+P^-$ and $P^0\bar{P}^0$. The analytical properties of the scattering amplitude in the $P$ wave require that the contribution of each channel to $\gamma$ and $\delta_1$ start at the corresponding meson-pair threshold as cubic in the corresponding momentum. Thus, at an energy close to both thresholds, one can parametrize these quantities as

$$\gamma(E) = c_1(p_+^3 + p_0^3), \quad \delta_1(E) = c_2(p_+^3 + p_0^3), \tag{16}$$

where $p_+$ and $p_0$ are the momenta of, respectively, the charged and neutral mesons at the energy $E$. Due to the very small mass difference between the $B$ mesons, the thresholds for the charged and neutral $B$ mesons practically coincide and so do the momenta $p_+$ and

[1)] A minor technical point is that formally the integral is logarithmically divergent at $r' \to \infty$ for the Coulomb potential, which might put into question the applicability of the condition $r > r'$ in the entire integration region. This is the standard infrared divergence of the Coulomb scattering phase, and it can be easily dealt with by temporarily introducing a small photon mass $\lambda$, so that the potential is cut off by the factor $\exp(-\lambda r)$ and the integral is convergent. In the final result, one can obviously take the limit $\lambda \to 0$.

$p_0$. However, for the $D\bar{D}$ pairs, the $D^+D^-$ threshold is 9.6 MeV higher than that for $D^0\bar{D}^0$, the difference of which is substantial at the energy of the $\psi(3770)$ resonance.

It should be noticed that neglecting the higher terms of expansion in the momenta for the parameters $\gamma$ and $\delta_1$ is justified only inasmuch as the momenta are small in the scale of the size of the strong interaction region $a_s$. In the calculation of the Coulomb interaction effect, it is essential that the distance parameter $a$ for the onset of the Coulomb interaction not be smaller than $a_s$: $a \geq a_s$. Otherwise, no restriction on the value of the product $pa$ is implied. In [9], these parameters were reasonably assumed to be approximately equal. However, it should be emphasized that, generally, the applicability of formula (3) for the Coulomb correction is separate from the applicability of the first terms of the threshold expansion in Eq. (16), and these two issues can be studied separately in experiments.[2)]

In conclusion of this section, we write the explicit formula for the correction described by Eq. (3) in the resonance region in terms of real quantities [9]:

$$F_c = 1 + \frac{\alpha}{v} \times \left[\frac{\Delta^2-\gamma^2}{\Delta^2+\gamma^2}\left(A\cos 2\delta_1 + B\sin 2\delta_1\right) - \frac{2\gamma\Delta}{\Delta^2+\gamma^2}\left(B\cos 2\delta_1 - A\sin 2\delta_1\right)\right]. \quad (17)$$

The coefficients $A$ and $B$ are found as the imaginary and the real parts of the integral in Eq. (3) with the Coulomb potential:

$$A = -\int\limits_{pa}^{\infty}\left[\left(1-\frac{1}{u^2}\right)\sin 2u + \frac{2\cos 2u}{u}\right]\frac{du}{u} \quad (18)$$
$$= \frac{\pi}{2} - \frac{\cos 2pa}{pa} + \frac{\sin 2pa}{2(pa)^2} - \mathrm{Si}(2pa),$$
$$B = \int\limits_{pa}^{\infty}\left[\frac{2\sin 2u}{u} - \left(1-\frac{1}{u^2}\right)\cos 2u\right]\frac{du}{u}$$
$$= \frac{\cos 2pa}{2(pa)^2} + \frac{\sin 2pa}{pa} - \mathrm{Ci}(2pa),$$

where

$$\mathrm{Si}(z) = \int\limits_0^z \sin t\frac{dt}{t} \quad \text{and} \quad \mathrm{Ci}(z) = -\int\limits_z^{\infty}\cos t\frac{dt}{t}.$$

[2)]In fact, an attempt at detecting terms higher than $p^3$ in the width parameter for the $\Upsilon(4S)$ resonance has been done by ARGUS Collaboration [15]. However, the deviation from the cubic behavior turned out to be too small within the experimental accuracy.

## 3. ESTIMATES OF THE EXPECTED VARIATION OF $R^{c/n}$ ACROSS THE $\psi(3770)$ RESONANCE

Here, we apply the formulas of the previous section, specifically Eqs. (17), (18), and (16), to an estimate of the scale of the variation in the ratio $R^{c/n}$ for the $D\bar{D}$ pair production at the $\psi(3770)$ resonance. In doing so, it should be clearly understood that, at present, one can only guess (in some reasonable range) the appropriate values of the scattering phase $\delta_1$ and the cutoff parameter $a$ for the Coulomb interaction (or, more generally, make a guess about an appropriate model for the onset at short distances of the Coulomb interaction between the charged $D$ mesons).

As different from the case of the $B$-meson production at the $\Upsilon(4S)$, in the charged-to-neutral ratio for the $D$-meson production at $\psi(3770)$, the Coulomb effect multiplies the ratio of the $p^3$ factors:

$$R^{c/n} = F_c\left(\frac{p_+}{p_0}\right)^3. \quad (19)$$

Also, according to the Particle Data Tables [8], the width of $\psi(3770)$ is somewhat larger than that of the $\Upsilon(4S)$: $\Gamma(\psi'') = 23.6 \pm 2.7$ MeV as opposed to $14 \pm 5$ MeV, which tends to smoothen the variation of $R^{c/n}$ near the $\psi(3770)$ peak.

A sample expected behavior of the ratio $R^{c/n}$ for the $D$ pair production near the $\psi(3770)$ resonance is shown in Fig. 1 for some representative values of the parameters $a$ and $\delta_1(E_0)$. For comparison, the same curves for the $B$ pairs in the vicinity of the $\Upsilon(4S)$ peak are also shown in a separate plot in Fig. 1. It should be noted in connection with this comparison that, generally, it is not expected that the phase $\delta_1$ is the same in these two cases, although the parameter $a$ viewed as characterizing the electric charge structure in a heavy meson is likely to be quite similar for $D$ and $B$ mesons, if one considers both $c$ and $b$ quarks as asymptotically heavy. With all the present uncertainty in the knowledge of the strong interaction parameters, one can see from this comparison that the expected variation of the ratio $R^{c/n}$ is quite less prominent for $D$ mesons at the $\psi(3770)$ peak than for $B$ mesons at the $\Upsilon(4S)$. Hopefully, the amplitude of the variation of $R^{c/n}$ at the $\psi(3770)$ is still sufficient for a study in the upcoming CLEO-c experiment.

## 4. DISCUSSION OF THE $\phi(1020)$ DATA AND SUMMARY

So far, the most detailed scan data for the production cross section of charged and neutral mesons are available only for the kaons in the vicinity of

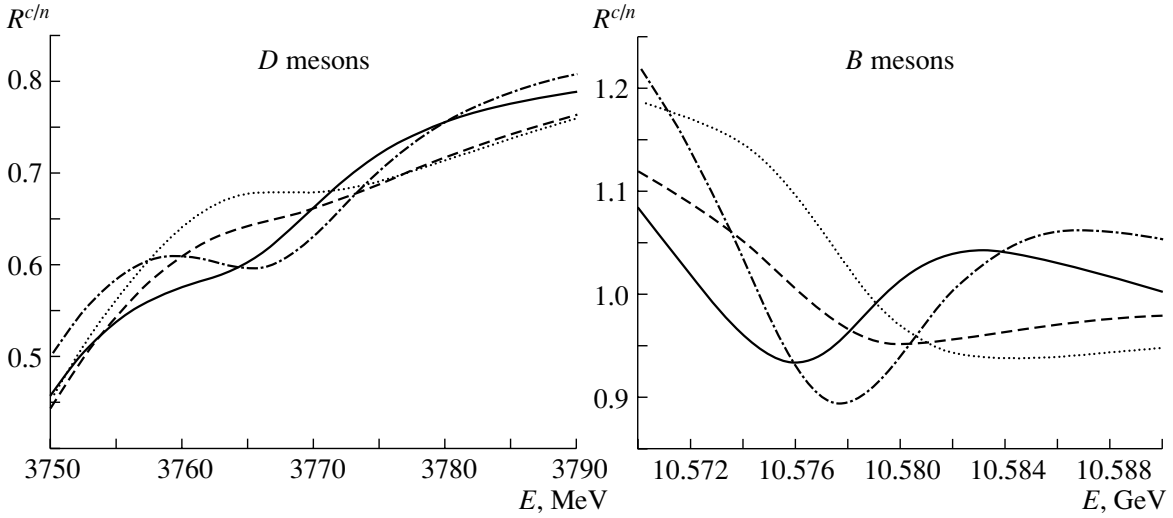


**Fig. 1.** The energy dependence of the ratio $R^{c/n}$ for $D$- and $B$-meson pair production in the region of the respective resonance: $\psi(3770)$ and $\Upsilon(4S)$ (the center positions are assumed at $E_0 = 3.770$ and 10.580 GeV, respectively) for some values of $a$ and $\delta_1(E_0)$: $a^{-1} = 200$ MeV, $\delta_1(E_0) = 30°$ (solid curve); $a^{-1} = 200$ MeV, $\delta_1(E_0) = -30°$ (dashed curve); $a^{-1} = 300$ MeV, $\delta_1(E_0) = 30°$ (dash-dotted curve); and $a^{-1} = 300$ MeV, $\delta_1(E_0) = -30°$ (dotted curve).

the $\phi(1020)$ resonance [7], and it is quite natural to discuss whether any hint at the discussed variation of $R^{c/n}$ is indicated by those data. A consideration of this ratio at the $\phi$ peak, however, encounters certain peculiarities. As mentioned above, the Coulomb interaction effect is relatively smaller for production of $K$ mesons at the $\phi$ resonance, and also the relativistic effects can play a certain role. The analysis of the ratio $R^{c/n}$ is further complicated by the fact that the production amplitude also receives an isovector contribution, which has opposite sign for $K^+K^-$ and $K_LK_S$ and thus changes the charged-to-neutral ratio. (In the vector dominance model this contribution is considered as the "tail" of the $\rho$ resonance.) Neglecting a smooth variation of the nonresonant $I = 1$ amplitude and also of the nonresonant part of the $I = 0$ amplitude (e.g., the tail of the $\omega$ resonance), one can approximate the formula for $R^{c/n}$ in this case as

$$R^{c/n} = F_c \left(\frac{p_+}{p_0}\right)^3 \left|\frac{1 + (A_0 - A_1)(\Delta + i\gamma)}{1 + (A_0 + A_1)(\Delta + i\gamma)}\right|^2, \quad (20)$$

where $A_1$ and $A_0$ stand for the appropriately normalized relative contribution of the nonresonant $I = 1$ and $I = 0$ amplitudes. The charged $K$ mesons are lighter than the neutral ones; thus, the cube of the ratio of the momenta is larger than one and decreases towards one at energies far above the threshold. The data [7], however, display a slight general increase in the ratio with energy, which shows that the effect of $A_1$ is not negligible. This makes any "absolute" prediction of $R^{c/n}$ in this case rather troublesome, as well as of its general variation with energy on a scale somewhat larger than the width of $\phi$. A real fit of the involved parameters, including the rapid variation of the Coulomb factor $F_c$ discussed here, requires knowledge of the raw data and of the correlation in the errors. For this reason, here in Fig. 2 is shown a comparison of the data sets listed in [7] with a representative theoretical curve.

More specifically, in [7] are listed (in Table IX) the scan data for the cross section for production of separately the $K^+K^-$ pairs and the $K_LK_S$ pairs acquired in two different scans (called PHI9801 and PHI9802). Furthermore, the detection of the $K_S$ mesons was done by two separate methods, i.e., by detecting their decay into charged pions and into neutral pions. Accordingly, paper [7] lists two separate sets of data for the $K_LK_S$ production measured by each of these two methods. The systematic error for the $K^+K^-$ cross section is listed at 7.1% and for the neutral kaons at 4 and 4.2% for the two identification techniques used. The statistical errors are listed for each individual entry for the cross section. The data points in the plot of Fig. 2 are the charged-to-neutral ratios calculated from the data listed in [7] with the errors corresponding to the listed statistical errors only. The representative theoretical curve corresponds to $1.38F_c$. In other words, the absolute normalization is set rather arbitrarily, given the described theoretical uncertainty in calculating the absolute normalization $R^{c/n}$ at the $\phi$ resonance and given the systematic uncertainty in the data. Also, any overall variation of the kinematical and amplitude factors in the shown energy interval is neglected. Changing the

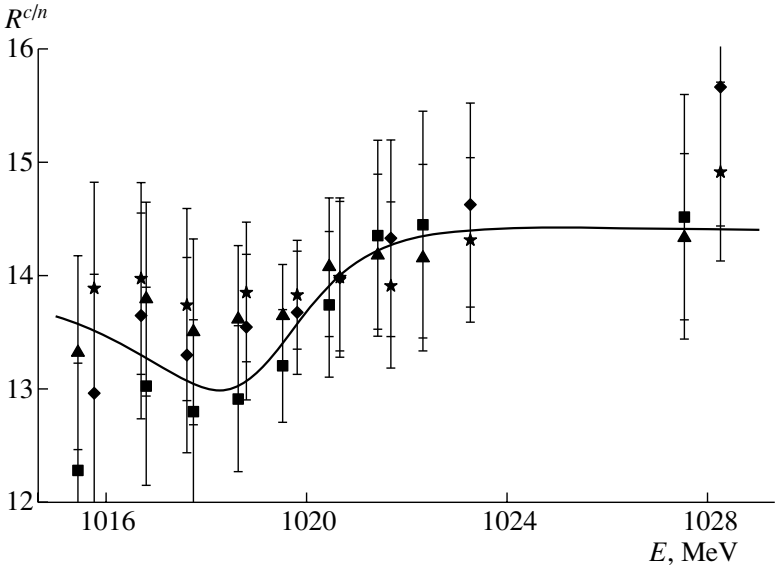


**Fig. 2.** The SND data [7] at the $\phi(1020)$ resonance represented as the charged-to-neutral yield ratio for the two reported scans (PHI9801 and PHI9802) in [7] and for two methods of identifying the $K_S$ mesons by their decay into neutral or into charged pions: PHI9801 neutral (stars), PHI9801 charged (diamonds), PHI9802 neutral (triangles), and PHI9802 charged (squares). The solid curve is explained in the text.

normalization of the curve and also including the overall energy dependence would result in a vertical shift and a slight tilt of the curve. The behavior of the Coulomb correction factor in the curve shown in Fig. 2 is calculated assuming $E_0 = 1019.5$ MeV, $\Gamma = 4.26$ MeV, $a^{-1} = 200$ MeV, and $\delta(E_0) = 40°$. The only purpose of the theoretical curve in Fig. 2 is to illustrate the approximate magnitude of the expected effect of the variation of $R^{c/n}$ at the $\phi(1020)$ resonance, and by no means should it be considered as a detailed prediction.

Due to large experimental and theoretical uncertainties as explained in the text and can be seen from Fig. 2, it is not entirely clear whether the available data [7] exclude or rather suggest a variation with the expected magnitude of the ratio $R^{c/n}$ within the width of the $\phi$ resonance. Perhaps, a detailed global fit of the raw data could provide a statistically significant evaluation of the amplitude of such variation.

In summary, the strong interaction phase significantly modifies the behavior of the cross section for production of pairs of charged pseudoscalar mesons near threshold in $e^+e^-$ annihilation. A simple calculation of this effect in the first order in the Coulomb interaction is presented in Section 2. The existence of a strong near-threshold resonance gives rise to a rapid variation of the charged-to-neutral yield ratio due to the interference of the resonance scattering phase with the Coulomb phase. The specific behavior of this variation is sensitive to details of the structure of the mesons and of their strong interaction. Thus, measuring the discussed effect for $B$ mesons at the $\Upsilon(4S)$ resonance and for $D$ mesons at the $\psi(3770)$ resonance might provide information on these details, which are not accessible by other means. The estimated effect for $D$ mesons in the region of the $\psi(3770)$ is less prominent than for $B$ mesons at the $\Upsilon(4S)$ but possibly is still measurable. The expected effect of the rapid variation of the charged-to-neutral ratio for the kaon production at the $\phi(1020)$ resonance is the smallest, and, conservatively, the available data appear to be not conclusive enough to confirm or exclude such variation with a statistical significance.

## ACKNOWLEDGMENTS

I thank J. Rosner for stimulating questions regarding the estimates of the behavior of $R^{c/n}$ at the $\psi(3770)$ resonance and A. Vainshtein for a useful discussion of the kaon production in the $\phi(1020)$ region.

This work is supported in part by the DOE, grant no. DE-FG02-94ER40823.

## REFERENCES

1. J. P. Alexander *et al.* (CLEO Collab.), Phys. Rev. Lett. **86**, 2737 (2001).
2. S. B. Athar *et al.* (CLEO Collab.), Phys. Rev. D **66**, 052003 (2002).
3. B. Aubert *et al.* (BABAR Collab.), Phys. Rev. D **65**, 032001 (2002).
4. B. Aubert *et al.* (BABAR Collab.), hep-ex/0401028.

5. N. C. Hastings *et al.* (Belle Collab.), Phys. Rev. D **67**, 052004 (2003).
6. The CLEO-c, http://www.lns.cornell.edu/public/CLEO/spoke/CLEOc/
7. M. N. Achasov *et al.* (SND Collab.), Phys. Rev. D **63**, 072002 (2001).
8. K. Hagivara *et al.* (Particle Data Group), Phys. Rev. D **66**, 010001 (2002).
9. M. B. Voloshin, Mod. Phys. Lett. A **18**, 1783 (2003).
10. D. Atwood and W. J. Marciano, Phys. Rev. D **41**, 1736 (1990).
11. G. P. Lepage, Phys. Rev. D **42**, 3251 (1990).
12. N. Byers and E. Eichten, Phys. Rev. D **42**, 3885 (1990).
13. R. Kaiser, A. V. Manohar, and T. Mehen, Phys. Rev. Lett. **90**, 142001 (2003).
14. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Non-Relativistic Theory* (Nauka, Moscow, 1989; Pergamon, Oxford, 1977).
15. H. Albrecht *et al.* (ARGUS Collab.), Z. Phys. C **65**, 619 (1995).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 779–783. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 812–816.*


# The Problem of "Saturation" in DIS and Heavy-Ion Collisions*

**A. B. Kaidalov****

*Institute of Theoretical and Experimental Physics,*
*Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received August 3, 2004

**Abstract**—The problem of saturation of parton densities in deep inelastic scattering (DIS) in the limit $x \to 0$ and in heavy-ion collisions at very high energies is investigated in the framework of Reggeon theory. It is emphasized that the shadowing effects are important in a definite kinematical region. The model which includes unitarity effects and gives a good description of structure functions and diffractive processes in DIS is discussed. In the limit $x \to 0$, the "saturation" of parton densities is achieved. The Gribov approach to interaction with nuclei allows one to describe structure functions of nuclei in the small-$x$ region. It is shown how the prediction of the Glauber model for the density of hadrons produced in the central rapidity region in nucleus–nucleus interactions is modified due to shadowing of small-$x$ partons. Calculations show that shadowing effects are important at RHIC energies, but the situation is far from saturation. Production of particles with large transverse momenta and jets is discussed. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Investigation of the deep inelastic scattering (DIS) in the region of small $x$-Bjorken gives important information on the behavior of partonic densities in the regime when these densities are large and nonlinear effects in their evolution are important. Very large densities can be obtained in heavy-ion collisions at very high energies and the dynamics of these processes has many features common to DIS in the limit $x \to 0$. Experimental data on small-$x$ DIS obtained at HERA [1] show that distributions of quarks and gluons have a fast increase as $x$ decreases up to values $\sim 10^{-4}$, which can be obtained at HERA. In Reggeon theory, this increase is related to an intercept of the leading Pomeron singularity $\alpha_{\mathrm{P}}(0)$: $xq(x)$ or $xg(x) \sim 1/x^{\Delta}$, where $\Delta \equiv \alpha_{\mathrm{P}}(0) - 1$ (see, for example, [2]). There are good reasons to believe (for reviews on this subject, see [2, 3]) that this fast increase will be modified for even smaller values of $x$ due to the influence of unitarity effects. In the limit $x \to 0$, partonic distributions $xq(x)(xg(x))$ reach the limit of saturation and increase only as powers of $\ln(1/x)$. The problem of saturation is even more important in heavy-ion collisions [4], where densities of partons in colliding nuclei are very large. This problem has been studied in QCD perturbation theory [4, 5], and partonic configurations, which include nonlinear effects in QCD evolution, were called "color glass condensate" [5]. However, the pattern of saturation is much more general.

In this paper, I first shall review theoretical approaches to the origin of the Pomeron in QCD. In Section 3, I shall consider the problem of small-$x$ behavior in DIS and shall discuss the model [6] based on Reggeon theory and the dipole picture of interaction of virtual photons with nucleons. This model gives a self-consistent description of structure functions of a proton and diffraction dissociation of a photon in a broad region of virtualities $Q^2$. I discuss the pattern of saturation in this model. It is shown in Section 4 that the Gribov approach to interactions with nuclei allows one to calculate shadowing effects in the small-$x$ region for structure functions of nuclei. Production of hadrons and jets in heavy-ion collisions is described in Section 5. It is emphasized that the coherence condition strongly limits the kinematic region where the shadowing effects for inclusive cross sections are important. In the same approach which leads to description of nuclear structure functions, it is shown that shadowing effects in heavy-ion collisions lead to a sizable reduction of inclusive particle spectra in comparison with the Glauber model even at RHIC energies and agree with existing experimental data. It is emphasized, however, that at RHIC (and even LHC) energies, the situation is far from the saturation limit. The effects which can lead to suppression of production for particles with large $p_{\perp}$ at RHIC are discussed.

*This article was submitted by the author in English.

** e-mail: kaidalov@itep.ru

1063-7788/05/6805-0779$26.00 

## 2. POMERON AND SMALL-$x$ PHYSICS

Scattering amplitude of a virtual photon on a proton target at fixed virtuality $Q^2$ and very high c.m. energy $W$ is described in Reggeon theory by exchange of the Pomeron pole and multi-Pomeron cuts. An increase in the structure function $F_2(x,Q^2)$ as $x \approx Q^2/W^2 \to 0$ in the pole approximation $\sim 1/x^{\Delta}$, where $\Delta \equiv \alpha_{\rm P}(0) - 1$. At large $Q^2$,

$$F_2(x,Q^2) \sim \sum_i e_i^2 x(q_i(x,Q^2) + \overline{q}_i(x,Q^2)),$$

where $q_i(x,Q^2)$ is the distribution of quarks of flavor $i$. Thus, for $\Delta > 0$, the density of partons increases as $x \to 0$. The value of $\Delta$ was calculated in QCD perturbation theory. Summation of the leading $\ln(1/x)$ leads to the BFKL [7] Pomeron with the value of $\Delta = (12\ln 2/\pi)\alpha_s \approx 0.5$, which corresponds to a very fast increase in parton densities. In this leading approximation, the Pomeron corresponds to a cut in the $j$ plane. In the next-to-leading approximation, the leading singularity is the Regge pole[1)] and the value of $\Delta$ substantially decreases to the values $\Delta = 0.15-0.25$ [8]. This leading pole strongly depends on the nonperturbative region of small momentum transfer [9]. A connection of the Pomeron with the spectrum of glueballs and the role of nonperturbative effects were studied in [10], using the method of vacuum correlators [11]. It was shown [10] that confinement effects and mixing of gluonic and $q\bar{q}$-Regge trajectories are important for the value of the intercept of the Pomeron trajectory. The Pomeron pole with $\Delta > 0$ leads for $s \equiv W^2 \to \infty$ to a violation of $s$-channel unitarity. It is well known (see, for example, [2]) that the unitarity is restored if the multi-Pomeron exchanges in the $t$ channel are taken into account. In the following, I will use for this intercept the value $\Delta \approx 0.2$, found from analysis of experimental data on high-energy hadronic interactions with account of multi-Pomeron contributions [12].

## 3. MODELS FOR SMALL-$x$ PHYSICS

Experimental data at HERA [1] show that distributions of quarks and gluons have a fast rise as $x$ decreases. For parametrization of $F_2(x,Q^2)$ at small $x \sim 1/x^{\Delta}$, data demonstrate that $\Delta$ increases with $Q^2$ and is close to 0.2 at $Q^2 \sim 10\ \mathrm{GeV}^2$. This was interpreted in [13] as a consequence of the decrease in multi-Pomeron contributions as $Q^2$ increases. A similar interpretation in terms of saturation of parton densities has been proposed by Golec-Biernat and Wüsthoff (G-BW) [14]. In the Pomeron theory, the multi-Pomeron contributions are related to diffractive production processes, and for self-consistency, it is important to describe simultaneously total cross sections (structure functions) and diffractive processes. At high energies, virtual photons dissociate first to $q\bar{q}$ pairs, which interact with a target. For small relative transverse distance $r_\perp$ between $q$ and $\bar{q}$, such a dipole has a small ($\sim r_\perp^2$) total interaction cross section and small diffraction dissociation cross section. On the other hand, for a large-size pair, these cross sections are not small. As a result, the role of multi-Pomeron rescatterings decreases as $Q^2$ increases. Explicit models which take these effects into account have been constructed in the framework of the Reggeon theory. The contribution of small-size configurations $\sigma_s^{(\rm tot)}$ has been described in QCD perturbation theory in a form similar to the one used in [14]:

$$\sigma_s^{({\rm tot})T(L)}(s,Q^2) \tag{1}$$
$$= 4\int d^2b \int_0^{r_0} d^2r \int_0^1 dz |\Psi^{T(L)}(r,z,Q^2)|^2 \times \sigma_s(b,r,s,Q^2),$$

where $\Psi^{T(L)}(r,z,Q^2)$ are wave functions of $q\bar{q}$ pair for transverse (longitudinal) photons. Integration on transverse size $r$ was limited by the value $r_0 \approx 0.2$ fm.

Contributions of large-distance dipoles were described in the same framework as hadronic interactions. An important difference from the G-BW model is that the model [6] corresponds to summation of a certain class of Reggeon diagrams and formulas for cross sections $\sigma(b,r,s,Q^2)$ are given in the impact parameter representation. Also, the fan-type diagrams with triple Pomeron interaction are included [6]. This leads to a qualitative difference from the G-BW model in the pattern of saturation for very small $x$ ($x \lesssim 10^{-5}$)[2)]. In particular, in the saturation region, cross sections of $q\bar{q}$ pairs with a target do not tend to a constant (as in the G-BW model), but have the Froissart-type increase $\sim \ln^2(1/x)$. This is due to a logarithmic increase in the interaction radius and is beyond the scope of QCD perturbation theory. It is worth emphasizing that the model of [6] predicts that an approach to the unitarity limit or saturation regime is very slow and, for large $Q^2$ ($Q \sim 10^2\ \mathrm{GeV}^2$), it is not achieved even for $x \sim 10^{-12}$.

[1)] There is a sequence of poles in the $j$ plane concentrated at point $j = 1$.

[2)] Models of [6], as well as the G-BW model [14], give a good description of HERA data on structure functions $F_2$ and diffractive production.

## 4. STRUCTURE FUNCTIONS OF NUCLEI IN THE SHADOWING REGION

Study of nuclear structure functions in the small-$x$ region can provide important information on mechanisms of shadowing. The total cross section for a virtual photon–nucleus interaction can be written as a series in number of rescatterings on nucleons of a nucleus

$$\sigma_{\gamma^* A}^{(\mathrm{tot})} = \sigma_{\gamma^* A}^{(1)} + \sigma_{\gamma^* A}^{(2)} + \dots. \quad (2)$$

The first term corresponds to a sum of incoherent interactions and is equal to[3)]

$$\sigma_{\gamma^* A}^{(1)} = A\sigma_{\gamma^* N}; \quad (3)$$

the second term describes the shadowing and according to Gribov theory [15] can be expressed in terms of the cross section of diffraction dissociation of a photon on a nucleon target

$$\sigma^{(2)} = -4\pi \int d^2 b T_A^2(b) \quad (4)$$
$$\times \int dM^2 \frac{d\sigma_{\gamma^* N}^{DD}(t=0)}{dM^2 dt} f_A(t_{\min}),$$

where $T_A(b) = \int dz \rho_A(r)$ is the nuclear profile function, $\rho(r)$ is nuclear density ($\int \rho_A(r) d^3 r = A$), and

$$f_A(t) = \frac{1}{A} \int e^{i\mathbf{q}\cdot\mathbf{r}} \rho_A(r) d^3 r$$

is the nuclear form factor. The minimal value of momentum transfer to nucleons

$$t_{\min} = -x_{\mathrm{P}}^2 m_N^2, \quad x_{\mathrm{P}} = \frac{Q^2 + M^2}{s} = \frac{x}{\beta}, \quad (5)$$

where $\beta = Q^2/(Q^2 + M^2)$. It is important to emphasize that the shadowing effects are different from zero only in the region of very small $x_{\mathrm{P}}$ or $x$, where $t_{\min} \ll 1/R_A^2$ ($R_A$ is the nuclear radius). The same condition of coherence corresponds to lifetimes of the initial hadronic fluctuation $\tau_h \sim 1/(m_N x)$ much larger than the radius of a nucleus. These well-known conditions for existence of shadowing are also important for heavy-ion collisions (see below). Higher order rescatterings in Eq. (4) are model dependent. In [16, 17], they were calculated in the Schwimmer model [18], where

$$F_{2A}/F_{2N} = \int d^2 b \frac{T_A(b)}{1 + F(x, Q^2) T_A(b)}, \quad (6)$$

with

$$F(x, Q^2) = 4\pi \int dM^2 \left( d\sigma_{\gamma^* N}^{DD}(t=0)/dM^2 dt \right)$$
$$\times \left( f_A(t_{\min})/\sigma_{\gamma^* N}(x, Q^2) \right).$$

Using information on diffraction dissociation in $\gamma^* N$ collisions, it is possible to calculate nuclear structure functions in good agreement with experimental data. These calculations show that the shadowing effects increase as $x$ decreases in the region $x < 0.1$. The shadowing correction $\sigma^{(2)}/\sigma^{(1)}$ increases with atomic number $\sim A^{1/3}$, but in the region of $x \sim 10^{-2}$, relevant for RHIC energies (see below), and for $Q^2 \sim 1\ \mathrm{GeV}^2$, a decrease in $F_{2A}$ due to shadowing is less than 25% even for heavy nuclei (Au).

Thus, the shadowing effects for distributions of quarks (which are directly measured in DIS) can be well described theoretically and prediction for smaller $x$, relevant for LHC energies, looks reliable. Information on shadowing effects for gluons is more limited. This information can be obtained from the data on "gluonic content" of the Pomeron [19]. Existing information indicates that, at least for $x \gtrsim 10^{-3}$, shadowing for gluons is of similar size as the one for quarks.

## 5. HEAVY-ION COLLISIONS

In heavy-ion collisions, the problem of shadowing of soft partons of colliding nuclei is very important. This phenomenon takes place only in the kinematical region where longitudinal momentum fractions of soft partons satisfy the condition $x_i \ll 1/(m_N R_A)$ ($i = 1, 2$) discussed in the previous section. Values of $x_i$ can be determined from the kinematical conditions

$$x_1 x_2 = M_\perp^2/s, \quad x_1 - x_2 = x_{\mathrm{F}},$$

where $M_\perp$ is the transverse mass of the system produced in the collision and $x_{\mathrm{F}}$ is its Feynman $x$.

For example, at RHIC energy $\sqrt{s} = 200$ GeV in the central rapidity region ($x_{\mathrm{F}} \approx 0$), the shadowing for Au–Au collisions is possible only for transverse momenta less than several GeV. In the forward region ($x_2 \ll x_1 \approx x_{\mathrm{F}}$), the shadowing is absent for partons of the nucleus 1 and extends to $p_\perp \sim 10$ GeV for partons of nucleus 2.

It is possible to calculate the shadowing effects for inclusive densities of produced particles using the Gribov approach [15] to nuclear interactions. For models of the Glauber type, which do not take into account interactions between Pomerons, there is the Abramovsky–Gribov–Kancheli (AGK) rule [20], which expresses inclusive spectra in $A_1 A_2$ collisions in terms of spectra in $NN$ collisions

$$\frac{d\sigma_{A_1 A_2}}{dy} = T_{A_1 A_2}(b) \frac{d\sigma_{NN}}{dy}, \quad (7)$$

[3)] In the small-$x$ region, it is possible to neglect a difference between interaction of $\gamma^*$ with $p$ and $n$.

where

$$T_{A_1A_2}(b) = \int d^2s T_{A_1}(\mathbf{b}-\mathbf{s}) T_{A_2}(\mathbf{s}).$$

For density of particles, we obtain from Eq. (7)

$$\frac{dn_{A_1A_2}(b)}{dy} = n_{A_1A_2}(b)\frac{dn_{NN}}{dy}, \qquad (8)$$

where

$$n_{A_1A_2}(b) = \frac{T_{A_1A_2}(b)\sigma_{NN}^{(\mathrm{tot})}}{\sigma_{A_1A_2}^{(\mathrm{tot})}}$$

is the number of collisions in the Glauber model.

Note that Eqs. (7), (8) are valid in the Glauber model for $x_\mathrm{F} \approx 0$ in the limit $s \to \infty$. For finite energies, there are kinematical corrections due to energy–momentum conservation effects [21].

In Reggeon theory, the shadowing effects discussed above are related to diagrams with interactions between Pomerons. They lead to a substantial decrease in particle densities at high energies. On the other hand, different cuttings of the same diagrams correspond to reinteractions between particles produced in $NN$ subcollisions and can lead to equilibration in the system produced in heavy-ion collisions. In string models, interactions between Pomerons can be considered as interactions of strings, which can lead to percolation in the string system [22].

The same model, which has been used for description of shadowing effects for nuclear structure functions, was used for prediction of inclusive particle densities in heavy-ion collisions [21]. In this approach, expression (8) for densities of particles in $A_1A_2$ collisions is modified as follows [21]:

$$\frac{dn_{A_1A_2}}{dy} = n_{A_1A_2}(b)\frac{dn_{NN}}{dy}\gamma_{A_1}\gamma_{A_2}, \qquad (9)$$

where

$$\gamma_A = \frac{1}{A}\int d^2b \frac{T_A(b)}{1+F(x,Q^2)T_A(b)}. \qquad (10)$$

For LHC, Eqs. (9), (10) predict a decrease in particle densities by a factor of $\sim$4 compared to the Glauber model [21], while for RHIC the suppression is $\approx$2. Detailed predictions for shadowing suppression are given in [17]. This approach agrees [2] with the results of experiments at RHIC [23, 24]. The dependence of number of produced particles on number of participant nucleons is also in agreement with data [25].

The kinematical borders for shadowing discussed above correspond also to an essential change in the spacetime picture of nuclear interactions and change [26] in AGK cutting rules. For example, for $x_{A_2} \ll 1/(m_N R_{A_2})$, the lifetime for the initial-state configuration of a nucleus $A_1$ is larger than the size of a nucleus $A_2$. In this case, conditions for coherence are satisfied and usual AGK cutting rules [20] are valid. For $x_{A_2} \gtrsim 1/(m_N R_A)$, the space-time picture is close to final-state reinteractions of produced particles with formation time [26]. The last situation corresponds to a picture used in most of the existing models of heavy-ion collisions. Note that it is valid only in the nonshadowing region. This situation is very general. For $J/\psi$ production in hadron–nucleus interactions, it was discussed in detail in [27]. It has important implications for production of particles and jets with large $p_\perp$ at RHIC. For $x_\mathrm{F} = 0$ and $p_\perp \approx 1$–$2$ GeV, the shadowing effects at $\sqrt{s} = 200$ GeV are small, as was emphasized below and is demonstrated by the data on D–Au collisions [28]. However, in this region of $x_{A_i} \gtrsim 1/(mR_A)$, the final-state interaction effects are important. In particular, account of the Cronin effect and interaction of a particle (jet) with comovers allowed one to describe [29] a suppression of particle production at large $p_\perp$ observed at RHIC [24] for Au–Au collisions. The data in D–Au collisions [28] can also be understood in the same model.

## 6. CONCLUSIONS

The Pomeron theory of high-energy interactions is able to describe a large class of physical processes including small-$x$ DIS and nuclear interactions.

An important problem for high-energy physics is the nature of the Pomeron. The analysis of this problem in QCD with inclusion of both nonperturbative and perturbative effects shows that the Pomeron has a very rich dynamical structure.

The problem of saturation of parton densities is related in the Pomeron theory to interactions between Pomerons. These effects are especially important for heavy-ion collisions. The shadowing effects are clearly seen in the RHIC data, but the saturation is not achieved at RHIC energies.

## ACKNOWLEDGMENTS

I would like to thank N. Armesto, K. Boreskov, A. Capella, V. Fadin, E.G. Ferreiro, O.V. Kancheli, V.A. Khoze, E. Levin, L.N. Lipatov, A. Martin, C. Merino, C.A. Salgado, K.A. Ter-Martirosyan, and J. Tran Thanh Van for useful discussions.

This work is supported in part by INTAS (grant no. 00-00366), the grant CRDF 13643, and the Russian Foundation for Basic Research (project no. 04-02-17263).

## REFERENCES


1. T. Ahmed *et al.* (H1 Collab.), Phys. Lett. B **299**, 374 (1993); C. Adloff *et al.* (H1 Collab.), Nucl. Phys. B **497**, 3 (1997); M. Derrick *et al.* (ZEUS Collab.), Phys. Lett. B **293**, 465 (1992); J. Breitweg *et al.* (ZEUS Collab.), Phys. Lett. B **407**, 432 (1997).
2. A. B. Kaidalov, Surveys High Energy Phys. **9**, 143 (1996); **16**, 267 (2001).
3. L. V. Gribov, E. M. Levin, and M. G. Ryskin, Phys. Rep. **100**, 1 (1983); E. Laenen and E. Levin, Annu. Rev. Nucl. Part. Sci. **44**, 199 (1994); A. H. Mueller, hep-ph/9911289.
4. L. Mc Lerran and R. Venugopalan, Phys. Rev. D **49**, 2233 (1994); **50**, 2225 (1994); **53**, 458 (1996); J. Jalilian-Marian *et al.*, Phys. Rev. D **59**, 014014, 034007 (1999); A. Kovner, L. Mc Lerran, and H. Weigert, Phys. Rev. D **52**, 3809, 6231 (1995); Yu. V. Kovchegov and A. H. Mueller, Nucl. Phys. B **529**, 451 (1998); Yu. V. Kovchegov, A. H. Mueller, and S. Wallon, Nucl. Phys. B **507**, 367 (1997); A. H. Mueller, Nucl. Phys. B **558**, 285 (1999).
5. L. Mc Lerran, Nucl. Phys. A **702**, 49 (2002).
6. A. Capella, E. G. Ferreiro, C. A. Salgado, and A. B. Kaidalov, Nucl. Phys. B **593**, 336 (2001); Phys. Rev. D **63**, 054010 (2001).
7. E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Zh. Éksp. Teor. Fiz. **71**, 840 (1976) [Sov. Phys. JETP **44**, 443 (1976)]; **72**, 377 (1977) [**45**, 199 (1977)]; I. I. Balitsky and L. N. Lipatov, Yad. Fiz. **28**, 1597 (1978) [Sov. J. Nucl. Phys. **28**, 822 (1978)].
8. V. S. Fadin and L. N. Lipatov, Phys. Lett. B **429**, 127 (1998); G. Camici and M. Ciafaloni, Phys. Lett. B **430**, 349 (1998); S. J. Brodsky *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 161 (1999) [JETP Lett. **70**, 155 (1999)].
9. L. P. A. Haakman, O. V. Kancheli, and J. H. Koch, Nucl. Phys. B **518**, 275 (1998).
10. A. B. Kaidalov and Yu. A. Simonov, Phys. Lett. B **477**, 163 (2000); Yad. Fiz. **63**, 1507 (2000) [Phys. At. Nucl. **63**, 1428 (2000)].
11. Yu. A. Simonov, Phys. Lett. B **249**, 514 (1990).
12. A. B. Kaidalov, L. A. Ponomarev, and K. A. Ter-Martirosian, Yad. Fiz. **44**, 722 (1986) [Sov. J. Nucl. Phys. **44**, 468 (1986)].
13. A. Capella, A. Kaidalov, C. Merino, and J. Tran Thanh Van, Phys. Lett. B **337**, 358 (1994).
14. K. Golec-Biernat and M. Wüsthoff, Phys. Rev. D **59**, 014017 (1999); **60**, 114023 (1999).
15. V. N. Gribov, Zh. Éksp. Teor. Fiz. **56**, 892 (1969) [Sov. Phys. JETP **29**, 483 (1969)]; **57**, 1306 (1969) [**30**, 709 (1969)].
16. A. Capella, A. Kaidalov, C. Merino, *et al.*, Eur. Phys. J. C **5**, 111 (1998).
17. J. L. Albacete, N. Armesto, A. Capella, *et al.*, hep-ph/0308050.
18. A. Schwimmer, Nucl. Phys. B **91**, 445 (1975).
19. C. Adloff *et al.* (H1 Collab.), Z. Phys. C **76**, 613 (1997); abstract 980 presented to ICHEP02.
20. V. A. Abramovsky, V. N. Gribov, and O. V. Kancheli, Yad. Fiz. **18**, 595 (1973) [Sov. J. Nucl. Phys. **18**, 308 (1973)].
21. A. Capella, A. Kaidalov, and J. Tran Thanh Van, Heavy Ion Phys. **9**, 169 (1999).
22. M. A. Braun, F. del Morel, and C. Pajares, Nucl. Phys. A **715**, 791 (2003).
23. B. B. Back *et al.* (PHOBOS Collab.), Phys. Rev. Lett. **87**, 102303 (2001).
24. S. S. Adler *et al.* (Phenix Collab.), nucl-ex/0304022.
25. A. Capella and D. Sousa, Phys. Lett. B **511**, 185 (2001).
26. K. G. Boreskov *et al.*, Yad. Fiz. **53**, 569 (1991) [Sov. J. Nucl. Phys. **53**, 356 (1991)].
27. K. Boreskov, A. Capella, A. Kaidalov, and J. Tran Tranh Van, Phys. Rev. D **47**, 919 (1993).
28. Phenix Collab., nucl-ex/0310019.
29. A. Capella, E. G. Ferreiro, A. B. Kaidalov, and D. Sousa, hep-ph/0403081.

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 784–807. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 817–840.*


# Properties of Doubly Heavy Baryons in the Relativistic Quark Model*

**D. Ebert[1], R. N. Faustov[2], V. O. Galkin[2], and A. P. Martynenko[3]**

Received May 5, 2004; in final form, November 2, 2004

**Abstract**—Mass spectra and semileptonic decay rates of baryons consisting of two heavy ($b$ or $c$) and one light quark are calculated in the framework of the relativistic quark model. The doubly heavy baryons are treated in the quark–diquark approximation. The ground and excited states of both the diquark and quark–diquark bound systems are considered. The quark–diquark potential is constructed. The light quark is treated completely relativistically, while the expansion in the inverse heavy-quark mass is used. The weak transition amplitudes of heavy diquarks $bb$ and $bc$ going, respectively, to $bc$ and $cc$ are explicitly expressed through the overlap integrals of the diquark wave functions in the whole accessible kinematic range. The relativistic baryon wave functions of the quark–diquark bound system are used for the calculation of the decay matrix elements, the Isgur–Wise function, and decay rates in the heavy-quark limit. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

The description of baryons within the constituent quark model and quantum chromodynamics (QCD) is a very important problem. Since the baryon is a three-body system, its theory [1] is much more complicated compared to the two-body meson system. Even now, it is not clear which of the two main QCD models, $Y$ law or $\Delta$ law, correctly describes the nonperturbative (long-range) part of the quark interaction in the baryon [2, 3]. The popular quark–diquark picture of a baryon is not universal and does not work in all cases [4]. The success of the heavy-quark effective theory (HQET) [5] in predicting some properties of the heavy-light $q\bar{Q}$ mesons ($B$ and $D$) suggests applying these methods to heavy-light baryons too. The simplest baryonic systems of this kind are the so-called doubly heavy baryons ($qQQ$) [1, 6–12]. The two heavy quarks ($b$ or $c$) compose in this case a bound diquark system in the antitriplet color state, which serves as a localized color source. The light quark orbits around this heavy source at a much larger ($\sim 1/m_q$) distance than the source size ($\sim 2/m_Q$). Thus, the doubly heavy baryons look effectively like a two-body bound system and strongly resemble the heavy-light $B$ and $D$ mesons [2, 13]. Then the HQET expansion in the inverse heavy-quark mass can be used. The main distinction of the $qQQ$ baryon from the $q\bar{Q}$ meson is that the $QQ$ color source, though being almost localized, still is a composite system bearing integer spin values $(0, 1, \dots)$. Hence, it follows that the interaction of the heavy diquark with the light quark is not pointlike, but is smeared by the form factor expressed through the overlap integrals of the diquark wave functions. Besides this, the diquark excitations contribute to the baryon excited states.

Recently, the first experimental indications of the existence of doubly charmed baryons were published by SELEX [14]. Although these data need further experimental confirmation and clarification, this manifests that, in the near future, the mass spectra and decay rates of doubly heavy baryons will be measured. This gives additional grounds for the theoretical investigation of the doubly heavy baryon properties. The energies necessary to produce these particles have already been reached. The main difficulty remains in their reconstruction, since these particles have in general a large number of decay modes and thus high statistics is required [15].

In previous approaches to the calculation of doubly heavy baryon properties, the expansion in inverse powers not only of the heavy-quark (diquark) mass $m_Q$ ($M_d$) but also of the light-quark mass $m_q$ was carried out. The estimates of the light-quark velocity in these baryons show that the light quark is highly relativistic ($v/c \sim 0.7$–$0.8$). Thus, the nonrelativistic approximation is not adequate for the light quark. Here, we present a consistent treatment of doubly

*This article was submitted by the authors in English.

[1] Institut für Physik, Humboldt-Universität zu Berlin, Germany.
[2] Scientific Council for Cybernetics, Russian Academy of Sciences, Russia.
[3] Samara State University, Samara, Russia.

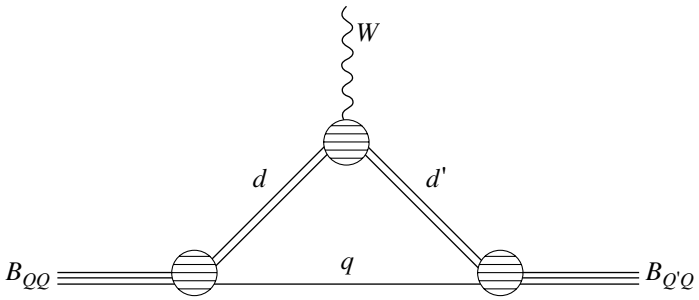


**Fig. 1.** Weak-transition matrix element of the doubly heavy baryon in the quark–diquark approximation.

heavy baryon properties in the framework of the relativistic quark model, based on the quasipotential wave equation without employing the expansion in $1/m_q$; namely, the light quark is treated fully relativistically. Concerning the heavy diquark (quark), we apply the expansion in $1/M_d$ ($1/m_Q$). Then we use the calculated wave functions for calculating the semileptonic decay rates of doubly heavy baryons in the quark–diquark approximation. The covariant expressions for the semileptonic decay amplitudes of the baryons with the spin 1/2, 3/2 are obtained in the limit $m_c, m_b \to \infty$ and compared with the predictions of HQET. The calculation of semileptonic decays of doubly heavy baryons ($bbq$) or ($bcq$) to doubly heavy baryons ($bcq$) or ($ccq$) can be divided into two steps (see Fig. 1). The first step is the study of form factors of the weak transition between initial and final doubly heavy diquarks. The second one consists in the inclusion of the light quark in order to compose a baryon with spin 1/2 or 3/2.

The paper is organized as follows. In Section 2, we describe our relativistic quark model, giving special emphasis to the construction of the quark–quark interaction potential in the diquark and the quark–diquark interaction potential in the baryon. In Section 3, we apply our model to the investigation of the heavy-diquark properties. The $cc$- and $bb$-diquark mass spectra are calculated. We also determine the diquark interaction vertex with the gluon, using the quasipotential approach, and calculate diquark wave functions. Thus, we take into account the internal structure of the diquark, which considerably modifies the quark–diquark potential at small distances and removes fictitious singularities. In Section 4, we construct the quasipotential of the interaction of a light quark with a heavy diquark. The light quark is treated fully relativistically. We use the expansion in inverse powers of the heavy-diquark mass to simplify the construction. First, we consider the infinitely heavy diquark limit and, then, include the $1/M^d_{QQ}$ corrections. In Section 5, we present our predictions for the mass spectra of the ground and excited states of $\Xi_{cc}$, $\Xi_{bb}$, $\Omega_{cc}$, and $\Omega_{bb}$ baryons. We consider the excitations of both the quark–diquark system and the diquark. The mixing between excited baryon states with the same total angular momentum and parity is discussed. For $\Xi_{cb}$ and $\Omega_{cb}$ baryons, composed of heavy quarks of different flavors, we give predictions only for ground states, since the excited states of the $cb$ diquark are unstable under the emission of soft gluons [11]. A detailed comparison of our predictions with other approaches is given. We reveal the close similarity of the excitations of the light quark in a doubly heavy baryon and a heavy-light meson. We also test the fulfillment of different relations between mass splittings of baryons with two $c$ or $b$ quarks, as well as the relations between splittings in the doubly heavy baryons and heavy-light mesons, following from the heavy-quark symmetry. Then we apply our model to the investigation of the heavy-diquark transition matrix elements in Section 6. The transition amplitudes of heavy diquarks are explicitly expressed in a covariant form through the overlap integrals of the diquark wave functions. The obtained general expressions reproduce in the appropriate limit the predictions of heavy-quark symmetry. Section 7 is devoted to the construction of transition matrix elements between doubly heavy baryons in the quark–diquark approximation. The corresponding Isgur–Wise function is determined. In Section 8, semileptonic decay rates of doubly heavy baryons are calculated in the nonrelativistic limit for heavy quarks. Section 9 contains our conclusions.

## 2. RELATIVISTIC QUARK MODEL

In the quasipotential approach and quark–diquark picture of doubly heavy baryons, the interaction of two heavy quarks in a diquark and the light-quark interaction with a heavy diquark in a baryon are described by the diquark wave function ($\Psi_d$) of the bound quark–quark state and by the baryon wave function ($\Psi_B$) of the bound quark–diquark state, respectively, which satisfy the quasipotential equation [16] of the

Schrödinger type [17]

$$\left(\frac{b^2(M)}{2\mu_R} - \frac{\mathbf{p}^2}{2\mu_R}\right)\Psi_{d,B}(\mathbf{p}) = \int \frac{d^3q}{(2\pi)^3} V(\mathbf{p},\mathbf{q};M)\Psi_{d,B}(\mathbf{q}), \quad (1)$$

where the relativistic reduced mass is

$$\mu_R = \frac{E_1E_2}{E_1+E_2} = \frac{M^4-(m_1^2-m_2^2)^2}{4M^3}, \quad (2)$$

and $E_1$ and $E_2$ are given by

$$E_1 = \frac{M^2-m_2^2+m_1^2}{2M}, \quad E_2 = \frac{M^2-m_1^2+m_2^2}{2M}. \quad (3)$$

Here, $M = E_1 + E_2$ is the bound-state (diquark or baryon) mass; $m_{1,2}$ are the masses of heavy quarks ($Q_1$ and $Q_2$), which form the diquark, or of the heavy diquark ($d$) and light quark ($q$), which form the doubly heavy baryon ($B$); and $\mathbf{p}$ is their relative momentum. In the center-of-mass system, the relative momentum squared on the mass shell reads

$$b^2(M) = \frac{[M^2-(m_1+m_2)^2][M^2-(m_1-m_2)^2]}{4M^2}. \quad (4)$$

The kernel $V(\mathbf{p},\mathbf{q};M)$ in Eq. (1) is the quasipotential operator of the quark–quark or quark–diquark interaction. It is constructed with the help of the off-mass-shell scattering amplitude, projected onto the positive energy states. In the following analysis, we closely follow the similar construction of the quark–antiquark interaction in mesons, which were extensively studied in our relativistic quark model [18, 19]. For the quark–quark interaction in a diquark, we use the relation $V_{QQ} = V_{Q\bar{Q}}/2$, arising under the assumption about the octet structure of the interaction from the difference of the $QQ$ and $Q\bar{Q}$ color states. An important role in this construction is played by the Lorentz structure of the confining interaction. In our analysis of mesons, while constructing the quasipotential of quark–antiquark interaction, we adopted that the effective interaction is the sum of the usual one-gluon-exchange term with the mixture of long-range vector and scalar linear confining potentials, where the vector confining potential contains the Pauli terms. We use the same conventions for the construction of the quark–quark and quark–diquark interactions in the baryon. The quasipotential is then defined as follows [10, 18]:

(a) for the quark–quark ($QQ$) interaction,

$$V(\mathbf{p},\mathbf{q};M) = \bar{u}_1(p)\bar{u}_2(-p)\mathcal{V}(\mathbf{p},\mathbf{q};M)u_1(q)u_2(-q), \quad (5)$$

with

$$\mathcal{V}(\mathbf{p},\mathbf{q};M) = \frac{2}{3}\alpha_s D_{\mu\nu}(\mathbf{k})\gamma_1^\mu\gamma_2^\nu + \frac{1}{2}V_{\rm conf}^V(\mathbf{k})\Gamma_1^\mu\Gamma_{2;\mu} + \frac{1}{2}V_{\rm conf}^S(\mathbf{k});$$

(b) for quark–diquark ($qd$) interaction,

$$V(\mathbf{p},\mathbf{q};M) = \frac{\langle d(P)|J_\mu|d(Q)\rangle}{2\sqrt{E_d(p)E_d(q)}}\bar{u}_q(p)\frac{4}{3}\alpha_s D_{\mu\nu}(\mathbf{k})\gamma^\nu u_q(q) + \psi_d^*(P)\bar{u}_q(p)J_{d;\mu}\Gamma_q^\mu V_{\rm conf}^V(\mathbf{k})u_q(q)\psi_d(Q) + \psi_d^*(P)\bar{u}_q(p)V_{\rm conf}^S(\mathbf{k})u_q(q)\psi_d(Q), \quad (6)$$

where $\alpha_s$ is the QCD coupling constant; the color factor is equal to $2/3$ for quark–quark and $4/3$ for quark–diquark interactions; $\langle d(P)|J_\mu|d(Q)\rangle$ is the vertex of the diquark–gluon interaction, which is discussed in detail below [$P = (E_d, -\mathbf{p})$ and $Q = (E_d, -\mathbf{q})$]; $D_{\mu\nu}$ is the gluon propagator in the Coulomb gauge:

$$D^{00}(\mathbf{k}) = -\frac{4\pi}{\mathbf{k}^2}, \quad D^{ij}(\mathbf{k}) = -\frac{4\pi}{k^2}\left(\delta^{ij} - \frac{k^ik^j}{\mathbf{k}^2}\right), \quad D^{0i} = D^{i0} = 0, \quad (7)$$

and $\mathbf{k} = \mathbf{p} - \mathbf{q}$; and $\gamma_\mu$ and $u(p)$ are the Dirac matrices and spinors:

$$u^\lambda(p) = \sqrt{\frac{\epsilon(p)+m}{2\epsilon(p)}}\begin{pmatrix} 1 \\ \dfrac{(\boldsymbol{\sigma}\cdot\mathbf{p})}{\epsilon(p)+m}\end{pmatrix}\chi^\lambda, \quad \lambda = \pm\frac{1}{2}, \quad (8)$$

with $\epsilon(p) = \sqrt{\mathbf{p}^2+m^2}$.

The diquark state in the confining part of the quark–diquark quasipotential (6) is described by the wave functions

$$\psi_d(P) = \begin{cases} 1, & \text{for scalar diquark,} \\ \varepsilon_d(p), & \text{for axial-vector diquark,} \end{cases} \quad (9)$$

where the four-vector

$$\varepsilon_d(p) = \left(\frac{(\boldsymbol{\varepsilon}_d\cdot\mathbf{p})}{M_d}, \boldsymbol{\varepsilon}_d + \frac{(\boldsymbol{\varepsilon}_d\cdot\mathbf{p})\mathbf{p}}{M_d(E_d(p)+M_d)}\right), \quad \varepsilon_d(p)\cdot p = 0, \quad (10)$$

is the polarization vector of the axial-vector diquark with momentum $\mathbf{p}$, $E_d(p) = \sqrt{\mathbf{p}^2+M_d^2}$, and $\varepsilon_d(0) = (0, \boldsymbol{\varepsilon}_d)$ is the polarization vector in the diquark rest frame. The effective long-range vector vertex of the diquark can be represented in the form

$$J_{d;\mu} = \begin{cases} \dfrac{(P+Q)_\mu}{2\sqrt{E_d(p)E_d(q)}}, & \text{for scalar diquark,} \\ \dfrac{(P+Q)_\mu}{2\sqrt{E_d(p)E_d(q)}} - \dfrac{i\mu_d}{2M_d}\Sigma^\nu_\mu \tilde{k}_\nu, & \text{for axial-vector diquark,} \end{cases} \tag{11}$$

where $\tilde{k} = (0, \mathbf{k})$ and we neglected the contribution of the chromoquadrupole moment of the axial-vector diquark, which is suppressed by an additional power of $k/M_d$. Here, the antisymmetric tensor

$$(\Sigma_{\rho\sigma})^\nu_\mu = -i(g_{\mu\rho}\delta^\nu_\sigma - g_{\mu\sigma}\delta^\nu_\rho), \tag{12}$$

and the axial-vector diquark spin $\mathbf{S}_d$ is given by $(S_{d;k})_{il} = -i\varepsilon_{kil}$. We choose the total chromomagnetic moment of the axial-vector diquark $\mu_d = 2$ [10, 20].

The effective long-range vector vertex of the quark is defined by [18, 19]

$$\Gamma_\mu(\mathbf{k}) = \gamma_\mu + \frac{i\kappa}{2m}\sigma_{\mu\nu}\tilde{k}^\nu, \tag{13}$$

where $\kappa$ is the Pauli interaction constant characterizing the anomalous chromomagnetic moment of quarks. In the configuration space, the vector and scalar confining potentials in the nonrelativistic limit reduce to

$$V^V_{\text{conf}}(r) = (1-\varepsilon)V_{\text{conf}}(r), \qquad V^S_{\text{conf}}(r) = \varepsilon V_{\text{conf}}(r), \tag{14}$$

with

$$V_{\text{conf}}(r) = V^S_{\text{conf}}(r) + V^V_{\text{conf}}(r) = Ar + B, \tag{15}$$

where $\varepsilon$ is the mixing coefficient.

All the parameters of our model, like quark masses, parameters of linear confining potential $A$ and $B$, mixing coefficient $\varepsilon$, and anomalous chromomagnetic quark moment $\kappa$, were fixed from the analysis of heavy-quarkonium masses [18] and radiative decays [21]. The quark masses $m_b = 4.88$ GeV, $m_c = 1.55$ GeV, $m_s = 0.50$ GeV, $m_{u,d} = 0.33$ GeV and the parameters of the linear potential $A = 0.18$ GeV$^2$ and $B = -0.30$ GeV have standard values of quark models. The value of the mixing coefficient of vector and scalar confining potentials $\varepsilon = -1$ has been determined from the consideration of the heavy-quark expansion [22] and meson radiative decays [21]. Finally, the universal Pauli interaction constant $\kappa = -1$ has been fixed from the analysis of the fine splitting of heavy-quarkonium $^3P_J$-states [18] and also from the heavy-quark expansion [22]. Note that the long-range magnetic contribution to the potential in our model is proportional to $(1+\kappa)$ and thus vanishes for the chosen value of $\kappa = -1$.

## 3. DIQUARKS IN THE RELATIVISTIC QUARK MODEL

The quark–quark interaction in the diquark consists of the sum of the spin-independent and spin-dependent parts:

$$V_{QQ}(r) = V^{\text{SI}}_{QQ}(r) + V^{\text{SD}}_{QQ}(r). \tag{16}$$

The spin-independent part with account of $v^2/c^2$ corrections including retardation effects [23] is given by

$$\begin{aligned} V^{\text{SI}}_{QQ}(r) = {} & -\frac{2}{3}\frac{\alpha_s(\mu^2)}{r} + \frac{1}{2}(Ar + B) \\ & + \frac{1}{8}\left(\frac{1}{m_1^2} + \frac{1}{m_2^2}\right)\Delta\left[-\frac{2}{3}\frac{\alpha_s(\mu^2)}{r} + \frac{1}{2}(1-\varepsilon)(1+2\kappa)Ar\right] \\ & + \frac{1}{2m_1m_2}\left\{-\frac{2}{3}\frac{\alpha_s}{r}\left[\mathbf{p}^2 + \frac{(\mathbf{p}\cdot\mathbf{r})^2}{r^2}\right]\right\}_W \\ & + \frac{1}{2}\left[\frac{1-\varepsilon}{2m_1m_2} - \frac{\varepsilon}{4}\left(\frac{1}{m_1^2} + \frac{1}{m_2^2}\right)\right] \times \left\{Ar\left[\mathbf{p}^2 - \frac{(\mathbf{p}\cdot\mathbf{r})^2}{r^2}\right]\right\}_W \\ & + \frac{1}{2}\left[\frac{1-\varepsilon}{2m_1m_2} - \frac{\varepsilon}{4}\left(\frac{1}{m_1^2} + \frac{1}{m_2^2}\right)\right]B\mathbf{p}^2, \end{aligned} \tag{17}$$

where $\{\dots\}_W$ denotes the Weyl ordering of operators and

$$\alpha_s(\mu^2) = \frac{4\pi}{\beta_0 \ln(\mu^2/\Lambda^2)} \tag{18}$$

with $\mu$ fixed to be equal to twice the reduced mass.

The spin-dependent part of the quark–quark potential can be represented in our model [18] as follows:

$$\begin{aligned} V^{\text{SD}}_{QQ}(r) = {} & a\mathbf{L}\cdot\tilde{\mathbf{S}} \\ & + b\left[\frac{3}{r^2}(\mathbf{S}_1\cdot\mathbf{r})(\mathbf{S}_2\cdot\mathbf{r}) - (\mathbf{S}_1\cdot\mathbf{S}_2)\right] \\ & + c\,\mathbf{S}_1\cdot\mathbf{S}_2 + d\,\mathbf{L}\cdot(\mathbf{S}_1 - \mathbf{S}_2), \end{aligned} \tag{19}$$

$$a = \frac{1}{m_1m_2}\left\{\left(1 + \frac{m_1^2 + m_2^2}{4m_1m_2}\right)\frac{2}{3}\frac{\alpha_s(\mu^2)}{r^3}\right. \tag{20}$$

$$-\frac{1}{2}\frac{m_1^2+m_2^2}{4m_1m_2}\frac{A}{r}+\frac{1}{2}(1+\kappa)\times\frac{(m_1+m_2)^2}{2m_1m_2}(1-\varepsilon)\frac{A}{r}\Bigg\},$$

$$b=\frac{1}{3m_1m_2}\left\{\frac{2\alpha_s(\mu^2)}{r^3}+\frac{1}{2}(1+\kappa)^2(1-\varepsilon)\frac{A}{r}\right\},$$

$$c=\frac{2}{3m_1m_2}\left\{\frac{8\pi\alpha_s(\mu^2)}{3}\delta^3(r)+\frac{1}{2}(1+\kappa)^2(1-\varepsilon)\frac{A}{r}\right\},$$

$$d=\frac{1}{m_1m_2}\left\{\frac{m_2^2-m_1^2}{4m_1m_2}\left[\frac{2}{3}\frac{\alpha_s(\mu^2)}{r^3}-\frac{1}{2}\frac{A}{r}\right]+\frac{1}{2}(1+\kappa)\frac{m_2^2-m_1^2}{2m_1m_2}(1-\varepsilon)\frac{A}{r}\right\},$$

where $\mathbf{L}$ is the orbital momentum and $\mathbf{S}_{1,2}$, $\tilde{\mathbf{S}}=\mathbf{S}_1+\mathbf{S}_2$ are the spin momenta. For $\kappa=-1$, the form of the spin-dependent potential (19) agrees with the expression based on QCD [24, 25].

Now we can calculate the mass spectra of heavy diquarks with account of all relativistic corrections (including retardation effects) of order $v^2/c^2$. For this purpose, we substitute the quasipotential, which is a sum of the spin-independent (17) and spin-dependent (19) parts, into the quasipotential Eq. (1). Then we multiply the resulting expression from the left by the quasipotential wave function of a bound state and integrate with respect to the relative momentum. Taking into account the accuracy of the calculations, we can use for the resulting matrix elements the wave functions of Eq. (1) with the static potential

$$V_{QQ}^{\rm NR}(r)=-\frac{2}{3}\frac{\alpha_s(\mu^2)}{r}+\frac{1}{2}(Ar+B). \qquad (21)$$

**Table 1.** Mass spectrum and mean squared radii of the *cc* diquark

| State | Mass, GeV | $\langle r^2\rangle^{1/2}$, fm |
|---|---|---|
| $1^3S_1$ | 3.226 | 0.56 |
| $2^3S_1$ | 3.535 | 1.02 |
| $3^3S_1$ | 3.782 | 1.37 |
| $1^1P_1$ | 3.460 | 0.82 |
| $2^1P_1$ | 3.712 | 1.22 |
| $3^1P_1$ | 3.928 | 1.54 |

As a result, we obtain the mass formula

$$\frac{b^2(M)}{2\mu_R}=W+\langle a\rangle\langle\mathbf{L}\cdot\tilde{\mathbf{S}}\rangle+\langle b\rangle\left\langle\left[\frac{3}{r^2}(\mathbf{S}_1\cdot\mathbf{r})(\mathbf{S}_2\cdot\mathbf{r})-(\mathbf{S}_1\cdot\mathbf{S}_2)\right]\right\rangle+\langle c\rangle\langle\mathbf{S}_1\cdot\mathbf{S}_2\rangle+\langle d\rangle\langle\mathbf{L}\cdot(\mathbf{S}_1-\mathbf{S}_2)\rangle, \qquad (22)$$

where

$$W=\langle V_{QQ}^{\rm SI}\rangle+\frac{\langle\mathbf{p}^2\rangle}{2\mu_R},$$

$$\langle\mathbf{L}\cdot\tilde{\mathbf{S}}\rangle=\frac{1}{2}(\tilde{J}(\tilde{J}+1)-L(L+1)-\tilde{S}(\tilde{S}+1)),$$

$$\left\langle\left[\frac{3}{r^2}(\mathbf{S}_1\cdot\mathbf{r})(\mathbf{S}_2\cdot\mathbf{r})-(\mathbf{S}_1\cdot\mathbf{S}_2)\right]\right\rangle=-\frac{6(\langle\mathbf{L}\cdot\tilde{\mathbf{S}}\rangle)^2+3\langle\mathbf{L}\cdot\tilde{\mathbf{S}}\rangle-2\tilde{S}(\tilde{S}+1)L(L+1)}{2(2L-1)(2L+3)},$$

$$\langle\mathbf{S}_1\cdot\mathbf{S}_2\rangle=\frac{1}{2}\left(\tilde{S}(\tilde{S}+1)-\frac{3}{2}\right),\qquad\tilde{\mathbf{S}}=\mathbf{S}_1+\mathbf{S}_2,$$

and $\langle a\rangle$, $\langle b\rangle$, $\langle c\rangle$, $\langle d\rangle$ are the appropriate averages over radial wave functions of Eq. (20). We use the notation for the heavy-diquark classification: $n^{2\tilde{S}+1}L_{\tilde{J}}$, where $n=1,2,\ldots$ is a radial quantum number, $L$ is the angular momentum, $\tilde{S}=0,1$ is the total spin of two heavy quarks, and $\tilde{J}=L-\tilde{S},L,L+\tilde{S}$ is the total angular momentum ($\tilde{\mathbf{J}}=\mathbf{L}+\tilde{\mathbf{S}}$), which is considered as the spin of a diquark ($\mathbf{S}_d$) in the following section. The first term on the right-hand side of the mass formula (22) contains all spin-independent contributions, the second and the last terms describe the spin–orbit interaction, the third term is responsible for the tensor interaction, and the fourth term gives the spin–spin interaction. The results of our numerical calculations of the mass spectra of *cc* and *bb*

**Table 2.** Mass spectrum and mean squared radii of the *bb* diquark

| State | Mass, GeV | $\langle r^2\rangle^{1/2}$, fm |
|---|---|---|
| $1^3S_1$ | 9.778 | 0.37 |
| $2^3S_1$ | 10.015 | 0.71 |
| $3^3S_1$ | 10.196 | 0.98 |
| $4^3S_1$ | 10.369 | 1.22 |
| $1^1P_1$ | 9.944 | 0.57 |
| $2^1P_1$ | 10.132 | 0.87 |
| $3^1P_1$ | 10.305 | 1.12 |
| $4^1P_1$ | 10.453 | 1.34 |

diquarks are presented in Tables 1 and 2. The mass of the ground state of the $bc$ diquark in the axial-vector $(1^3S_1)$ state is

$$M_{bc}^A = 6.526 \text{ GeV}$$

and in the scalar $(1^1S_1)$ state is

$$M_{bc}^S = 6.519 \text{ GeV}.$$

In order to determine the diquark interaction with the gluon field, which takes into account the diquark structure, it is necessary to calculate the corresponding matrix element of the quark current between diquark states. This diagonal matrix element can be parametrized by the following set of elastic form factors:

(a) scalar diquark ($S$):

$$\langle S(P)|J_\mu|S(Q)\rangle = H_+(k^2)(P+Q)_\mu; \qquad (23)$$

(b) axial-vector diquark ($A$):

$$\langle A(P)|J_\mu|A(Q)\rangle \qquad (24)$$
$$= -[\varepsilon_d^*(P)\cdot\varepsilon_d(Q)]H_1(k^2)(P+Q)_\mu + H_2(k^2)$$
$$\times\left\{[\varepsilon_d^*(P)\cdot Q]\varepsilon_{d;\mu}(Q) + [\varepsilon_d(Q)\cdot P]\varepsilon_{d;\mu}^*(P)\right\}$$
$$+ H_3(k^2)\frac{1}{M_V^2}[\varepsilon_d^*(P)\cdot Q][\varepsilon_d(Q)\cdot P](P+Q)_\mu,$$

where $k = P - Q$ and $\varepsilon_d(P)$ is the polarization vector of the axial-vector diquark (10).

In our quark model, we find the following relation between these diquark transition form factors in the nonrelativistic limit [26]:

$$H_+(k^2) = H_1(k^2) = H_2(k^2) = 2\mathcal{F}(\mathbf{k}^2), \qquad (25)$$
$$H_3(k^2) = 0,$$
$$\mathcal{F}(\mathbf{k}^2) = \frac{\sqrt{E_d M_d}}{E_d + M_d}\left[\int\frac{d^3p}{(2\pi)^3}\right.$$
$$\left.\times\,\bar\Psi_d\left(\mathbf{p} + \frac{2m_{Q_2}}{E_d+M_d}\mathbf{k}\right)\Psi_d(\mathbf{p}) + (1\leftrightarrow 2)\right],$$

where $\Psi_d$ are the diquark wave functions in the rest reference frame. We calculated corresponding form factors $\mathcal{F}(r)/r$, which are the Fourier transforms of $\mathcal{F}(\mathbf{k}^2)/\mathbf{k}^2$, using the diquark wave functions found by numerically solving the quasipotential equation. In Fig. 2, the functions $\mathcal{F}(r)$ for the $cc$ diquark in the $1S$, $1P$, $2S$, $2P$ states are shown as an example. We see that the slope of $\mathcal{F}(r)$ decreases with the increase in the diquark excitation. Our estimates show that this form factor can be approximated with a high accuracy by the expression

$$\mathcal{F}(r) = 1 - e^{-\xi r - \zeta r^2}, \qquad (26)$$

which agrees with previously used approximations [27]. The values of the parameters $\xi$ and $\zeta$ for different $cc$ and $bb$ diquark states are given in Tables 3 and 4. As we see, the functions $\mathcal{F}(r)$ vanish in the limit $r \to 0$ and become unity for large values of $r$. Such a behavior can be easily understood intuitively. At large distances, a diquark can be well approximated by a pointlike object and its internal structure cannot be resolved. When the distance to the diquark decreases, the internal structure plays a more important role. As the distance approaches zero, the interaction weakens and goes to zero for $r = 0$, since this point coincides with the center of gravity of the two heavy quarks forming the diquark. Thus, the function $\mathcal{F}(r)$ gives an important contribution to the short-range part of the interaction of the light quark with the heavy diquark in the baryon and can be neglected for the long-range (confining) interaction. It is important to note that the inclusion of such a function removes a fictitious singularity $1/r^3$ at the origin arising from the one-gluon-exchange part of the quark–diquark potential when the expansion in inverse powers of the heavy-quark mass is used.

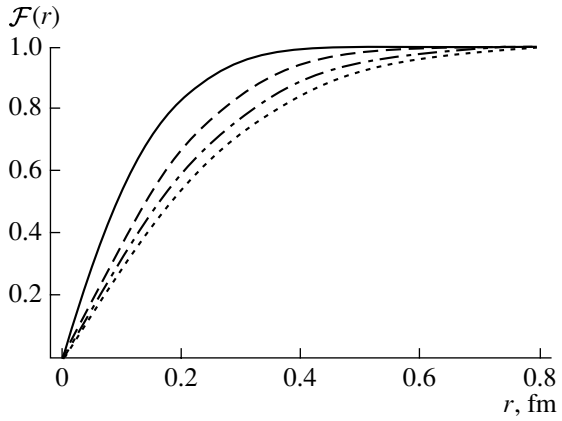


**Fig. 2.** The form factors $\mathcal{F}(r)$ for the $cc$ diquark. The solid curve is for the $1S$ state, the dashed curve for the $1P$ state, the dash-dotted curve for the $2S$ state, and the dotted curve for the $2P$ state.

**Table 3.** Parameters $\xi$ and $\zeta$ for ground and excited states of the $cc$ diquark

| State | $\xi$, GeV | $\zeta$,GeV$^2$ |
|---|---|---|
| $1S$ | 1.30 | 0.42 |
| $2S$ | 0.67 | 0.19 |
| $3S$ | 0.57 | 0.12 |
| $1P$ | 0.74 | 0.315 |
| $2P$ | 0.60 | 0.155 |
| $3P$ | 0.55 | 0.075 |

**Table 4.** Parameters $\xi$ and $\zeta$ for ground and excited states of the $bb$ diquark

| State | $\xi$, GeV | $\zeta$, GeV$^2$ |
|---|---|---|
| $1S$ | 1.30 | 1.60 |
| $2S$ | 0.85 | 0.31 |
| $3S$ | 0.66 | 0.155 |
| $4S$ | 0.56 | 0.09 |
| $1P$ | 0.90 | 0.59 |
| $2P$ | 0.65 | 0.215 |
| $3P$ | 0.58 | 0.120 |
| $4P$ | 0.51 | 0.085 |

## 4. QUASIPOTENTIAL OF THE INTERACTION OF A LIGHT QUARK WITH A HEAVY DIQUARK

The expression for the quasipotential (6) can, in principle, be used for arbitrary quark and diquark masses. The substitution of the Dirac spinors (8) and diquark form factors (23) and (24) into (6) results in an extremely nonlocal potential in the configuration space. Clearly, it is very hard to deal with such potentials without any simplifying expansion. Fortunately, in the case of the heavy-diquark–light-quark picture of the baryon, one can carry out (following HQET) the expansion in inverse powers of the heavy-diquark mass $M_d$. The leading terms then follow in the limit $M_d \to \infty$.

### *4.1. Infinitely Heavy Diquark Limit*

In the limit $M_d \to \infty$, the heavy-diquark vertices (23) and (24) have only the zeroth component, and the diquark mass and spin decouple from the consideration. As a result, we get in this limit the quasipotential for the light quark similar to the one in the heavy-light meson in the limit of an infinitely heavy antiquark [19]. The only difference consists in the extra factor $\mathcal{F}(k^2)$, defined in (25), in the one-gluon-exchange part, which accounts for the heavy-diquark structure. The quasipotential in this limit is given by

$$V(\mathbf{p},\mathbf{q};M) = \bar{u}_q(p)\Bigg\{-\frac{4}{3}\alpha_s\mathcal{F}(k^2)\frac{4\pi}{\mathbf{k}^2}\gamma_q^0 \qquad (27)$$

$$+ V^V_{\rm conf}(\mathbf{k})\left[\gamma_q^0 + \frac{\kappa}{2m_q}\gamma_q^0(\boldsymbol{\gamma}\cdot\mathbf{k})\right] + V^S_{\rm conf}(\mathbf{k})\Bigg\}u_q(q).$$

The resulting interaction is still nonlocal in configuration space. However, taking into account that doubly heavy baryons are weakly bound, we can replace $\epsilon_q(p) \to E_q = (M^2 - M_d^2 + m_q^2)/(2M)$ in the Dirac spinors (8) [19]. Such a simplifying substitution is widely used in quantum electrodynamics [28–30] and introduces only minor corrections of order of the ratio of the binding energy $\langle V\rangle$ to $E_q$. This substitution makes the Fourier transformation of the potential (27) local. In contrast with the heavy-light meson case, no special consideration of the one-gluon-exchange term is necessary, since the presence of the diquark structure described by an extra function $\mathcal{F}(k^2)$ in Eq. (27) removes the fictitious $1/r^3$ singularity at the origin in configuration space.

The resulting local quark–diquark potential for $M_d \to \infty$ can be represented in configuration space in the following form:

$$V_{M_d\to\infty}(r) = \frac{E_q + m_q}{2E_q} \qquad (28)$$

$$\times\Bigg[V_{\rm Coul}(r) + V_{\rm conf}(r) + \frac{1}{(E_q+m_q)^2}$$

$$\times\Bigg\{\mathbf{p}[V_{\rm Coul}(r) + V^V_{\rm conf}(r) - V^S_{\rm conf}(r)]\mathbf{p}$$

$$-\frac{E_q+m_q}{2m_q}\Delta V^V_{\rm conf}(r)[1-(1+\kappa)]$$

$$+\frac{2}{r}\Bigg(V'_{\rm Coul}(r) - V'^S_{\rm conf}(r) - V'^V_{\rm conf}(r)$$

$$\times\left[\frac{E_q}{m_q} - 2(1+\kappa)\frac{E_q+m_q}{2m_q}\right]\Bigg)\mathbf{l}\cdot\mathbf{S}_q\Bigg\}\Bigg],$$

where $V_{\rm Coul}(r) = -(4/3)\alpha_s\mathcal{F}(r)/r$ is the smeared Coulomb potential. The prime denotes differentiation with respect to $r$; $\mathbf{l}$ is the orbital momentum and $\mathbf{S}_q$ is the spin operator of the light quark. Note that the last term in (28) is of the same order as the first two terms and thus cannot be treated perturbatively. It is important to note that the quark–diquark potential $V_{M_d\to\infty}(r)$ almost coincides with the quark–antiquark potential in heavy-light ($B$ and $D$) mesons for $m_Q \to \infty$ [19]. The only difference is the presence of the extra factor $\mathcal{F}(r)$ in $V_{\rm Coul}(r)$, which accounts for the internal structure of the diquark. This is the consequence of the heavy-quark (diquark) limit, in which its spin and mass decouple from the consideration.

In the infinitely heavy diquark limit, the quasipotential Eq. (1) in configuration space becomes

$$\left(\frac{E_q^2 - m_q^2}{2E_q} - \frac{\mathbf{p}^2}{2E_q}\right)\Psi_B(r) = V_{M_d\to\infty}(r)\Psi_B(r), \qquad (29)$$

and the mass of the baryon is given by $M = M_d + E_q$.

Solving (29) numerically, we get the eigenvalues $E_q$ and the baryon wave functions $\Psi_B$. The obtained results are presented in Table 5. We use the notation $n_d L n_q l(j)$ for the classification of baryon states in the infinitely heavy diquark limit. Here, we first give the radial quantum number $n_d$ and the angular momentum $L$ of the heavy diquark. Then the radial quantum number $n_q$, the angular momentum $l$, and the value $j$ of the total angular momentum ($\mathbf{j} = \mathbf{l} + \mathbf{S}_q$) of the light quark are shown. We see that the heavy-diquark spin and mass decouple in the limit $M_d \to \infty$, and thus we get the number of degenerated states in accord with the heavy-quark symmetry prediction. This symmetry also predicts an almost equality of corresponding light-quark energies $E_q$ for *bbq* and *ccq* baryons and their nearness in the same limit to the light-quark energies $E_q$ of $B$ and $D$ mesons [19]. The small deviations of the baryon energies from values of the meson energies are connected with the different forms of the singularity smearing at $r = 0$ in the baryon and meson cases.

### *4.2. $1/M_d$ Corrections*

The heavy-quark symmetry degeneracy of states is broken by $1/M_d$ corrections. The corrections of order $1/M_d$ to the potential (28) arise from the spatial components of the heavy-diquark vertex. Other contributions at first order in $1/M_d$ come from the one-gluon-exchange potential and the vector confining potential, while the scalar potential gives no contribution at first order. The resulting $1/M_d$ corrections to the quark–diquark potential (28) are given by the following expressions:

(a) scalar diquark:

$$\delta V_{1/M_d}(r) = \frac{1}{E_q M_d}\Bigg\{\mathbf{p}\left[V_{\rm Coul}(r) + V^V_{\rm conf}(r)\right]\mathbf{p} + V'_{\rm Coul}(r)\frac{\mathbf{l}^2}{2r} - \frac{1}{4}\Delta V^V_{\rm conf}(r) + \left[\frac{1}{r}V'_{\rm Coul}(r) + \frac{1+\kappa}{r}V'^V_{\rm conf}(r)\right]\mathbf{l}\cdot\mathbf{S}_q\Bigg\}; \quad (30)$$

(b) axial-vector diquark:

$$\delta V_{1/M_d}(r) = \frac{1}{E_q M_d}\Bigg\{\mathbf{p}\left[V_{\rm Coul}(r) + V^V_{\rm conf}(r)\right]\mathbf{p} + V'_{\rm Coul}(r)\frac{\mathbf{l}^2}{2r} - \frac{1}{4}\Delta V^V_{\rm conf}(r) + \left[\frac{1}{r}V'_{\rm Coul}(r) + \frac{1+\kappa}{r}V'^V_{\rm conf}(r)\right]\mathbf{l}\cdot\mathbf{S}_q + \frac{1}{2}\left[\frac{1}{r}V'_{\rm Coul}(r) + \frac{1+\kappa}{r}V'^V_{\rm conf}(r)\right]\mathbf{l}\cdot\mathbf{S}_d + \frac{1}{3}\Bigg(\frac{1}{r}V'_{\rm Coul}(r) - V''_{\rm Coul}(r) + (1+\kappa) \times \left[\frac{1}{r}V'^V_{\rm conf}(r) - V''^V_{\rm conf}(r)\right]\Bigg) \times \left[-\mathbf{S}_q\cdot\mathbf{S}_d + \frac{3}{r^2}(\mathbf{S}_q\cdot\mathbf{r})(\mathbf{S}_d\cdot\mathbf{r})\right] + \frac{2}{3}\left[\Delta V_{\rm Coul}(r) + (1+\kappa)\Delta V^V_{\rm conf}(r)\right]\mathbf{S}_d\cdot\mathbf{S}_q\Bigg\}, \quad (31)$$

where $\mathbf{S} = \mathbf{S}_q + \mathbf{S}_d$ being the total spin, $\mathbf{S}_d$ being the diquark spin (which is equal to the total angular momentum $\tilde{\mathbf{J}}$ of two heavy quarks forming the diquark). The first three terms in (31) represent spin-independent corrections, the fourth and the fifth terms are responsible for the spin–orbit interaction, the sixth one is the tensor interaction, and the last one is the spin–spin interaction. It is necessary to note that the confining vector interaction gives a contribution to the spin-dependent part, which is proportional to $(1+\kappa)$. Thus, it vanishes for the chosen value of $\kappa = -1$, while the confining vector contribution to the spin-independent part is nonzero.

**Table 5.** The values of $E_q$ in the limit $M_d \to \infty$ (in GeV)

| Baryon state | $E_q$ ($ccq$) | $E_s$ ($ccs$) | $E_q$ ($bbq$) | $E_s$ ($bbs$) |
|---|---|---|---|---|
| $1S1s(1/2)$ | 0.491 | 0.638 | 0.492 | 0.641 |
| $1S1p(3/2)$ | 0.788 | 0.906 | 0.785 | 0.904 |
| $1S1p(1/2)$ | 0.877 | 0.968 | 0.880 | 0.969 |
| $1S2s(1/2)$ | 0.987 | 1.080 | 0.993 | 1.084 |
| $1P1s(1/2)$ | 0.484 | 0.633 | 0.489 | 0.636 |
| $1P1p(3/2)$ | 0.793 | 0.909 | 0.789 | 0.906 |
| $1P1p(1/2)$ | 0.873 | 0.965 | 0.876 | 0.967 |
| $1P2s(1/2)$ | 0.980 | 1.075 | 0.984 | 1.078 |
| $2S1s(1/2)$ | 0.481 | 0.631 | 0.486 | 0.634 |
| $2S1p(3/2)$ | 0.794 | 0.909 | 0.791 | 0.908 |
| $2S1p(1/2)$ | 0.871 | 0.963 | 0.874 | 0.965 |
| $2S2s(1/2)$ | 0.979 | 1.074 | 0.982 | 1.076 |
| $2P1s(1/2)$ | 0.479 | 0.630 | 0.481 | 0.631 |
| $3S1s(1/2)$ | 0.478 | 0.630 | 0.480 | 0.630 |

In order to estimate the matrix elements of spin-dependent terms in the $1/M_d$ corrections to the quark–diquark potential (30) and (31), as well as different mixings of baryon states, it is convenient to

**Table 6.** Mass spectrum of $\Xi_{cc}$ baryons

| State $(n_dLn_ql)J^P$ | Mass, GeV | | State $(n_dLn_ql)J^P$ | Mass, GeV | |
|---|---|---|---|---|---|
| | this work | [11] | | this work | [11] |
| $(1S1s)1/2^+$ | 3.620 | 3.478 | $(1P1s)1/2^-$ | 3.838 | 3.702 |
| $(1S1s)3/2^+$ | 3.727 | 3.61 | $(1P1s)3/2^-$ | 3.959 | 3.834 |
| $(1S1p)1/2^-$ | 4.053 | 3.927 | $(2S1s)1/2^+$ | 3.910 | 3.812 |
| $(1S1p)3/2^-$ | 4.101 | 4.039 | $(2S1s)3/2^+$ | 4.027 | 3.944 |
| $(1S1p)1/2'^-$ | 4.136 | 4.052 | $(2P1s)1/2^-$ | 4.085 | 3.972 |
| $(1S1p)5/2^-$ | 4.155 | 4.047 | $(2P1s)3/2^-$ | 4.197 | 4.104 |
| $(1S1p)3/2'^-$ | 4.196 | 4.034 | $(3S1s)1/2^+$ | 4.154 | 4.072 |

**Table 7.** Mass spectrum of $\Xi_{bb}$ baryons

| State $(n_dLn_ql)J^P$ | Mass, GeV | | State $(n_dLn_ql)J^P$ | Mass, GeV | |
|---|---|---|---|---|---|
| | this work | [11] | | this work | [11] |
| $(1S1s)1/2^+$ | 10.202 | 10.093 | $(2S1s)1/2^+$ | 10.441 | 10.373 |
| $(1S1s)3/2^+$ | 10.237 | 10.133 | $(2S1s)3/2^+$ | 10.482 | 10.413 |
| $(1S1p)1/2^-$ | 10.632 | 10.541 | $(2S1p)1/2^-$ | 10.873 | |
| $(1S1p)3/2^-$ | 10.647 | 10.567 | $(2S1p)3/2^-$ | 10.888 | |
| $(1S1p)5/2^-$ | 10.661 | 10.580 | $(2S1p)1/2'^-$ | 10.902 | |
| $(1S1p)1/2'^-$ | 10.675 | 10.578 | $(2S1p)5/2^-$ | 10.905 | |
| $(1S1p)3/2'^-$ | 10.694 | 10.581 | $(2S1p)3/2'^-$ | 10.920 | |
| $(1S2s)1/2^+$ | 10.832 | | $(2P1s)1/2^-$ | 10.563 | 10.493 |
| $(1S2s)3/2^+$ | 10.860 | | $(2P1s)3/2^-$ | 10.607 | 10.533 |
| $(1P1s)1/2^-$ | 10.368 | 10.310 | $(3S1s)1/2^+$ | 10.630 | 10.563 |
| $(1P1s)3/2^-$ | 10.408 | 10.343 | $(3S1s)3/2^+$ | 10.673 | |
| $(1P1p)1/2^+$ | 10.763 | | $(3P1s)1/2^-$ | 10.744 | |
| $(1P1p)3/2^+$ | 10.779 | | $(3P1s)3/2^-$ | 10.788 | |
| $(1P1p)5/2^+$ | 10.786 | | $(4S1s)1/2^+$ | 10.812 | |
| $(1P1p)1/2'^+$ | 10.838 | | $(4S1s)3/2^+$ | 10.856 | |
| $(1P1p)3/2'^+$ | 10.856 | | $(4P1s)1/2^-$ | 10.900 | |

use the following relations:

$$|J,j\rangle = \sum_S (-1)^{J+l+S_d+S_q} \times \sqrt{(2S+1)(2j+1)} \begin{Bmatrix} S_d & S_q & S \\ l & J & j \end{Bmatrix} |J,S\rangle \qquad (32)$$

and

$$|J,j\rangle = \sum_{J_d} (-1)^{J+l+S_d+S_q} \times \sqrt{(2J_d+1)(2j+1)} \begin{Bmatrix} S_d & l & J_d \\ S_q & J & j \end{Bmatrix} |J,J_d\rangle, \qquad (33)$$

where $\mathbf{J} = \mathbf{j} + \mathbf{S}_d$ is the baryon total angular mo-

**Table 8.** Mass spectrum of $\Omega_{cc}$ baryons

| State $(n_dLn_ql)J^P$ | Mass, GeV | | State $(n_dLn_ql)J^P$ | Mass, GeV | |
|---|---|---|---|---|---|
| | this work | [32] | | this work | [32] |
| $(1S1s)1/2^+$ | 3.778 | 3.594 | $(1P1s)1/2^-$ | 4.002 | 3.812 |
| $(1S1s)3/2^+$ | 3.872 | 3.730 | $(1P1s)3/2^-$ | 4.102 | 3.949 |
| $(1S1p)1/2^-$ | 4.208 | 4.050 | $(2S1s)1/2^+$ | 4.075 | 3.925 |
| $(1S1p)3/2^-$ | 4.252 | 4.102 | $(2S1s)3/2^+$ | 4.174 | 4.064 |
| $(1S1p)1/2'^-$ | 4.271 | 4.145 | $(2P1s)1/2^-$ | 4.251 | 4.073 |
| $(1S1p)5/2^-$ | 4.303 | 4.134 | $(2P1s)3/2^-$ | 4.345 | 4.213 |
| $(1S1p)3/2'^-$ | 4.325 | 4.176 | $(3S1s)1/2^+$ | 4.321 | 4.172 |

**Table 9.** Mass spectrum of $\Omega_{bb}$ baryons

| State $(n_dLn_ql)J^P$ | Mass, GeV | | State $(n_dLn_ql)J^P$ | Mass, GeV | |
|---|---|---|---|---|---|
| | this work | [32] | | this work | [32] |
| $(1S1s)1/2^+$ | 10.359 | 10.210 | $(2S1s)1/2^+$ | 10.610 | 10.493 |
| $(1S1s)3/2^+$ | 10.389 | 10.257 | $(2S1s)3/2^+$ | 10.645 | 10.540 |
| $(1S1p)1/2^-$ | 10.771 | 10.651 | $(2S1p)1/2^-$ | 11.011 | |
| $(1S1p)3/2^-$ | 10.785 | 10.661 | $(2S1p)3/2^-$ | 11.025 | |
| $(1S1p)5/2^-$ | 10.798 | 10.670 | $(2S1p)1/2'^-$ | 11.035 | |
| $(1S1p)1/2'^-$ | 10.804 | 10.700 | $(2S1p)5/2^-$ | 11.040 | |
| $(1S1p)3/2'^-$ | 10.821 | 10.720 | $(2S1p)3/2'^-$ | 11.051 | |
| $(1S2s)1/2^+$ | 10.970 | | $(2P1s)1/2^-$ | 10.738 | 10.617 |
| $(1S2s)3/2^+$ | 10.992 | | $(2P1s)3/2^-$ | 10.775 | 10.663 |
| $(1P1s)1/2^-$ | 10.532 | 10.416 | $(3S1s)1/2^+$ | 10.806 | 10.682 |
| $(1P1s)3/2^-$ | 10.566 | 10.462 | $(3S1s)3/2^+$ | 10.843 | |
| $(1P1p)1/2^+$ | 10.914 | | $(3P1s)1/2^-$ | 10.924 | |
| $(1P1p)3/2^+$ | 10.928 | | $(3P1s)3/2^-$ | 10.961 | |
| $(1P1p)5/2^+$ | 10.937 | | $(4S1s)1/2^+$ | 10.994 | |
| $(1P1p)1/2'^+$ | 10.971 | | $(4S1s)3/2^+$ | 11.031 | |
| $(1P1p)3/2'^+$ | 10.986 | | $(4P1s)1/2^-$ | 11.083 | |

mentum, $\mathbf{j} = \mathbf{l} + \mathbf{S}_q$ is the light-quark total angular momentum, $\mathbf{S} = \mathbf{S}_q + \mathbf{S}_d$ is the baryon total spin, and $\mathbf{J}_d = \mathbf{l} + \mathbf{S}_d$.

## 5. MASS SPECTRA OF DOUBLY HEAVY BARYONS

For the description of the quantum numbers of baryons, we use the notation $(n_dLn_ql)J^P$, where we first show the radial quantum number of the diquark ($n_d = 1, 2, 3, \dots$) and its orbital momentum by a capital letter ($L = S, P, D, \dots$), then the radial quantum number of the light quark ($n_q = 1, 2, 3, \dots$) and its orbital momentum by a lowercase letter ($l = s, p, d, \dots$), and at the end the total angular momentum $J$ and parity $P$ of the baryon.

The presence of the spin–orbit interaction proportional to $\mathbf{l} \cdot \mathbf{S}_d$ and of the tensor interaction in the quark–diquark potential at $1/M_d$ order (31) results in a mixing of states, which have the same total angular momentum $J$ and parity, but different light-quark total momentum $j$. For example, the baryon states with a diquark in the ground state and light quark in the $p$ wave $(1S1p)$ for $J = 1/2$ or $3/2$ have different values

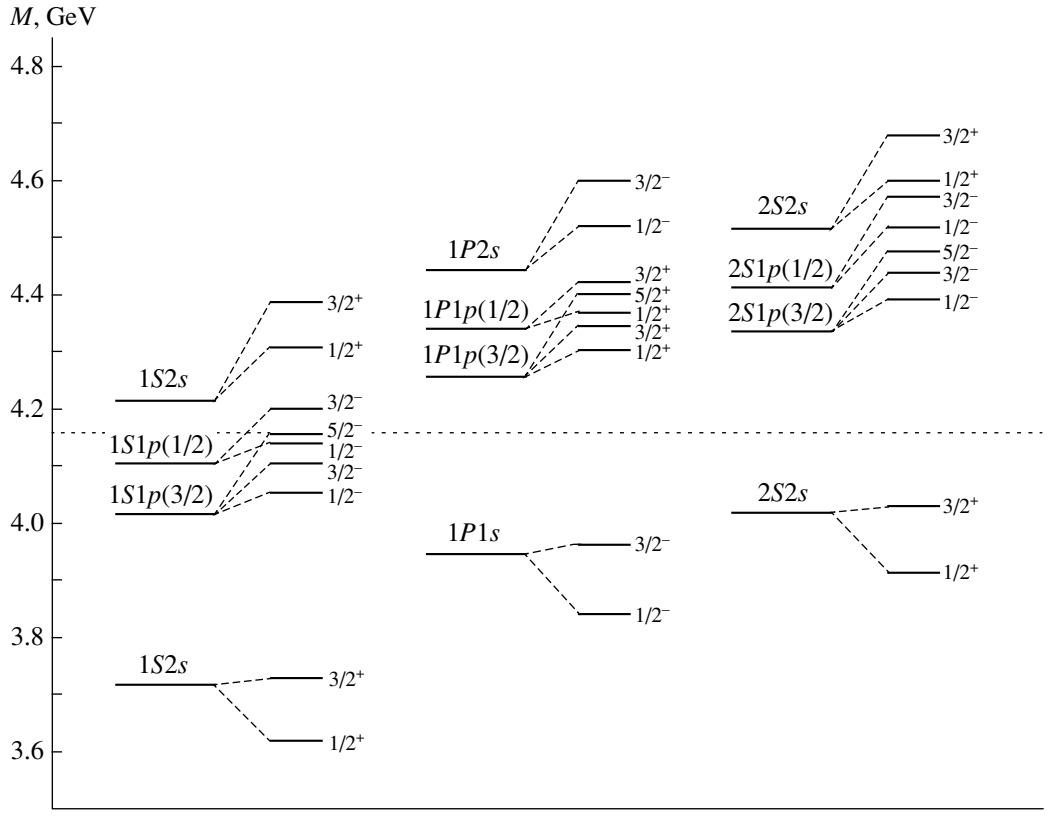


**Fig. 3.** Mass spectrum of $\Xi_{cc}$ baryons. The horizontal dashed line shows the $\Lambda_c D$ threshold.

of the light-quark angular momentum $j = 1/2$ and $3/2$, which mix between themselves. In the case of the *ccq* baryon, we have the mixing matrix for $J = 1/2$

$$\begin{pmatrix} -55.6 & -7.3 \\ -8.5 & -37.9 \end{pmatrix} \text{[MeV]} \tag{34}$$

with the following eigenvectors:

$$|(1S1p)1/2^{'-}\rangle = -0.334|j = 3/2\rangle + 0.943|j = 1/2\rangle, \tag{35}$$

$$|(1S1p)1/2^{-}\rangle = 0.925|j = 3/2\rangle + 0.380|j = 1/2\rangle.$$

For the *ccq* baryon with $J = 3/2$, the mixing matrix is given by

$$\begin{pmatrix} -23.0 & 18.1 \\ 21.3 & 18.9 \end{pmatrix} \text{[MeV]}, \tag{36}$$

and the eigenvectors are equal to

$$|(1S1p)3/2^{'-}\rangle = 0.343|j = 3/2\rangle + 0.939|j = 1/2\rangle, \tag{37}$$

$$|(1S1p)3/2^{-}\rangle = 0.919|j = 3/2\rangle - 0.394|j = 1/2\rangle.$$

For the *bbq* baryon, we get the mixing matrix for $J = 1/2$

$$\begin{pmatrix} -18.0 & -2.4 \\ -2.8 & -12.6 \end{pmatrix} \text{[MeV]}, \tag{38}$$

which has eigenvalues

$$|(1S1p)1/2^{'-}\rangle = 0.349|j = 3/2\rangle + 0.937|j = 1/2\rangle, \tag{39}$$

$$|(1S1p)1/2^{-}\rangle = 0.915|j = 3/2\rangle + 0.402|j = 1/2\rangle,$$

and for $J = 3/2$, the mixing matrix is

$$\begin{pmatrix} -7.4 & 5.9 \\ 7.1 & 6.3 \end{pmatrix} \text{[MeV]}, \tag{40}$$

so that

$$|(1S1p)3/2^{'-}\rangle = 0.341|j = 3/2\rangle + 0.940|j = 1/2\rangle, \tag{41}$$

$$|(1S1p)3/2^{-}\rangle = 0.917|j = 3/2\rangle + 0.400|j = 1/2\rangle.$$

The quasipotential with $1/m_Q$ corrections is given by the sum of $V_{m_Q\to\infty}(r)$ from (28) and $\delta V_{1/m_Q}(r)$ from (30) and (31). By substituting it into the

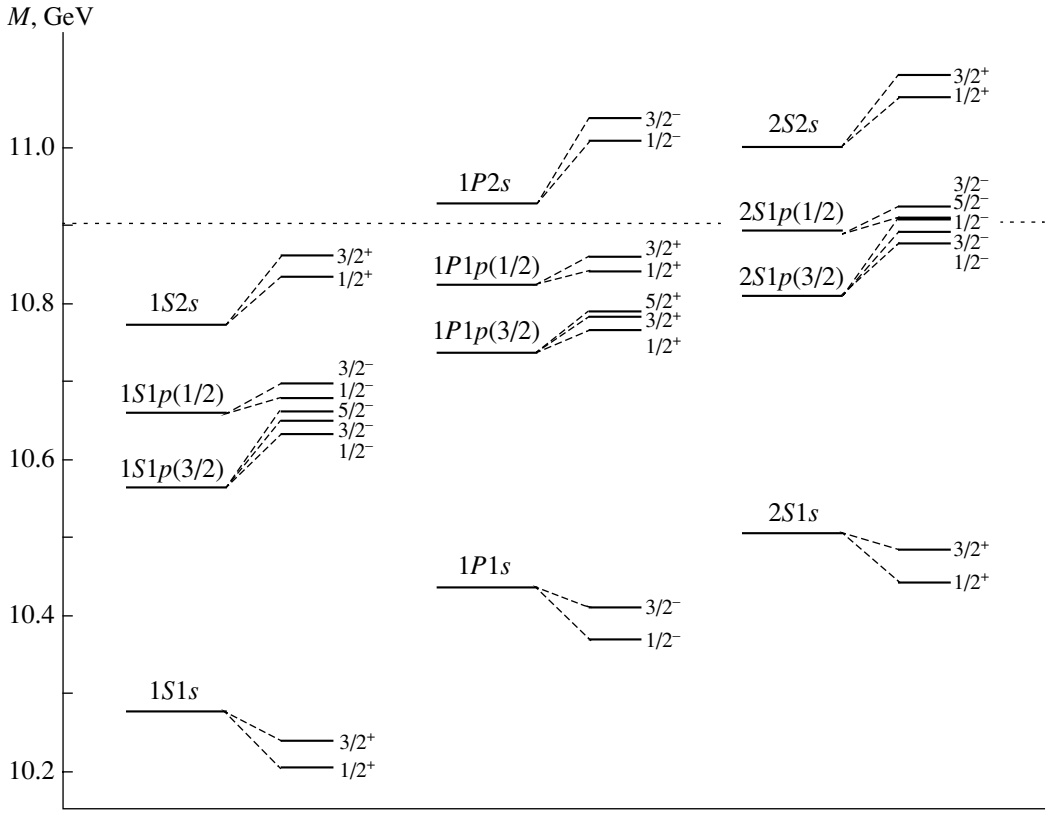


**Fig. 4.** Mass spectrum of $\Xi_{bb}$ baryons. The horizontal dashed line shows the $\Lambda_b D$ threshold.

quasipotential Eq. (1) and treating the $1/m_Q$ correction term $\delta V_{1/m_Q}(r)$ using perturbation theory, we are now able to calculate the mass spectra of $\Xi_{cc}$, $\Xi_{bb}$, $\Xi_{cb}$, $\Omega_{cc}$, $\Omega_{bb}$, $\Omega_{cb}$ baryons with account of $1/m_Q$ corrections. In Tables 6–9, we present mass spectra of ground and excited states of doubly heavy baryons containing both heavy quarks of the same flavor ($c$ and $b$). The corresponding level orderings are schematically shown in Figs. 3–6. In these figures, we first show our predictions for doubly heavy baryon spectra in the limit when all $1/M_d$ corrections are neglected (denoting baryon states by $n_d L n_q l(j)$). We see that, in this limit, the $p$-wave excitations of the light quark are inverted. This means that the mass of the state with higher angular momentum $j = 3/2$ is smaller than the mass of the state with lower angular momentum $j = 1/2$ [13, 31]. The similar $p$-level inversion was found previously in the mass spectra of heavy-light mesons in the infinitely heavy-quark limit [19]. Note that the pattern of levels of the light quark and level separation in doubly heavy baryons and heavy-light mesons almost coincide in these limits. Next, we switch on $1/M_d$ corrections. This results in splitting of the degenerate states and mixing of states with different $j$, which have the same total angular momentum $J$ and parity, as was discussed above. Since the diquark has spin 1, the states with $j = 1/2$ split into two different states with $J = 1/2$ or $3/2$, while the states with $j = 3/2$ split into three different states with $J = 1/2$, $3/2$, or $5/2$. The fine splitting between $p$ levels turns out to be of the same order of magnitude as the gap between $j = 1/2$ and $j = 3/2$ degenerate multiplets in the infinitely heavy diquark limit. The inclusion of $1/M_d$ corrections leads also to the relative shifts of the baryon levels, further decreasing this gap. As a result, some of the $p$ levels from different (initially degenerate) multiplets overlap; however, the heavy-diquark spin-averaged centers remain inverted. The resulting picture for the diquark in the ground state is very similar to the one for heavy-light mesons [19]. The purely inverted pattern of $p$ levels is observed only for the $B$ meson and $\Xi_{bb}$, $\Omega_{bb}$ baryons, while in other heavy-light mesons ($D$, $D_s$, $B_s$) and doubly heavy baryons ($\Xi_{cc}$, $\Omega_{cc}$) $p$ levels from different $j$ multiplets overlap. The absence of the $p$-level overlap for the $\Omega_{bb}$ baryon in contrast to the $B_s$ meson (where we predict a very small overlap of these levels [19]) is explained by the fact that the ratio $m_s/M_{bb}^d$ is approximately two times smaller than $m_s/m_b$ and thus it is of order $m_q/m_b$. As was argued in [19], these ratios determine the applicability of the heavy-quark limit.

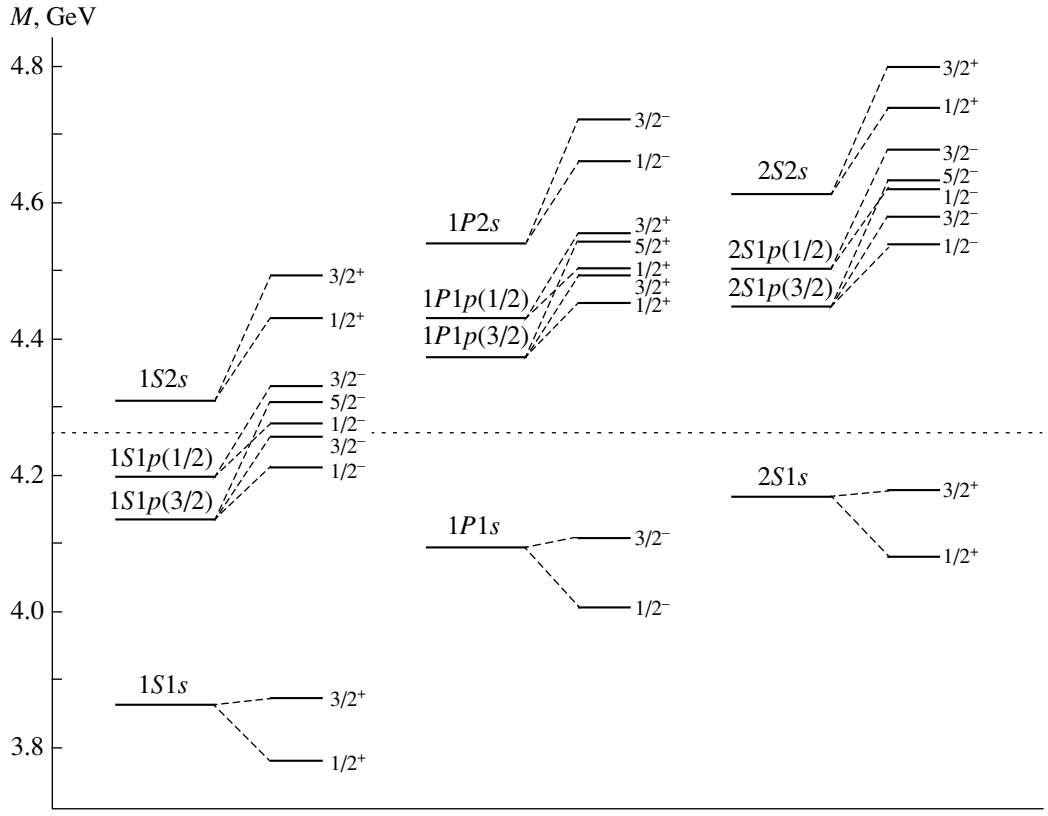


**Fig. 5.** Mass spectrum of $\Omega_{cc}$ baryons. The horizontal dashed line shows the $\Lambda_c D_s$ threshold.

In Tables 6–9, we compare our predictions for the ground- and excited-state masses of $\Xi_{cc}$, $\Xi_{bb}$ and $\Omega_{cc}$, $\Omega_{bb}$ baryons with the results of [11, 32]. As we see from these tables, our predictions are approximately 50–150 MeV higher than the estimates of [11, 32]. One of the reasons for the difference between these two predictions for the masses of $\Xi_{cc}$ and $\Omega_{cc}$ baryons (which is the largest) is the difference in the $c$-quark masses. The mass of the $c$ quark in [11] is determined from fitting the charmonium spectrum in the quark model, where all spin-independent relativistic corrections were ignored. However, our estimates show that, due to the rather large average value of $v^2/c^2$ in charmonium,[4)] such corrections play an important role and give contributions to the charmonium masses of order of 100 MeV. As a result, the $c$-quark mass found in [11] is approximately 70 MeV less than in our model. For the calculation of the diquark masses, we also take into account the spin-independent corrections (17) to the $QQ$ potential. We find that their contribution is less than that in the case of charmonium, since $V_{QQ} = V_{Q\bar{Q}}/2$. Thus, the $cc$-diquark masses in [11] are approximately 50 MeV smaller than in our model. The other main source of the difference is the expansion in inverse powers of the light-quark mass, which was used in [11, 32] but is not applied in our approach, where the light quark is treated fully relativistically.

In Table 10, we compare our model predictions for the ground-state masses of doubly heavy baryons with some other predictions [6, 9, 11, 33] as well as our previous prediction [10], where the expansion in inverse powers of the heavy- and light-quark masses was used. In general, we find reasonable agreement within 100 MeV between different predictions [6, 9–11, 33] for the ground-state masses of the doubly heavy baryons. The main advantage of our present approach is the completely relativistic treatment of the light quark and account of the nonlocal composite structure of the diquark. The authors of [33] do not exploit the assumption about the quark–diquark structure of the baryon. The consideration in [9] is based on some semiempirical regularities of the baryon mass spectra and the Feynman–Hellmann theorem. The predictions of [6] are obtained within the nonrelativistic quark model.

For the $\Xi_{cb}$ and $\Omega_{cb}$ baryons, containing heavy quarks of different flavors ($c$ and $b$), we calculate only the ground-state masses. As was argued in [11], the excited states of heavy diquarks, composed of the

[4)] The spin-dependent relativistic corrections, which are of the same order in $v^2/c^2$, produce level splittings of the order of 120 MeV.

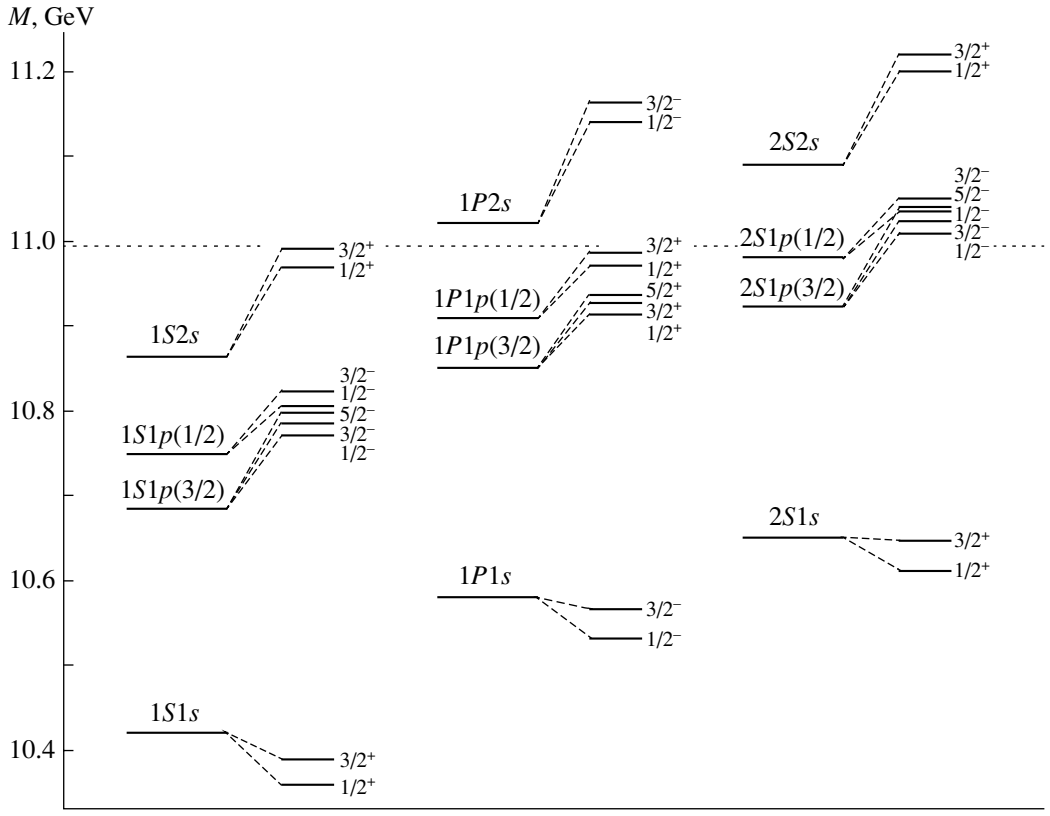


**Fig. 6.** Mass spectrum of $\Omega_{bb}$ baryons. The horizontal dashed line shows the $\Lambda_b D_s$ threshold.

quarks with different flavors, are unstable under the emission of soft gluons, and thus the calculation of the excited baryon (*cbq* and *cbs*) masses is not justified in the quark–diquark scheme. We get the following predictions for the masses of the ground-state *cbq* baryons:

$(1S1s)1/2^+$ states with the axial-vector and scalar *cb* diquarks, respectively:

$$M(\Xi_{cb}) = 6.933 \text{ GeV}, \quad M(\Xi'_{cb}) = 6.963 \text{ GeV};$$

$(1S1s)3/2^+$ state:

$$M(\Xi^*_{cb}) = 6.980 \text{ GeV};$$

and for *cbs* baryons:

$(1S1s)1/2^+$ states with the axial-vector and scalar *cb* diquarks, respectively:

$$M(\Omega_{cb}) = 7.088 \text{ GeV}, \quad M(\Omega'_{cb}) = 7.116 \text{ GeV},$$

$(1S1s)3/2^+$ state:

$$M(\Omega^*_{cb}) = 7.130 \text{ GeV}.$$

Now we compare our results with the model-independent predictions of HQET. The heavy-quark symmetry predicts simple relations between the spin-averaged masses of doubly heavy baryons with the accuracy of order $1/M_d$:

$$\Delta\bar{M}_{1,2} \equiv \bar{M}_1(\Xi_{bb}) - \bar{M}_1(\Xi_{cc}) \qquad (42)$$
$$= \bar{M}_2(\Xi_{bb}) - \bar{M}_2(\Xi_{cc}) = \bar{M}_1(\Omega_{bb}) - \bar{M}_1(\Omega_{cc})$$
$$= \bar{M}_2(\Omega_{bb}) - \bar{M}_2(\Omega_{cc}) = M^d_{bb} - M^d_{cc} \equiv \Delta M^d,$$

where the spin-averaged masses are

$$\bar{M}_1 = (M_{1/2} + 2M_{3/2})/3,$$
$$\bar{M}_2 = (M_{1/2} + 2M_{3/2} + 3M_{5/2})/6,$$

$M_J$ are the masses of baryons with total angular momentum $J$, and $M^d_{QQ}$ are the masses of diquarks in definite states. The numerical results are presented in Table 11 (only the states below threshold are considered). We see that the equalities in Eq. (42) are satisfied with good accuracy for the baryons in the ground and excited states.

It follows from the heavy-quark symmetry that the hyperfine mass splittings of initially degenerate light-quark states

$$\Delta M(\Xi_{QQ}) \equiv M_{3/2}(\Xi_{QQ}) - M_{1/2}(\Xi_{QQ}), \qquad (43)$$
$$\Delta M(\Omega_{QQ}) \equiv M_{3/2}(\Omega_{QQ}) - M_{1/2}(\Omega_{QQ})$$

**Table 10.** Comparison of different predictions for mass spectra of ground states of doubly heavy baryons (in GeV) ($\{QQ\}$ denotes the diquark in the axial-vector state and $[QQ]$ denotes the diquark in the scalar state)

| Baryon | Quark content | $J^P$ | This work | [11] | [10] | [9] | [6] | [33] |
|---|---|---|---|---|---|---|---|---|
| $\Xi_{cc}$ | $\{cc\}q$ | $1/2^+$ | 3.620 | 3.478 | 3.66 | 3.66 | 3.61 | 3.69 |
| $\Xi^*_{cc}$ | $\{cc\}q$ | $3/2^+$ | 3.727 | 3.61 | 3.81 | 3.74 | 3.68 | |
| $\Omega_{cc}$ | $\{cc\}s$ | $1/2^+$ | 3.778 | 3.59 | 3.76 | 3.74 | 3.71 | 3.86 |
| $\Omega^*_{cc}$ | $\{cc\}s$ | $3/2^+$ | 3.872 | 3.69 | 3.89 | 3.82 | 3.76 | |
| $\Xi_{bb}$ | $\{bb\}q$ | $1/2^+$ | 10.202 | 10.093 | 10.23 | 10.34 | | 10.16 |
| $\Xi^*_{bb}$ | $\{bb\}q$ | $3/2^+$ | 10.237 | 10.133 | 10.28 | 10.37 | | |
| $\Omega_{bb}$ | $\{bb\}s$ | $1/2^+$ | 10.359 | 10.18 | 10.32 | 10.37 | | 10.34 |
| $\Omega^*_{bb}$ | $\{bb\}s$ | $3/2^+$ | 10.389 | 10.20 | 10.36 | 10.40 | | |
| $\Xi_{cb}$ | $\{cb\}q$ | $1/2^+$ | 6.933 | 6.82 | 6.95 | 7.04 | | 6.96 |
| $\Xi'_{cb}$ | $[cb]q$ | $1/2^+$ | 6.963 | 6.85 | 7.00 | 6.99 | | |
| $\Xi^*_{cb}$ | $\{cb\}q$ | $3/2^+$ | 6.980 | 6.90 | 7.02 | 7.06 | | |
| $\Omega_{cb}$ | $\{cb\}s$ | $1/2^+$ | 7.088 | 6.91 | 7.05 | 7.09 | | 7.13 |
| $\Omega'_{cb}$ | $[cb]s$ | $1/2^+$ | 7.116 | 6.93 | 7.09 | 7.06 | | |
| $\Omega^*_{cb}$ | $\{cb\}s$ | $3/2^+$ | 7.130 | 6.99 | 7.11 | 7.12 | | |

should scale with the diquark masses:

$$\Delta M(\Xi_{cc}) = R\Delta M(\Xi_{bb}), \qquad (44)$$
$$\Delta M(\Omega_{cc}) = R\Delta M(\Omega_{bb}),$$

where $R = M^d_{bb}/M^d_{cc}$ is the ratio of diquark masses. Our model predictions for these splittings are displayed in Table 12. Again, we see that heavy-quark symmetry relations are satisfied with high accuracy.

The close similarity of the interaction of the light quark with the heavy quark in the heavy-light mesons and with the heavy diquark in the doubly heavy baryons produces very simple relations between the meson and baryon mass splittings [7, 34, 35]. In fact,

**Table 11.** Differences between spin-averaged masses of doubly heavy baryons defined in Eq. (42) (in GeV)

| Baryon state | $\Delta\bar{M}_1(\Xi)$ | $\Delta\bar{M}_2(\Xi)$ | $\Delta\bar{M}_1(\Omega)$ | $\Delta\bar{M}_2(\Omega)$ | $\Delta M^d$ |
|---|---|---|---|---|---|
| $1S1s$ | 6.534 | | 6.538 | | 6.552 |
| $1S1p$ | 6.512 | 6.532 | | 6.519 | 6.552 |
| $1P1s$ | 6.476 | | 6.486 | | 6.484 |
| $2S1s$ | 6.480 | | 6.492 | | 6.480 |

for the ground-state hyperfine splittings of mesons and baryons, we obtain in the approximation $M^d_{QQ} \cong 2m_Q$

$$\Delta M(\Xi_{QQ}) = \frac{3}{2}\frac{m_Q}{M^d_{QQ}}\Delta M_{B,D} \cong \frac{3}{4}\Delta M_{B,D}, \qquad (45)$$
$$\Delta M(\Omega_{QQ}) \cong \frac{3}{4}\Delta M_{B_s,D_s}, \qquad Q = b, c,$$

where the factor $3/2$ is just the ratio of the baryon and meson spin matrix elements. The numerical fulfillment of relations (45) is shown in Table 13.

## 6. HEAVY-DIQUARK TRANSITION FORM FACTORS

Now we can apply the calculated masses and wave functions of heavy diquarks and doubly heavy baryons for studying their semileptonic decays. We use the two-step picture schematically shown in Fig. 1. First, we consider the heavy-diquark weak transition and, then, include the light quark in consideration.

The form factors of the subprocess $d(QQ_s) \to d'(Q'Q_s)e\bar{\nu}$, where one heavy quark $Q_s$ is a spectator, are determined by the weak decay of the active heavy

**Table 12.** Hyperfine splittings of the doubly heavy baryons for the states with the light-quark angular momentum $j = 1/2$ (in MeV)

| Baryon state | $R$ | $\Delta M(\Xi_{bb})$ | $R\Delta M(\Xi_{bb})$ | $\Delta M(\Xi_{cc})$ | $\Delta M(\Omega_{bb})$ | $R\Delta M(\Omega_{bb})$ | $\Delta M(\Omega_{cc})$ |
|---|---|---|---|---|---|---|---|
| $1S1s[3/2-1/2]$ | 3.03 | 35 | 106 | 107 | 30 | 91 | 94 |
| $1S1p[3/2-1/2]$ | 3.03 | 19 | 58 | 60 | 17 | 52 | 54 |
| $1P1s[3/2-1/2]$ | 2.87 | 40 | 115 | 121 | 34 | 98 | 100 |
| $2S1s[3/2-1/2]$ | 2.83 | 41 | 116 | 117 | 35 | 99 | 99 |

**Table 13.** Comparison of hyperfine splittings (in MeV) in doubly heavy baryons and heavy-light mesons (experimental values for the hyperfine splittings in mesons are taken from [36])

| $\Delta M(\Xi_{cc})$ | $\frac{3}{4}\Delta M_D^{\rm exp}$ | $\Delta M(\Xi_{bb})$ | $\frac{3}{4}\Delta M_B^{\rm exp}$ | $\Delta M(\Omega_{cc})$ | $\frac{3}{4}\Delta M_{D_s}^{\rm exp}$ | $\Delta M(\Omega_{bb})$ | $\frac{3}{4}\Delta M_{B_s}^{\rm exp}$ |
|---|---|---|---|---|---|---|---|
| 107 | 106 | 35 | 34 | 94 | 108 | 30 | 35 |

quark $Q \to Q' e\bar{\nu}$. The local effective Hamiltonian is given by

$$H_{\rm eff} = \frac{G_{\rm F}}{\sqrt{2}} V_{QQ'} \left(\bar{Q}'\gamma_\mu(1-\gamma_5)Q\right)\left(\bar{l}\gamma_\mu(1-\gamma_5)\nu_e\right), \tag{46}$$

where $G_{\rm F}$ is the Fermi constant and $V_{QQ'}$ is the CKM matrix element. In the relativistic quark model, the transition matrix element of the weak current between two diquark states is determined by the contraction of the wave functions $\Psi_d$ of the initial and final diquarks with the two-particle vertex function $\Gamma$,

$$\langle d'(Q)|J_\mu^W|d(P)\rangle \tag{47}$$
$$= \int \frac{d^3p\,d^3q}{(2\pi)^6} \bar{\Psi}^{\lambda\sigma}_{d',Q}(\mathbf{q})\Gamma_\mu^{\lambda\sigma,\rho\omega}(\mathbf{p},\mathbf{q})\Psi^{\rho\omega}_{d,P}(\mathbf{p}),$$
$$\lambda,\sigma,\rho,\omega = \pm\frac{1}{2}.$$

Here, we denote mass, energy, and four-velocity of the initial diquark ($Q_bQ_s$, index $b$ stands for the initial active quark and index $s$ for spectator) by $M_i$, $E_i = M_i v^0$, and $v = P/M_i$, and mass, energy, and four-velocity of the final diquark ($Q_aQ_s$, index $a$ denotes the final active quark) by $M_f$, $E_f = M_f v'^0$, and $v' = Q/M_f$.

The leading contribution to the vertex function $\Gamma_\mu$ comes from the diagram in Fig. 7 [22, 37, 38] (we explicitly show spin indices):

$$\Gamma_\mu^{\lambda\sigma,\rho\omega}(\mathbf{p},\mathbf{q}) = \Gamma_\mu^{(1)} = \bar{u}_a^\lambda(\mathbf{q}_1)\gamma_\mu(1-\gamma_5) \tag{48}$$
$$\times\, u_b^\rho(\mathbf{p}_1)\bar{u}_s^\sigma(\mathbf{q}_2)u_s^\omega(\mathbf{p}_2)(2\pi)^3\delta(\mathbf{p}_2-\mathbf{q}_2)\delta^{\sigma\omega}.$$

Relativistic four-momenta of the particles in the initial and final states are defined as follows:

$$p_{1,2} = \epsilon_{1,2}(p)v \pm \sum_{i=1}^{3} n^{(i)}(v)p^i, \tag{49}$$
$$v = \frac{P}{\mathcal{M}_i}, \quad \mathcal{M}_i = \epsilon_1(p) + \epsilon_2(p),$$
$$q_{1,2} = \epsilon_{1,2}(q)v' \pm \sum_{i=1}^{3} n^{(i)}(v')q^i,$$
$$v' = \frac{Q}{\mathcal{M}_f}, \quad \mathcal{M}_f = \epsilon_1(q) + \epsilon_2(q),$$

and $n^{(i)}$ are three four-vectors defined by

$$n^{(i)}(v) = \left\{v^i, \delta^{ij} + \frac{1}{1+v^0}v^iv^j\right\}.$$

After making necessary transformations, the expression for $\Gamma$ should be continued in $\mathcal{M}_i$ and $\mathcal{M}_f$ to the values of initial $M_i$ and final $M_f$ bound-state masses.

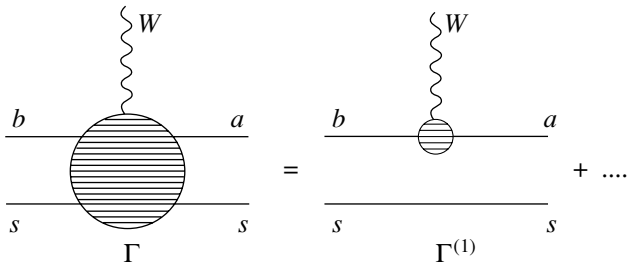


**Fig. 7.** The leading-order contribution $\Gamma^{(1)}$ to the diquark vertex function $\Gamma$.

The transformation of the bound-state wave functions from the rest frame to the moving one with four-momenta $P$ and $Q$ is given by [22, 37]

$$\Psi^{\rho\omega}_{d,P}(\mathbf{p}) = D^{1/2,\rho\alpha}_b(R^W_{L_P})D^{1/2,\omega\beta}_s(R^W_{L_P})\Psi^{\alpha\beta}_{d,0}(\mathbf{p}), \tag{50}$$

$$\bar\Psi^{\lambda\sigma}_{d',Q}(\mathbf{q}) = \bar\Psi^{\varepsilon\tau}_{d',0}(\mathbf{q})D^{+1/2,\,\varepsilon\lambda}_a(R^W_{L_Q})D^{+1/2,\tau\sigma}_s(R^W_{L_Q}),$$

where $R^W$ is the Wigner rotation, $L_P$ is the Lorentz boost from the diquark rest frame to a moving one, and the rotation matrix $D^{1/2}(R)$ is defined by

$$\begin{pmatrix}1&0\\0&1\end{pmatrix}D^{1/2}_{b,s}(R^W_{L_P}) = S^{-1}(\mathbf{p}_{1,2})S(\mathbf{P})S(\mathbf{p}), \tag{51}$$

where

$$S(\mathbf{p}) = \sqrt{\frac{\epsilon(p)+m}{2m}}\left(1+\frac{(\boldsymbol{\alpha}\cdot\mathbf{p})}{\epsilon(p)+m}\right)$$

is the usual Lorentz transformation matrix of the four-spinor.

Using relations (49)–(51), we can express the matrix element (47) in the form of the trace over spinor indices of both particles. As a result, we get the covariant (see [37]) expression for the transition matrix element:

$$\langle d'(Q)|J^W_\mu|d(P)\rangle = 2\sqrt{M_iM_f}\int\frac{d^3p\,d^3q}{(2\pi)^3} \tag{52}$$
$$\times\,\mathrm{Tr}\{\bar\Psi_{d'}(Q,q)\gamma_\mu(1-\gamma_5)\Psi_d(P,p)\}\delta^3(\mathbf{p}_2-\mathbf{q}_2),$$

where the amplitudes $\Psi_d$ for the scalar ($S$) and axial-vector ($A$) diquarks ($d$) are given by

$$\Psi_S(P,p) = \sqrt{\frac{\epsilon_b(p)+m_b}{2\epsilon_b(p)}} \tag{53}$$
$$\times\sqrt{\frac{\epsilon_s(p)+m_s}{2\epsilon_s(p)}}\left[\frac{\hat v+1}{2\sqrt2}+\frac{\hat v-1}{2\sqrt2}\right.$$
$$\times\frac{\tilde p^2}{(\epsilon_b(p)+m_b)(\epsilon_s(p)+m_s)}-\left(\frac{\hat v+1}{2\sqrt2}\right.$$
$$\left.\left.\times\frac{1}{\epsilon_s(p)+m_s}+\frac{\hat v-1}{2\sqrt2}\frac{1}{\epsilon_b(p)+m_b}\right)\hat{\tilde p}\right]\gamma_0\Phi_S(p),$$

$$\Psi_A(P,p) = \sqrt{\frac{\epsilon_b(p)+m_b}{2\epsilon_b(p)}}\sqrt{\frac{\epsilon_s(p)+m_s}{2\epsilon_s(p)}} \tag{54}$$
$$\times\left[\frac{\hat v+1}{2\sqrt2}\hat\varepsilon+\frac{\hat v-1}{2\sqrt2}\frac{\tilde p^2}{(\epsilon_b(p)+m_b)(\epsilon_s(p)+m_s)}\hat\varepsilon\right.$$
$$-\frac{\hat v-1}{2\sqrt2}\frac{2(\varepsilon\cdot\tilde p)\hat{\tilde p}}{(\epsilon_b(p)+m_b)(\epsilon_s(p)+m_s)}+\frac{\hat v+1}{2\sqrt2}$$
$$\left.\times\frac{\hat\varepsilon\hat{\tilde p}}{\epsilon_s(p)+m_s}-\frac{\hat v-1}{2\sqrt2}\frac{\hat{\tilde p}\hat\varepsilon}{\epsilon_b(p)+m_b}\right]\gamma_0\gamma^5\Phi_A(p).$$

Here, $\Phi_d(p)\equiv\Psi_{d,0}(\mathbf{p})/\sqrt{2M_d}$ is the diquark wave function in the rest frame, normalized to unity. The four-vector

$$\tilde p = L_P(0,\mathbf{p}) = \left(\frac{(\mathbf{p}\cdot\mathbf{P})}{M},\mathbf{p}+\frac{(\mathbf{p}\cdot\mathbf{P})\mathbf{P}}{M(E+M)}\right) \tag{55}$$

has the following properties:

$$\tilde p^2 = -\mathbf{p}^2, \tag{56}$$
$$(\varepsilon\cdot\tilde p) = -(\boldsymbol{\varepsilon}\cdot\mathbf{p}),$$
$$(\tilde p\cdot v) = 0.$$

The presence of $\delta^3(\mathbf{p}_2-\mathbf{q}_2)$, with momenta $\mathbf{p}_2$ and $\mathbf{q}_2$ given by Eq. (49), in the decay matrix element (52) leads to the following additional relations valid up to the terms of order $1/m_Q$:

$$\tilde p_\mu = \frac{w\epsilon_s(p)-\epsilon_s(q)}{w^2-1}(wv_\mu-v'_\mu) \tag{57}$$
$$=\frac{\sqrt{\mathbf{p}^2}}{\sqrt{w^2-1}}(wv_\mu-v'_\mu),$$

$$\tilde q_\mu = \frac{w\epsilon_s(q)-\epsilon_s(p)}{w^2-1}(wv'_\mu-v_\mu) \tag{58}$$
$$=\frac{\sqrt{\mathbf{q}^2}}{\sqrt{w^2-1}}(wv'_\mu-v_\mu),$$

$$\epsilon_s(q) = w\epsilon_s(p)-\sqrt{w^2-1}\sqrt{\mathbf{p}^2}, \tag{59}$$
$$\mathbf{q}^2 = \left(\sqrt{w^2-1}\epsilon_s(p)-w\sqrt{\mathbf{p}^2}\right)^2,$$

$$\epsilon_s(p) = w\epsilon_s(q)-\sqrt{w^2-1}\sqrt{\mathbf{q}^2}, \tag{60}$$
$$\mathbf{p}^2 = \left(\sqrt{w^2-1}\epsilon_s(q)-w\sqrt{\mathbf{q}^2}\right)^2,$$

which allow one to express either $\mathbf{q}$ through $\mathbf{p}$ or $\mathbf{p}$ through $\mathbf{q}$. The argument of the $\delta$ function can then be rewritten as

$$\mathbf{p}_2-\mathbf{q}_2 = \mathbf{q}-\mathbf{p}-\frac{\epsilon_s(p)+\epsilon_s(q)}{w+1}(\mathbf{v}'-\mathbf{v}), \tag{61}$$

where $w=(v\cdot v')$.

Calculating traces in Eq. (52) and using relations (57)–(60), one can see that the spectator-quark contribution factors out in all decay matrix elements. Indeed, all transition matrix elements have a common factor

$$\sqrt{\frac{\epsilon_s(p)+m_s}{2\epsilon_s(p)}}\sqrt{\frac{\epsilon_s(q)+m_s}{2\epsilon_s(q)}} \tag{62}$$
$$\times\left[1-\sqrt{\frac{w-1}{w+1}}\left(\frac{\sqrt{\mathbf{p}^2}}{\epsilon_s(p)+m_s}+\frac{\sqrt{\mathbf{q}^2}}{\epsilon_s(q)+m_s}\right)\right.$$

$$+\frac{\sqrt{\mathbf{p}^2}\sqrt{\mathbf{q}^2}}{[\epsilon_s(q)+m_s][\epsilon_s(p)+m_s]}\Bigg] = \sqrt{\frac{2}{w+1}}I_s(p,q).$$

If the $\delta$ function is used to express $\mathbf{q}$ through $\mathbf{p}$ or $\mathbf{p}$ through $\mathbf{q}$, then $I_s(p,q)=\mathcal{I}_s(p)$ or $I_s(p,q)=\mathcal{I}_s(q)$, respectively, with

$$\mathcal{I}_s(p) = \sqrt{\frac{w\epsilon_s(p)-\sqrt{w^2-1}\sqrt{\mathbf{p}^2}}{\epsilon_s(p)}} \tag{63}$$
$$\times\,\theta\left(\sqrt{\epsilon_s(p)-m_s}-\sqrt{\frac{w-1}{w+1}}\sqrt{\epsilon_s(p)+m_s}\right)$$
$$+\frac{m_s}{\sqrt{\epsilon_s(p)[w\epsilon_s(p)-\sqrt{w^2-1}\sqrt{\mathbf{p}^2}]}}$$
$$\times\,\theta\left(\sqrt{\frac{w-1}{w+1}}\sqrt{\epsilon_s(p)+m_s}-\sqrt{\epsilon_s(p)-m_s}\right).$$

The weak-current matrix elements have the following covariant decomposition:

(a) scalar-to-scalar diquark transition ($bc\to bc$)

$$\frac{\langle S_f(v')|J^V_\mu|S_i(v)\rangle}{\sqrt{M_{S_i}M_{S_f}}} = h_+(w)(v+v')_\mu + h_-(w)(v-v')_\mu; \tag{64}$$

(b) scalar-to-axial-vector diquark transition ($bc\to cc$)

$$\frac{\langle A(v',\varepsilon')|J^V_\mu|S(v)\rangle}{\sqrt{M_AM_S}} = ih_V(w)\epsilon_{\mu\alpha\beta\gamma}\varepsilon'^{*\alpha}v'^\beta v^\gamma, \tag{65}$$

$$\frac{\langle A(v',\varepsilon')|J^A_\mu|S(v)\rangle}{\sqrt{M_AM_S}} = h_{A_1}(w)(w+1)\varepsilon'^*_\mu - h_{A_2}(w)(v\cdot\varepsilon'^*)v_\mu - h_{A_3}(w)(v\cdot\varepsilon'^*)v'_\mu; \tag{66}$$

(c) axial-vector-to-scalar diquark transition ($bb\to bc$)

$$\frac{\langle S(v')|J^V_\mu|A(v,\varepsilon)\rangle}{\sqrt{M_AM_S}} = ih_V(w)\epsilon_{\mu\alpha\beta\gamma}\varepsilon^\alpha v'^\beta v^\gamma, \tag{67}$$

$$\frac{\langle S(v')|J^A_\mu|A(v,\varepsilon)\rangle}{\sqrt{M_AM_S}} = h_{A_1}(w)(w+1)\varepsilon_\mu - \tilde{h}_{A_2}(w)(v'\cdot\varepsilon)v'_\mu - \tilde{h}_{A_3}(w)(v\cdot\varepsilon')v_\mu; \tag{68}$$

(d) axial-vector-to-axial-vector diquark transition ($bb\to bc$, $bc\to cc$)

$$\frac{\langle A_f(v',\varepsilon'))|J^V_\mu|A_i(v,\varepsilon)\rangle}{\sqrt{M_{A_i}M_{A_f}}} \tag{69}$$
$$= -(\varepsilon'^*\cdot\varepsilon)[h_1(w)(v+v')_\mu + h_2(w)(v-v')_\mu] + h_3(w)(v\cdot\varepsilon'^*)\varepsilon_\mu + h_4(w)(v'\cdot\varepsilon)\varepsilon'^*_\mu - (v\cdot\varepsilon'^*)(v'\cdot\varepsilon)[h_5(w)v_\mu + h_6(w)v'_\mu],$$

$$\frac{\langle A_f(v',\varepsilon'))|J^A_\mu|A_i(v,\varepsilon)\rangle}{\sqrt{M_{A_i}M_{A_f}}} = i\epsilon_{\mu\alpha\beta\gamma} \tag{70}$$
$$\times\left\{\varepsilon^\beta\varepsilon'^{*\gamma}[h_7(w)(v+v')^\alpha + h_8(w)(v-v')^\alpha] + v'^\beta v^\gamma[h_9(w)(v\cdot\varepsilon'^*)\varepsilon^\alpha + h_{10}(w)(v'\cdot\varepsilon)\varepsilon'^{*\alpha}]\right\}.$$

The transition form factors are expressed through the overlap integrals of the diquark wave functions:

$$h_+(w) = h_1(w) = h_7(w) \tag{71}$$
$$= \sqrt{\frac{2}{w+1}}\int\frac{d^3p\,d^3q}{(2\pi)^3}\bar{\Phi}_{d'}(\mathbf{q}) \times\sqrt{\frac{\epsilon_a(q)+m_a}{2\epsilon_a(q)}}\sqrt{\frac{\epsilon_b(p)+m_b}{2\epsilon_b(p)}} \times\left(1-\frac{\sqrt{\mathbf{p}^2}\sqrt{\mathbf{q}^2}}{[\epsilon_a(q)+m_a][\epsilon_b(p)+m_b]}\right)\Phi_d(\mathbf{p}) \times I_s(p,q)\delta^3\left(\mathbf{p}-\mathbf{q}+\frac{\epsilon_s(p)+\epsilon_s(q)}{w+1}(\mathbf{v}'-\mathbf{v})\right),$$

$$h_V(w) = h_3(w) = h_4(w) \tag{72}$$
$$= \sqrt{\frac{2}{w+1}}\int\frac{d^3p\,d^3q}{(2\pi)^3}\bar{\Phi}_{d'}(\mathbf{q}) \times\sqrt{\frac{\epsilon_a(q)+m_a}{2\epsilon_a(q)}}\sqrt{\frac{\epsilon_b(p)+m_b}{2\epsilon_b(p)}} \times\left[1+\sqrt{\frac{w+1}{w-1}}\left(\frac{\sqrt{\mathbf{q}^2}}{\epsilon_a(q)+m_a}+\frac{\sqrt{\mathbf{p}^2}}{\epsilon_b(p)+m_b}\right) +\frac{\sqrt{\mathbf{p}^2}\sqrt{\mathbf{q}^2}}{[\epsilon_a(q)+m_a][\epsilon_b(p)+m_b]}\right]\Phi_d(\mathbf{p}) \times I_s(p,q)\delta^3\left(\mathbf{p}-\mathbf{q}+\frac{\epsilon_s(p)+\epsilon_s(q)}{w+1}(\mathbf{v}'-\mathbf{v})\right).$$

Explicit formulas for other form factors are given in [39].

If we consider the spectator quark to be light and then take the limit of an infinitely heavy active-quark mass, $m_{a,b}\to\infty$, we can explicitly obtain heavy-quark symmetry relations for the decay matrix elements of heavy-light diquarks, which are analogous to those of heavy-light mesons [22, 40]:

$$h_+(w) = h_V(w) = h_{A_1}(w) = h_{A_3}(w) = \tilde{h}_{A_3}(w) = h_1(w) = h_3(w) = h_4(w) = h_7(w) = \xi(w),$$
$$h_-(w) = h_{A_2}(w) = \tilde{h}_{A_2}(w) = h_2(w) = h_5(w) = h_6(w) = h_8(w) = h_9(w) = h_{10}(w) = 0,$$

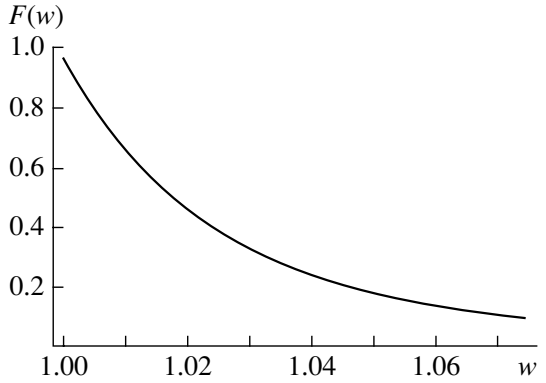


**Fig. 8.** The function $F(w)$ for the $bb \to bc$ quark transition.

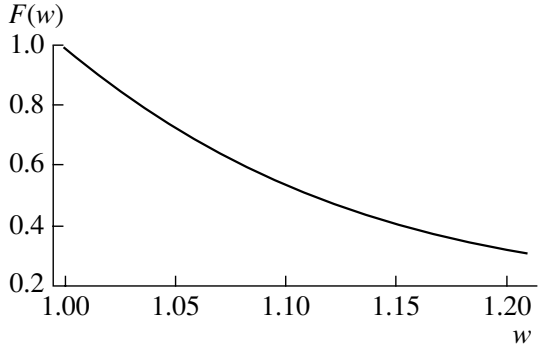


**Fig. 9.** The function $F(w)$ for the $bc \to cc$ quark transition.

with the Isgur–Wise function

$$\xi(w) = \sqrt{\frac{2}{w+1}} \int \frac{d^3p\, d^3q}{(2\pi)^3} \bar{\Phi}_f(\mathbf{q}) \Phi_i(\mathbf{p}) \qquad (73)$$
$$\times\, I_s(p,q)\delta^3\left(\mathbf{p} - \mathbf{q} + \frac{\epsilon_s(p) + \epsilon_s(q)}{w+1}(\mathbf{v}' - \mathbf{v})\right).$$

The diquark transition matrix element should be multiplied by a factor of 2 if either the initial or final diquark is composed of two identical heavy quarks.

For the heavy-diquark system, we can now apply the $v/c$ expansion. First, we perform the integration over $q$ in the form factors (71), (72) and use relations (59). Then, applying the nonrelativistic limit, we get the following expressions for the form factors:

(a) scalar-to-scalar diquark transition:

$$h_+(w) = F(w), \qquad (74)$$
$$h_-(w) = -(w+1)f(w)F(w);$$

(b) scalar-to-axial-vector diquark transition:

$$h_V(w) = [1 + (w+1)f(w)]F(w), \qquad (75)$$
$$h_{A_1}(w) = h_{A_3}(w) = [1 + (w-1)f(w)]F(w),$$
$$h_{A_2}(w) = -2f(w)F(w);$$

(c) axial-vector-to-scalar diquark transition:

$$h_V(w) = \tilde{h}_{A_3}(w) = [1 + (w+1)f(w)]F(w), \quad (76)$$
$$h_{A_1}(w) = [1 + (w-1)f(w)]F(w),$$
$$\tilde{h}_{A_2}(w) = 0;$$

(d) axial-vector-to-axial-vector diquark transition:

$$h_1(w) = h_7(w) = F(w), \qquad (77)$$
$$h_2(w) = h_8(w) = -(w+1)f(w)F(w),$$
$$h_3(w) = h_4(w) = (1 + (w+1)f(w))F(w),$$
$$h_5(w) = h_9(w) = 2f(w)F(w),$$
$$h_6(w) = h_{10}(w) = 0,$$

where

$$F(w) = \sqrt{\frac{1}{w(w+1)}} \qquad (78)$$
$$\times \left(1 + \frac{m_a}{\sqrt{m_a^2 + (w^2-1)m_s^2}}\right)^{1/2}$$
$$\times \int \frac{d^3p}{(2\pi)^3} \bar{\Phi}_f\left(\mathbf{p} + \frac{2m_s}{w+1}(\mathbf{v}' - \mathbf{v})\right) \Phi_i(\mathbf{p})$$

and

$$f(w) = \frac{m_s}{\sqrt{m_a^2 + (w^2-1)m_s^2} + m_a}. \qquad (79)$$

The appearance of the terms proportional to the function $f(w)$ is the result of the account of the spectator-quark recoil. Their contribution is important and distinguishes our approach from previous considerations [41, 42]. We plot the function $F(w)$ for $bb \to bc$ and $bc \to cc$ diquark transitions in Figs. 8 and 9.

## 7. DOUBLY-HEAVY-BARYON TRANSITIONS

The second step in studying weak transitions of doubly heavy baryons is the inclusion of the spectator light quark in the consideration. We carry out all further calculations in the limit of the infinitely heavy-diquark mass, $M_d \to \infty$, treating the light quark relativistically. The transition matrix element between doubly heavy baryon states in the quark–diquark approximation (see Figs. 1 and 7) is given by [cf. Eqs. (47) and (48)]

$$\langle B'(Q)|J_\mu^W|B(P)\rangle \qquad (80)$$
$$= 2\sqrt{M_i M_f} \int \frac{d^3p\, d^3q}{(2\pi)^3} \bar{\Psi}_{B',Q}(\mathbf{q})$$
$$\times \langle d'(Q)|J_\mu^W|d(P)\rangle \Psi_{B,P}(\mathbf{p})\delta^3(\mathbf{p}_q - \mathbf{q}_q),$$

where $\Psi_{B,P}(p)$ is the doubly heavy baryon wave function; $\mathbf{p}$ and $\mathbf{q}$ are the relative quark–diquark momenta; and $\mathbf{p}_q$ and $\mathbf{q}_q$ are light-quark momenta, expressed in the form similar to (49).

The baryon ground-state wave function $\Psi_{B,P}(p)$ is a product of the spin-independent part $\Psi_B(p)$, satisfying the related quasipotential Eq. (1), and the spin part $U_B(v)$:

$$\Psi_{B,P}(\mathbf{p}) = \Psi_B(\mathbf{p})U_B(v). \tag{81}$$

The baryon spin wave function is constructed from the Dirac spinor $u_q(v)$ of the light spectator quark and the diquark wave function. The ground-state spin-1/2 baryons can contain either the scalar or axial-vector diquark. The former baryon is denoted by $\Xi'_{QQ'}$ and the latter one by $\Xi_{QQ'}$. The ground-state spin-3/2 baryon can be formed only from the axial-vector diquark and is denoted by $\Xi^*_{QQ'}$. To obtain the corresponding baryon spin states, we use in the baryon matrix elements the following replacements:

$$u_q(v) \to U_{\Xi'_{QQ'}}(v), \tag{82}$$

$$[\varepsilon_\mu(v)u_q(v)]_{\text{spin-1/2}} \to \frac{i}{\sqrt{3}}(\gamma_\mu + v_\mu)\gamma_5 U_{\Xi_{QQ'}}(v),$$

$$[\varepsilon^\mu(v)u_q(v)]_{\text{spin-3/2}} \to U^\mu_{\Xi^*_{QQ'}}(v),$$

where baryon spinor wave functions are normalized by $U^*_B U_B = 1\,(B = \Xi', \Xi)$ and the Rarita–Schwinger wave functions are normalized by $U^{*\mu}_{\Xi^*}U_{\Xi^*\mu} = -1$.

Then the decay amplitudes of doubly heavy baryons in the infinitely heavy diquark limit are given by the following expressions.

(a) $\Xi'_{QQ_s} \to \Xi'_{Q'Q_s}$ transition:

$$\frac{\langle \Xi'_{Q'Q_s}(v')|J^W_\mu|\Xi'_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} \tag{83}$$
$$= [h_+(w)(v+v')_\mu + h_-(w)(v-v')_\mu]\bar U_{\Xi'_{Q'Q_s}}(v')$$
$$\times U_{\Xi'_{QQ_s}}(v)\eta(w);$$

(b) $\Xi'_{QQ_s} \to \Xi_{Q'Q_s}$ and $\Xi_{QQ_s} \to \Xi'_{Q'Q_s}$ transitions:

$$\frac{\langle \Xi_{Q'Q_s}(v')|J^W_\mu|\Xi'_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} = \frac{i}{\sqrt{3}} \tag{84}$$
$$\times [ih_V(w)\epsilon_{\mu\alpha\beta\gamma}v'^\beta v^\gamma - g_{\mu\alpha}h_{A_1}(w+1)$$
$$+ v_\mu v_\alpha h_{A_2}(w) + v'_\mu v_\alpha h_{A_3}(w)]\bar U_{\Xi_{Q'Q_s}}(v')$$
$$\times \gamma_5(\gamma^\alpha + v'^\alpha)U_{\Xi'_{QQ_s}}(v)\eta(w),$$

$$\frac{\langle \Xi'_{Q'Q_s}(v')|J^W_\mu|\Xi_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} = \frac{i}{\sqrt{3}} \tag{85}$$
$$\times [ih_V(w)\epsilon_{\mu\alpha\beta\gamma}v'^\beta v^\gamma - g_{\mu\alpha}h_{A_1}(w+1)$$
$$+ v'_\mu v'_\alpha \tilde h_{A_2}(w) + v_\mu v'_\alpha \tilde h_{A_3}(w)]\bar U_{\Xi'_{Q'Q_s}}(v')$$
$$\times (\gamma^\alpha + v^\alpha)\gamma_5 U_{\Xi_{QQ_s}}(v)\eta(w);$$

(c) $\Xi'_{QQ_s} \to \Xi^*_{Q'Q_s}$ and $\Xi^*_{QQ_s} \to \Xi'_{Q'Q_s}$ transitions:

$$\frac{\langle \Xi^*_{Q'Q_s}(v')|J^W_\mu|\Xi'_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} \tag{86}$$
$$= [ih_V(w)\epsilon_{\mu\alpha\beta\gamma}v'^\beta v^\gamma - g_{\mu\alpha}h_{A_1}(w+1)$$
$$+ v_\mu v_\alpha h_{A_2}(w) + v'_\mu v_\alpha h_{A_3}(w)]\bar U^\alpha_{\Xi^*_{Q'Q_s}}(v')$$
$$\times U_{\Xi'_{QQ_s}}(v)\eta(w),$$

$$\frac{\langle \Xi'_{Q'Q_s}(v')|J^W_\mu|\Xi^*_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} \tag{87}$$
$$= [ih_V(w)\epsilon_{\mu\alpha\beta\gamma}v'^\beta v^\gamma - g_{\mu\alpha}h_{A_1}(w+1)$$
$$+ v'_\mu v'_\alpha \tilde h_{A_2}(w) + v_\mu v'_\alpha \tilde h_{A_3}(w)]\bar U_{\Xi'_{Q'Q_s}}(v')$$
$$\times U^\alpha_{\Xi^*_{QQ_s}}(v)\eta(w);$$

(d) $\Xi_{QQ_s} \to \Xi_{Q'Q_s}$ transition:

$$\frac{\langle \Xi_{Q'Q_s}(v')|J^W_\mu|\Xi_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} = -\frac{1}{3} \tag{88}$$
$$\times \Big\{ g_{\rho\lambda}[h_1(w)(v+v')_\mu + h_2(w)(v-v')_\mu]$$
$$- h_3(w)g_{\mu\rho}v_\lambda - h_4(w)g_{\mu\lambda}v'_\rho + v'_\rho v_\lambda[h_5(w)v_\mu$$
$$+ h_6(w)v'_\mu] + i\epsilon_{\mu\alpha\beta\gamma}\Big(g^\beta_\rho g^\gamma_\lambda[h_7(w)(v+v')^\alpha$$
$$+ h_8(w)(v-v')^\alpha] + v'^\beta v^\gamma[h_9(w)g^\alpha_\rho v_\lambda$$
$$+ h_{10}(w)g^\alpha_\lambda v'_\rho]\Big)\Big\}\bar U_{\Xi_{Q'Q_s}}(v')\gamma_5(\gamma^\lambda + v'^\lambda)$$
$$\times (\gamma^\rho + v^\rho)\gamma_5 U_{\Xi_{QQ_s}}(v)\eta(w);$$

(e) $\Xi_{QQ_s} \to \Xi^*_{Q'Q_s}$ and $\Xi^*_{QQ_s} \to \Xi_{Q'Q_s}$ transitions:

$$\frac{\langle \Xi^*_{Q'Q_s}(v')|J^W_\mu|\Xi_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} = -\frac{i}{\sqrt{3}} \tag{89}$$
$$\times \Big\{ g_{\rho\lambda}[h_1(w)(v+v')_\mu + h_2(w)(v-v')_\mu]$$
$$- h_3(w)g_{\mu\rho}v_\lambda - h_4(w)g_{\mu\lambda}v'_\rho + v'_\rho v_\lambda[h_5(w)v_\mu$$
$$+ h_6(w)v'_\mu] + i\epsilon_{\mu\alpha\beta\gamma}\Big(g^\beta_\rho g^\gamma_\lambda[h_7(w)(v+v')^\alpha$$
$$+ h_8(w)(v-v')^\alpha] + v'^\beta v^\gamma[h_9(w)g^\alpha_\rho v_\lambda$$
$$+ h_{10}(w)g^\alpha_\lambda v'_\rho]\Big)\Big\}\bar U^\lambda_{\Xi^*_{Q'Q_s}}(v')$$

$$\times (\gamma^\rho + v^\rho)\gamma_5 U_{\Xi_{QQ_s}}(v)\eta(w),$$

$$\frac{\langle \Xi_{Q'Q_s}(v')|J_\mu^W|\Xi^*_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} = -\frac{i}{\sqrt{3}} \tag{90}$$

$$\times \Big\{ g_{\rho\lambda}[h_1(w)(v+v')_\mu + h_2(w)(v-v')_\mu]$$

$$- h_3(w) g_{\mu\rho} v_\lambda - h_4(w) g_{\mu\lambda} v'_\rho + v'_\rho v_\lambda [h_5(w) v_\mu$$

$$+ h_6(w) v'_\mu] + i\epsilon_{\mu\alpha\beta\gamma}\Big( g^\beta_\rho g^\gamma_\lambda [h_7(w)(v+v')^\alpha$$

$$+ h_8(w)(v-v')^\alpha] + v'^\beta v^\gamma [h_9(w) g^\alpha_\rho v_\lambda$$

$$+ h_{10}(w) g^\alpha_\lambda v'_\rho]\Big)\Big\} \bar{U}_{\Xi_{Q'Q_s}}(v')$$

$$\times \gamma_5 (\gamma^\lambda + v'^\lambda) U^\rho_{\Xi^*_{QQ_s}}(v)\eta(w);$$

(f) $\Xi^*_{QQ_s} \to \Xi^*_{Q'Q_s}$ transition:

$$\frac{\langle \Xi^*_{Q'Q_s}(v')|J_\mu^W|\Xi^*_{QQ_s}(v)\rangle}{2\sqrt{M_i M_f}} \tag{91}$$

$$= -\Big\{ g_{\rho\lambda}[h_1(w)(v+v')_\mu + h_2(w)(v-v')_\mu]$$

$$- h_3(w) g_{\mu\rho} v_\lambda - h_4(w) g_{\mu\lambda} v'_\rho + v'_\rho v_\lambda [h_5(w) v_\mu$$

$$+ h_6(w) v'_\mu] + i\epsilon_{\mu\alpha\beta\gamma}\Big( g^\beta_\rho g^\gamma_\lambda [h_7(w)(v+v')^\alpha$$

$$+ h_8(w)(v-v')^\alpha] + v'^\beta v^\gamma [h_9(w) g^\alpha_\rho v_\lambda$$

$$+ h_{10}(w) g^\alpha_\lambda v'_\rho]\Big)\Big\} \bar{U}^\lambda_{\Xi^*_{Q'Q_s}}(v') U^\rho_{\Xi^*_{QQ_s}}(v)\eta(w).$$

Here, $\eta(w)$ is the heavy-diquark–light-quark Isgur–Wise function, which is determined by the dynamics of the light spectator quark $q$ and is calculated similarly to Eqs. (73) and (62):

$$\eta(w) = \sqrt{\frac{2}{w+1}} \int \frac{d^3p\, d^3q}{(2\pi)^3} \bar{\Psi}_B(\mathbf{q}) \Psi_B(\mathbf{p}) \tag{92}$$

$$\times I_q(p,q)\delta^3\left(\mathbf{p} - \mathbf{q} + \frac{\epsilon_q(p) + \epsilon_q(q)}{w+1}(\mathbf{v}' - \mathbf{v})\right).$$

We plot the Isgur–Wise function $\eta(w)$ in Fig. 10. In the nonrelativistic limit for heavy quarks, the diquark form factors $h_i(w)$ obey relations (74)–(77). In this limit, the baryon transition matrix elements contain the common factor $F(w)\eta(w)$ (cf. [41]).

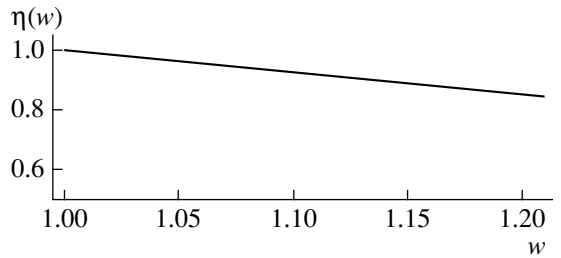


**Fig. 10.** The Isgur–Wise function $\eta(w)$ of the light-quark–heavy-diquark bound system.

## 8. SEMILEPTONIC DECAY RATES OF DOUBLY HEAVY BARYONS

The exclusive differential rate of the doubly heavy baryon semileptonic decay $B \to B' e\bar{\nu}$ can be written in the form

$$\frac{d\Gamma}{dw} = \frac{G_F^2 |V_{bc}|^2 M_f^3 \sqrt{w^2-1}}{48\pi^3}\Omega(w), \tag{93}$$

where $w = (v\cdot v') = (M_i^2 + M_f^2 - k^2)/(2M_i M_f)$, $k = P - Q$, and the function $\Omega(w)$ is the contraction of the hadronic transition matrix elements and the leptonic tensor. For the massless leptons, the differential decay rates of the transitions $\Xi'_{QQ_s} \to \Xi_{Q'Q_s}$ and $\Xi_{QQ_s} \to \Xi'_{Q'Q_s}$ are as follows:

$$\frac{d\Gamma}{dw}(\Xi'_{QQ_s} \to \Xi_{Q'Q_s}) = \frac{G_F^2 |V_{QQ'}|^2}{72\pi^3} \tag{94}$$

$$\times (w^2-1)^{1/2}(w+1)^3 M_f^3 \Big\{ 2(M_i^2 + M_f^2$$

$$- 2M_i M_f w)\left[h^2_{A_1}(w) + \frac{w-1}{w+1} h^2_V(w)\right]$$

$$+ \Big[(M_f - M_i w) h_{A_1}(w) + (w-1)$$

$$\times (M_f h_{A_2}(w) + M_i h_{A_3}(w))\Big]^2\Big\}\eta^2(w),$$

$$\frac{d\Gamma}{dw}(\Xi_{QQ_s} \to \Xi'_{Q'Q_s}) = \frac{M_f^3}{M_i^3} \tag{95}$$

$$\times \frac{d\Gamma}{dw}(\Xi'_{QQ_s} \to \Xi_{Q'Q_s}, M_i \leftrightarrow M_f, h_{A_2} \to \tilde{h}_{A_2},$$

$$h_{A_3} \to \tilde{h}_{A_3}).$$

The differential decay rates for other transitions are given in [39].

The semileptonic decay rates of doubly heavy baryons are calculated in the nonrelativistic limit for heavy quarks and presented in Tables 14 and 15. Our results for the semileptonic decay rates of doubly heavy baryons $\Xi_{bb}$ and $\Xi_{bc}$ are compared with previous predictions. It is seen from Table 14 that results of different approaches vary substantially. Most of the previous papers [42–44] give their predictions only for selected decay modes. Their values agree with ours in order of magnitude. Our predictions are smaller than the QCD sum rule results [44] by a factor of ∼2. This can be a result of our treatment of the heavy-spectator-quark recoil in the diquark. On the other

**Table 14.** Semileptonic decay rates of doubly heavy baryons $\Xi_{bb}$ and $\Xi_{bc}$ (in $10^{-14}$ GeV)

| Decay | This work | [45] | [42] | [44] | [43] |
|---|---|---|---|---|---|
| $\Xi_{bb} \to \Xi'_{bc}$ | 1.64 | 4.28 | | | |
| $\Xi_{bb} \to \Xi_{bc}$ | 3.26 | 28.5 | | 8.99 | |
| $\Xi_{bb} \to \Xi^*_{bc}$ | 1.05 | 27.2 | | 2.70 | |
| $\Xi^*_{bb} \to \Xi'_{bc}$ | 1.63 | 8.57 | | | |
| $\Xi^*_{bb} \to \Xi_{bc}$ | 0.55 | 52.0 | | | |
| $\Xi^*_{bb} \to \Xi^*_{bc}$ | 3.83 | 12.9 | | | |
| $\Xi'_{bc} \to \Xi_{cc}$ | 1.76 | 7.76 | | | |
| $\Xi'_{bc} \to \Xi^*_{cc}$ | 3.40 | 28.8 | | | |
| $\Xi_{bc} \to \Xi_{cc}$ | 4.59 | 8.93 | 4.0 | 8.87 | 0.8 |
| $\Xi_{bc} \to \Xi^*_{cc}$ | 1.43 | 14.1 | 1.2 | 2.66 | |
| $\Xi^*_{bc} \to \Xi_{cc}$ | 0.75 | 27.5 | | | |
| $\Xi^*_{bc} \to \Xi^*_{cc}$ | 5.37 | 17.2 | | | |

hand, the authors of [45], where the Bethe–Salpeter equation is used, give more decay channels. Their results are substantially higher than ours; for some decays, the difference reaches almost two orders of magnitude, which seems quite strange. For example, for the sum of the semileptonic decays $\Xi_{bb} \to \Xi_{bc}^{(\prime,*)}$, the authors of [45] predict $\sim 6 \times 10^{-13}$ GeV, which almost saturates the estimate of the total decay rate $\Gamma^{\text{total}}_{\Xi_{bb}} \sim (8.3 \pm 0.3) \times 10^{-13}$ GeV [12] and thus is unlikely.

## 9. CONCLUSIONS

In this paper, we calculated the masses and semileptonic decay rates of the doubly heavy baryons on the basis of the quark–diquark approximation in the framework of the relativistic quark model. The mass spectra of orbital and radial excitations of both the heavy diquark and the light quark were considered. The main advantage of the proposed approach consists in the fully relativistic treatment of the light-quark $(u, d, s)$ dynamics and in account of the internal structure of the diquark in the short-range quark–diquark interaction. We apply only the expansion in inverse powers of the heavy-(di)quark mass, which considerably simplifies calculations. The infinitely heavy-(di)quark limit, as well as the first-order $1/M_d$ spin-independent and spin-dependent contributions, was considered. A close similarity between excitations of the light quark in the doubly heavy baryons and heavy-light mesons was demonstrated. In the infinitely heavy-(di)quark limit, the only difference originates from the internal structure of the diquark, which is important at small distances. The first-order contributions to the heavy-(di)quark expansion explicitly depend on the values of the heavy-diquark and heavy-quark spins and masses. This results in the different number of levels to which the initially degenerate states split, as well as their ordering. Our model respects the constraints imposed by heavy-quark symmetry on the number of levels and their splittings.

**Table 15.** Semileptonic decay rates of doubly heavy baryons $\Omega_{bb}$ and $\Omega_{bc}$ (in $10^{-14}$ GeV)

| Decay | $\Gamma$ | Decay | $\Gamma$ |
|---|---|---|---|
| $\Omega_{bb} \to \Omega'_{bc}$ | 1.66 | $\Omega'_{bc} \to \Omega_{cc}$ | 1.90 |
| $\Omega_{bb} \to \Omega_{bc}$ | 3.40 | $\Omega'_{bc} \to \Omega^*_{cc}$ | 3.66 |
| $\Omega_{bb} \to \Omega^*_{bc}$ | 1.10 | $\Omega_{bc} \to \Omega_{cc}$ | 4.95 |
| $\Omega^*_{bb} \to \Omega'_{bc}$ | 1.70 | $\Omega_{bc} \to \Omega^*_{cc}$ | 1.48 |
| $\Omega^*_{bb} \to \Omega_{bc}$ | 0.57 | $\Omega^*_{bc} \to \Omega_{cc}$ | 0.80 |
| $\Omega^*_{bb} \to \Omega^*_{bc}$ | 3.99 | $\Omega^*_{bc} \to \Omega^*_{cc}$ | 5.76 |

We find that the $p$-wave levels of the light quark with $j = 1/2$ and $j = 3/2$, are inverted in the infinitely heavy (di)quark limit. The origin of this inversion is the following. The confining potential contribution to the spin–orbit term in (28) exceeds the one-gluon-exchange contribution. Thus, the sign before the spin–orbit term is negative [46, 47], and the level inversion emerges. However, the $1/M_d$ corrections, which produce the hyperfine splittings of these multiplets, are substantial. As a result, the purely inverted pattern of $p$ levels for the heavy diquark in the ground state occurs only for the doubly heavy baryons $\Xi_{bb}$ and $\Omega_{bb}$. For $\Xi_{cc}$ and $\Omega_{cc}$ baryons, the levels from these multiplets overlap. A similar pattern was previously found in our model for the heavy-light mesons [19].

Then we used the baryon wave functions for the calculation of semileptonic decay rates of doubly heavy baryons. The weak transition matrix elements between heavy-diquark states were calculated with the self-consistent account of the spectator-quark recoil. It was shown that recoil effects lead to additional contributions to the transition matrix elements. Such terms were missed in the previous quark-model calculations. If we neglect these recoil contributions, the previously obtained expressions [41, 42] for heavy-diquark transition matrix elements are reproduced. It was found that these recoil terms, which are proportional to the ratio of the heavy-spectator to the final active-quark mass [see Eqs. (74)–(79)], give important contributions to transition matrix elements of doubly heavy diquarks even in the nonrelativistic

limit. In this limit, these weak-transition matrix elements are proportional to the function $F(w)$ (78), which is expressed through the overlap integral of the heavy-diquark wave functions. The function $F(w)$ falls off rather rapidly, especially for the $bb \to bc$ diquark transition, where the spectator quark is the $b$ quark (see Figs. 8, 9). Such a decrease is the consequence of the large mass of the spectator quark and high recoil momenta ($q_{\max} \approx m_b - m_c \sim 3.33$ GeV) transferred.

We calculated the doubly heavy baryon transition matrix elements in the infinitely heavy diquark limit. The expressions for transition amplitudes and decay rates were obtained for the most general parametrization of the diquark transition matrix elements. The Isgur–Wise function $\eta(w)$ (92) for the light-quark–heavy-diquark bound system was determined. This function is very similar to the Isgur–Wise function of the heavy-light meson in our model [22], as is required by the heavy-quark symmetry. In the heavy-quark limit, the baryon transition matrix elements contain the common factor which is the product of the diquark form factor $F(w)$ and the Isgur–Wise function $\eta(w)$.

## ACKNOWLEDGMENTS

We are grateful to M. Ivanov, V. Kiselev, A. Likhoded, V. Lyubovitskij, M. Müller-Preussker, I. Narodetskii, A. Onishchenko, V. Savrin, and Yu. Simonov for the interest in our work and useful discussions.

Two of us (R.N.F. and V.O.G.) were supported in part by the Deutsche Forschungsgemeinschaft under contract Eb 139/2-2. A.P.M. was supported in part by the German Academic Exchange Service (DAAD).

## REFERENCES

1. J.-M. Richard, Phys. Rep. **212**, 1 (1992).
2. G. S. Bali, Phys. Rep. **343**, 1 (2001).
3. D. S. Kuzmenko and Yu. A. Simonov, Phys. Lett. B **494**, 81 (2000); Yad. Fiz. **64**, 110 (2001) [Phys. At. Nucl. **64**, 107 (2001)]; **66**, 983 (2003) [**66**, 950 (2003)].
4. M. Anselmino, E. Predazzi, S. Ekelin, *et al.*, Rev. Mod. Phys. **65**, 1199 (1993).
5. A. V. Manohar and M. B. Wise, *Heavy Quark Physics* (Cambridge Univ. Press, Cambridge, 2000).
6. J. G. Körner, M. Kramer, and D. Pirjol, Prog. Part. Nucl. Phys. **33**, 787 (1994).
7. M. J. Savage and M. B. Wise, Phys. Lett. B **248**, 177 (1990).
8. E. Bagan, H. G. Dosch, P. Godzinsky, and J.-M. Richard, Z. Phys. C **64**, 57 (1994).
9. R. Roncaglia, A. Dzierba, D. B. Lichtenberg, and E. Predazzi, Phys. Rev. D **51**, 1248 (1995); R. Roncaglia, D. B. Lichtenberg, and E. Predazzi, Phys. Rev. D **52**, 1722 (1995).
10. D. Ebert, R. N. Faustov, V. O. Galkin, *et al.*, Z. Phys. C **76**, 111 (1997).
11. S. S. Gershtein, V. V. Kiselev, A. K. Likhoded, and A. I. Onishchenko, Phys. Rev. D **62**, 054021 (2000).
12. V. V. Kiselev and A. K. Likhoded, Usp. Fiz. Nauk **172**, 497 (2002) [Phys. Usp. **45**, 455 (2002)].
13. N. Isgur, Phys. Rev. D **62**, 014025 (2000).
14. M. Mattson *et al.*, Phys. Rev. Lett. **89**, 112001 (2002).
15. K. Anikeev *et al.*, hep-ph/0201071.
16. A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **29**, 380 (1963).
17. A. P. Martynenko and R. N. Faustov, Teor. Mat. Fiz. **64**, 179 (1985) [Theor. Math. Phys. **64**, 765 (1985)].
18. V. O. Galkin, A. Yu. Mishurov, and R. N. Faustov, Yad. Fiz. **55**, 2175 (1992) [Sov. J. Nucl. Phys. **55**, 1207 (1992)].
19. D. Ebert, V. O. Galkin, and R. N. Faustov, Phys. Rev. D **57**, 5663 (1998); **59**, 019902(E) (1999).
20. S. Weinberg, in *Lectures on Elementary Particles and Quantum Field Theory*, Ed. by S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, MA, 1970); I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **96**, 385 (1989) [Sov. Phys. JETP **69**, 217 (1989)]; S. Ferrara, M. Poratti, and V. L. Telegdi, Phys. Rev. D **46**, 3529 (1992).
21. V. O. Galkin and R. N. Faustov, Yad. Fiz. **44**, 1575 (1986) [Sov. J. Nucl. Phys. **44**, 1023 (1986)]; V. O. Galkin, A. Yu. Mishurov, and R. N. Faustov, Yad. Fiz. **51**, 1101 (1990) [Sov. J. Nucl. Phys. **51**, 705 (1990)]; D. Ebert, R. N. Faustov, and V. O. Galkin, Phys. Lett. B **537**, 241 (2002).
22. R. N. Faustov and V. O. Galkin, Z. Phys. C **66**, 119 (1995).
23. D. Ebert, R. N. Faustov, and V. O. Galkin, Phys. Rev. D **62**, 034014 (2000).
24. A. M. Badalian and Yu. A. Simonov, Yad. Fiz. **59**, 2247 (1996) [Phys. At. Nucl. **59**, 2164 (1996)].
25. A. Di Giacomo, H. G. Dosch, V. I. Shevchenko, and Yu. A. Simonov, Phys. Rep. **372**, 319 (2002).
26. D. Ebert, R. N. Faustov, V. O. Galkin, and A. P. Martynenko, Phys. Rev. D **66**, 014008 (2002).
27. A. P. Martynenko and V. A. Saleev, Phys. Lett. B **385**, 297 (1996).
28. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press, New York, 1957; Fizmatgiz, Moscow, 1960).
29. A. P. Martynenko and R. N. Faustov, Teor. Mat. Fiz. **66**, 399 (1986).
30. I. B. Khriplovich and A. I. Milstein, Preprint No. 96-49, BINP; hep-ph/9607374; P. Labelle and G. P. Lepage, Phys. Rev. Lett. **72**, 2006 (1994).
31. H. J. Schnitzer, Phys. Lett. B **76B**, 461 (1978).
32. V. V. Kiselev, A. K. Likhoded, O. N. Pakhomova, and V. A. Saleev, Phys. Rev. D **66**, 034030 (2002).
33. I. M. Narodetskii and M. A. Trusov, Yad. Fiz. **65**, 949 (2002) [Phys. At. Nucl. **65**, 917 (2002)]; I. M. Narodetskii, A. N. Plekhanov, and A. I. Veselov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 64 (2003) [JETP Lett. **77**, 58 (2003)].

34. R. Lewis, N. Mathur, and R. M. Woloshyn, Phys. Rev. D **64**, 094509 (2001).
35. H. J. Lipkin, Phys. Lett. B **171**, 293 (1986).
36. Particle Data Group (D. E. Groom *et al.*), Eur. Phys. J. C **15**, 1 (2001).
37. R. N. Faustov, Ann. Phys. (N.Y.) **78**, 176 (1973).
38. R. N. Faustov, V. O. Galkin, and A. Yu. Mishurov, Phys. Rev. D **53**, 6302 (1996).
39. D. Ebert, R. N. Faustov, V. O. Galkin, and A. P. Martynenko, Phys. Rev. D **70**, 014018 (2004).
40. A. F. Falk and M. Neubert, Phys. Rev. D **47**, 2965 (1993).
41. M. J. White and M. J. Savage, Phys. Lett. B **271**, 410 (1991).
42. M. A. Sanchis-Lozano, Nucl. Phys. B **440**, 251 (1995); Phys. Lett. B **321**, 407 (1994).
43. V. E. Lyubovitskij, A. Faessler, Th. Gutsche, *et al.*, Prog. Part. Nucl. Phys. **50**, 329 (2003).
44. A. I. Onishchenko, hep-ph/0006271; hep-ph/0006295, and private communication.
45. X.-H. Guo, H.-Y. Jin, and X.-Q. Li, Phys. Rev. D **58**, 114007 (1998).
46. A. Yu. Dubin, A. B. Kaidalov, and Yu. A. Simonov, Phys. Lett. B **323**, 41 (1994).
47. Yu. S. Kalashnikova, A. V. Nefediev, and Yu. A. Simonov, Phys. Rev. D **64**, 014037 (2001).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 808–813. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 841–846.*


# Multiquark States*

**F. Wang[1], J. L. Ping[1,2], D. Qing[3], and T. Goldman[4]**

Received May 21, 2004; in final form, November 24, 2004

**Abstract**—The pentaquark state recently discovered has been discussed based on various quark model calculations. Odd parity for the state cannot be ruled out theoretically because the contributions related to nontrivial color structures have not been studied completely. Other multiquark states, especially dibaryons, have been discussed also. A strangeness $-3$ $N\Omega$ dibaryon has been shown to have a width as small as 12–22 keV and should be detectable in Ω-high-productivity reactions such as at RHIC, COMPAS, and the planned JHF and FAIR projects. © *2005 Pleiades Publishing, Inc.*

## 1. MULTIQUARK STATE SEARCH

Multiquark states were studied even before the advent of QCD. The development of QCD accelerated multiquark studies because it is natural in QCD that there should be multiquark states, including glueballs and quark–gluon hybrids. Prof. Yu.A. Simonov is one of the pioneers of multiquark studies; he led an ITEP group that developed the quark compound bag model in the early 1980s to study hadron interactions and multiquark states [1].

For a long time, multiquark states were only a theoretical speculation; experimental searches had not obtained definite evidence, even though there were various claims of the discovery of multiquark states.

## 2. DISCOVERY OF THE PENTAQUARK

Eleven groups [2] claimed recently that they found a pentaquark state, now called $\Theta^+$, with mass $\sim$1540 MeV and width $\Gamma < 25$ MeV. Five measurements used a real or virtual photon–nucleus ($p$, $d$, or other nuclei) reaction and the resonance is inferred from the final state $nK^+$ or $pK^0_S$ invariant mass. A reanalysis of 1986 bubble chamber $K$–nucleus reaction data also found the $\Theta^+$. Three most recent experiments used $pp$ or $p$–nucleus reactions. One used old neutrino reaction data and the eleventh used deep-inelastic $ep$ scattering data. In addition, the NA49 Collaboration claimed that they found the antidecuplet partner $\Xi^{--}$ of $\Theta^+$ [3]. The HERA-H1 Collaboration claimed that they found the charm pentaquark $\Theta_c$ [4].

If we had not had bitter previous experience in the search for multiquark states, one would have accepted that the pentaquark state has been discovered. Taking into account the historical lessons that many low-statistics multiquark signals eventually disappeared, the consensus is that high-statistics data are necessary to confirm this state.

Moreover, there are weaknesses among these measurements. Some "pentaquark signals" of the real or virtual photon reactions might be due to the kinematical reflection of the following normal meson-production processes [5]:

$$\gamma n \to f_2 n \to K^- K^+ n, \qquad (1)$$
$$\gamma p \to a_2^+ n \to K\bar{K} n,$$
$$\gamma p \to f_2 p \to K\bar{K} p.$$

Some measurements only measure the charged $\pi^+$, $\pi^-$ decay products of $K^0_S$ to obtain the $pK^0_S$ invariant mass and there might be kinematical reflections in these events too [6]. Reanalysis of the vast $K^+n$ and $K^0_L p$ scattering data could not find the $\Theta^+$ signal with a width greater than or equal to a few MeV [7], except the recent time delay and speed plot analysis, where a signal appears around 1.57 GeV in the $P_{01}$ (with the notation $L_{I(2J)}$) channel [8]. In addition, a very tiny bump had appeared in 1973 CERN $K^+p \to pK^0_S\pi^+$ inelastic-scattering data [9].

The NA49 claim has been challenged by another CERN group based on Ξ-spectroscopy data with higher statistics [10]. From HERA-B $p$–nucleus reaction data, one has not found the $\Theta^+$ [11]. BES

*This article was submitted by the authors in English.

[1]Department of Physics and Center for Theoretical Physics, Nanjing University, Nanjing, China.
[2]Department of Physics, Nanjing Normal University, Nanjing; School of Physics and Microelectronics, Shandong University, Jinan, China.
[3]CERN, Geneva, Switzerland.
[4]Theoretical Division, Los Alamos National Laboratory, Los Alamos, USA.

$J/\psi$-decay data analysis has not found the $\Theta^+$ either [12]. There are other groups that have not found the $\Theta^+$ signal but did not claim so publicly.

## 3. WHAT IS THE PENTAQUARK

Theoretical studies based on the chiral soliton model played an important role in triggering the $\Theta^+$ searches [13]. In the chiral soliton quark model, the $\Theta^+$ is a member of the antidecuplet rotational excitation following the well-established octet and decuplet baryons. The model predicted an $I = 0$, $J^P = \frac{1}{2}^+$ state with mass about 1540 MeV and width less than 15 MeV, quite close to the later experimental results. The QCD background of this model has been criticized by Jaffe and Wilczek [14], and the difficulties of the flavor $SU(3)$ extension in the description of the meson–baryon scattering with strangeness was discussed by Karliner and Mattis [15].

Various quark models have been proposed to understand the $\Theta^+$, mainly aimed at explaining the low mass and narrow width, which is not a serious problem in the chiral soliton quark model but hard to understand in the usual quark model. First, the ground state should be an $S$-wave one in the naive quark model [16, 17] and this means $\Theta^+$ should be a negative parity state. The $S$-wave $uudd\bar{s}$ configuration will have $S$-wave $KN$ components. However, both the $I = 0, 1$ $KN$ $S$-wave phase shifts are negative in the $\Theta^+$ energy region and this means that $\Theta^+$ cannot be an $S$-wave $KN$ resonance. On the other hand, since the $I = 0$ $KN$ $P$-wave $P_{01}$ phase shifts are positive, there might be a resonance in this channel and this is consistent with the $J^P = \frac{1}{2}^+$ predicted by the chiral soliton quark model of the spin–parity of $\Theta^+$. If the $\Theta^+$ is confirmed to be a positive-parity state and there is no negative-parity pentaquark state with lower energy, then it will provide another example of an inverted energy-level structure which has been a weak point of the naive quark model of the baryon spectrum: the first excited states of $N$ and $\Delta$ have positive parity instead of negative parity as predicted by the naive quark model.

Jaffe and Wilczek [14] proposed a $\{ud\}\{ud\}\bar{s}$ structure for the pentaquark state, where $\{ud\}$ means a diquark in the color antitriplet, spin-0, isospin-0 $S$-wave state. They assume a strong color–spin force to argue for such a diquark structure, a color Cooper pair. The overall color singlet condition dictates that the four quarks must be in a color triplet state (in a color $SU(3)$ [211] representation), which is antisymmetric with respect to the diquark exchange in color space. The two diquarks (quasibosons) should be symmetric with respect to the overall diquark exchange and this dictates the relative motion of the two diquarks in a $P$ wave. In this way, Jaffe and Wilczek get a positive-parity $\Theta^+$ state, the same as the chiral soliton quark model one. This model gives an antidecuplet spectrum different from the chiral soliton quark model ones and predicts a pentaquark octet based on phenomenological assumptions. Simonov did a quantitative calculation based on the Jaffe–Wilczek configuration by means of the effective Hamiltonian approach. This method obtained a quantitative fit of single-baryon masses, but the calculated $\Theta^+$ mass is about 400 MeV higher than the observed 1540 MeV [18]. Dudek and Close [19] pointed out that there should be a spin–orbit partner $\Theta^*$ with $J^P = \frac{3}{2}^+$ not too far from the $\Theta^+$ mass.

An even more serious problem is the decay width. Buccella and Sorba [20] suggested an approximate flavor symmetry selection rule to explain the narrow width of $\Theta^+$ based on the Jaffe–Wilczek model. However, if the Jaffe–Wilczek correlated diquark–diquark wave function is totally antisymmetrized, then there will still be an $SU(6)$ totally symmetric component and no flavor symmetry selection rule to forbid the pentaquark to decay to a $KN$ final state. Diakonov and Petrov [21] even raised the criticism that the nonrelativistic quark model might not make any sense for light flavor pentaquark states.

Stancu, Riska, and Glozman [22] gave another explanation of the special properties of the pentaquark state using their Goldstone boson exchange (GBE) quark model. The flavor–spin dependence of the GBE $qq$ interaction favors a totally symmetric flavor–spin wave function of the four light quarks. The overall color singlet dictates the four light quarks to be in a color triplet state, i.e., in a color $SU(3)$ [211] representation which is the same as discussed in the Jaffe–Wilczek model. The Pauli principle further dictates the four light quarks to be in a [31] representation of $SO(3)$; therefore, they must be in a $s^3p$ configuration. The kinetic-energy increase might be compensated by a potential-energy decrease. However, the GBE interaction itself might not be enough to form the $\Theta^+$ and Stancu, Riska, and Glozman [22] have to introduce an additional $q\bar{s}$ interaction due to $\eta$-meson exchange, which was argued not to exist in the GBE approach [23]. The narrow width of $\Theta^+$ has not been discussed.

Jennings and Maltman [24] did a comparative study of the three models (chiral soliton model, diquark model, and GBE model) of the $\Theta^+$ mentioned above and related exotics. They concluded that the three models appear to be different but might describe the same physics. They also pointed out that the

narrow width of $\Theta^+$ may be explained by the small overlap of the five-quark-model wave function and the $KN$ one. But Jennings–Maltman avoid discussion of the mass of the pentaquark because it is dependent on estimates of the confinement and kinetic energy.

Our group has done three-quark-model calculations. The first one is an application of the Fock space expansion model which we developed to explain the nucleon spin structure [25]. The naive quark model assumes that the baryon has a pure valence $q^3$ configuration. This is certainly an approximation. One expects there should be higher Fock components,

$$B = aq^3 + bq^3q\bar{q} + \ldots. \quad (2)$$

The nucleon spin structure discovered in polarized lepton–nucleon deep-inelastic scattering can be explained by allowing the nucleon ground state to have about a 15% $q^3q\bar{q}$ component, where the $q\bar{q}$ parts have pseudoscalar-meson quantum numbers. In the $\Theta^+$-mass calculation, we assume $\Theta^+$ is a pure $uudd\bar{s}$ five-quark state but with channel coupling. Our preliminary results are listed in the table. In the table, $K_8N_8$ means that the $K$ and $N$ are both in color octets but coupled to an overall color singlet. These numerical results are under further check. However, the following results will not change. The $S$-wave state will have a lower mass than that of the $P$ wave; the channel coupling plays a vital role in reducing the calculated $S$-wave $\Theta^+$ mass; and it is possible to obtain a mass close to the observed mass 1540 MeV of the $\Theta^+$ if the overestimation of the $K$ mass in this model is taken into account.

Quite possibly, confinement is due to gluon flux tube (or gluon string) formation in the QCD vacuum. Simonov proposed a field correlator method to study the non-Abelian nonperturbative properties of QCD and found a $Y$-shaped gluon flux tube (or string), but not a $\Delta$-shaped one, in a baryon. The ground-state energy can be approximated by a potential [26]

$$V_{3q} = -A_{3q}\sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sigma_{3q}L_{\min} + C_{3q}, \quad (3)$$

$$L_{\min} = \sum_i L_i,$$

where $L_i$ is the distance between the quark $i$ and the $Y$-shaped gluon junction $\mathbf{r}_i$ is the position of quark $i$. The first term in Eq. (4) is the color Coulomb interaction and the second term is similar to a linear confinement potential.

Most of the constituent quark models use a quadratic or linear potential to model the quark confinement,

$$V_{\rm conf}(\mathbf{r}_{ij}) = -a\lambda_i\cdot\lambda_j r_{ij}^n, \quad (4)$$
$$r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|, \quad n = 1, 2.$$

Here, $\lambda_i^a$ $(a = 1, \ldots, 8)$ is the color $SU(3)$-group generator. For a single hadron, $q\bar{q}$ mesons, or $q^3$ baryons, such a modeling can be achieved by adjusting the strength constant $a$ of the confinement potential. The color factor $\lambda_i\cdot\lambda_j$ gives rise to a strength ratio 1/2 for baryon and meson, which is almost the ratio for the minimum length of the flux tube to the circumference of the triangle formed by three valence quarks of a baryon.

How to extend the confinement potential to multiquark systems is an open question. There are a few lattice QCD calculations of pentaquarks which obtained an $S$-wave ground state [27]. However, the lattice calculations have not given the color flux tube or string structure as the Simonov group did for mesons, baryons, and glueballs. From general $SU(3)$ color group considerations, there might be the following color structures: $q^3(1)q\bar{s}(1)$, $q^3(8)q\bar{s}(8)$, $qq(\bar{3})qq(\bar{3})\bar{s}(\bar{3})$, etc. Here, the number in parentheses denotes the color $SU(3)$ representation labeled by its dimensions. The first one is a color singlet meson–baryon channel; the second is the hidden color meson–baryon channel; the third is the color structure used in the Jaffe–Wilczek model. New color structures will give rise to additional interactions which have not been taken into account in the quark model calculations of the pentaquark so far. And our first model calculation mentioned above shows that hidden color channel coupling reduces the calculated pentaquark mass.

In the 1990s, based on the quark cluster model, we developed a model, called the quark delocalization, color screening model (QDCSM), to take into account the additional interaction in multiquark systems induced by various color structures, which are not possible for a $q\bar{q}$ meson and $q^3$ baryon, by a reparametrization of color confinement [28]. This model explains the existing baryon–baryon interaction data (bound-state deuteron and $NN$, $N\Lambda$, $N\Sigma$ scatterings) well with all model parameters fixed in hadron spectroscopy except for only one additional parameter, the color screening constant $\mu$. More importantly, *it is the unique model, so far, which explains a long-standing fact: The nuclear force and the molecular force are similar except for the obvious difference of length and energy scales; the nucleus is approximately an A-nucleon system rather than a 3A-quark system* [29].

A preliminary calculation of the pentaquark mass has been done with the QDCSM in the color singlet $KN$ configuration. In the $I = 1$ $S$-wave $KN$ channel, a pure repulsive effective interaction is obtained. This helps to rule out the $I = 1$ possibility for the $\Theta^+$. In the $I = 0$ $S$-wave $KN$ channel, a $\Theta^+$ mass

The calculated mass of $\Theta^+$ (in MeV)

| | Pure $KN$ | $KN + K^*N$ | $KN + K_8N_8$ | $KN + K^*N + K_8N_8$ |
|---|---|---|---|---|
| $S_{01}$ parity $= -$ | 2282 | 2157 | 1943 | 1766 |
| $P_{01}$ parity $= +$ | 2357.1 | 2356.3 | 2357.0 | 2336.8 |

of 1615 MeV is obtained in an adiabatic approximation. More precise dynamical calculation might reduce the calculated mass further. For the $P$-wave channels, we only obtain spin-averaged effective $KN$ interactions because the spin–orbit coupling has not been included yet. In the $I = 0$ channel, there is a strong attraction, as wanted in other quark models. However, in our model, the $P$-wave attraction is not strong enough to overcome the kinetic-energy increase to make the $P$-wave state be a ground state. This is consistent with the lattice and QCD sum-rule results [27, 30]. In the $I = 1$ channel, only a very weak attraction is obtained. This rules out the $I = 1$ $\Theta^+$ again.

In a third model, we use the $\{ud\}\{ud\}\bar{s}$ configuration. The color structure is the same as the Jaffe–Wilczek one, but the four nonstrange quarks are totally antisymmetrized. The space part is fixed to be an equilateral triangle with the two diquarks sitting at the bottom corners and the $\bar{s}$ at top. The height and the length of the bottom side of the triangle are taken as variational parameters in addition to the quark delocalization. A three-body variational calculation with the QDCSM has been done. The minimum of this variational calculation is around 1.3–1.4 GeV, corresponding to a color screening parameter $\mu = 1.0$–$0.8\ \mathrm{fm}^{-2}$. This gives rise to an effective attraction with a minimum at around 50–150 MeV, qualitatively similar to our second model result but with the possibility of further reducing the calculated $\Theta^+$ mass to be closer to the observed value.

As mentioned before, these numerical results have to be checked further and the QDCSM is only a model of QCD. Based on these results, we cannot definitively claim that the ground state of the $\Theta^+$ is $IJ^P = 0\frac{1}{2}^-$. However, this possibility has not been ruled out because color confinement contributions of various nontrivial color structures for a pentaquark system have not been studied thoroughly. The results of QDCSM can be checked by devising pentaquark interpolating field operators with these nontrivial color structures in lattice QCD calculations.

Suppose the $\Theta^+$ is finally verified to be a 1540-MeV narrow-width ($\sim$1 MeV) $IJ^P = 0\frac{1}{2}^+$ state. Then an interesting scenario similar to that of nuclear structure study from the 1940s to the 1950s will recur. The low-lying, even parity rotational excitation of nuclei is hard to explain by the naive Mayer–Jenson nuclear shell model; A. Bohr and B. Mottelson had to introduce the rotational excitation of a deformed liquid-drop model. Later, nucleon Cooper pairs were introduced because of the strong short-range pairing correlation. In the 1970s–1980s, an $S$–$D$ Cooper pair interaction boson model was developed and the collective rotation was rederived from this model, which is based on Mayer–Jenson's nuclear shell model but with nucleon pair correlation. In the description of the pentaquark, one has introduced the chiral soliton rotational excitation, quark color Cooper pairs, and much more. The historical lessons of nuclear structure study might be a good mirror to light the way for the study of hadron structure.

## 4. TETRAQUARKS, HYBRIDS, AND GLUEBALLS

By the end of the 1970s, Jaffe [31] had already suggested that the scalar mesons with masses less than 1 GeV ($\sigma(600)$, $f_0(980)$, $a_0(980)$, $\kappa(900)$) might be $qq\bar{q}\bar{q}$ states [31]. The "discovery" of a pentaquark enhances this possibility. In addition, there are a few new candidates for tetraquarks: the BABAR experiment observed a resonance of $M = 2317$ MeV, $\Gamma < 10$ MeV in the $D_S^+\pi^0$ invariant mass analysis [32]. CLEO confirmed this resonance and observed a new $D_{SJ}^+(2463\ \mathrm{MeV})$ state [33]. These might be $q\bar{q}c\bar{s}$ tetraquark states. BELLE observed a resonance of 3872 MeV in the $J/\psi\pi^+\pi^-$ channel [34] in $B$-meson decay. CDF-II confirmed this resonance in $p\bar{p}$ collisions [35]. This might be a $D^0\bar{D}^{*0}$ molecular state.

There have been glueball and hybrid candidates. In general, these states will be mixed and also mixed with tetraquark states. The "discovery" (not confirmed) of a pentaquark state will make the identification of glueballs and hybrids even harder [36].

## 5. DIBARYONS

Immediately after the development of MIT bag model, Jaffe predicted the deeply bound dibaryon $H$, an $I = S = 0$ $uuddss$ flavor singlet state [37]. Long-term extensive searches for the $H$ cannot confirm

its existence. Simonov [38] proposed the quark compound bag model and predicted various dibaryon resonances. Lomon [39] extended the $R$-matrix formalism to dibaryon studies. The $R$-matrix formalism model explains the $NN$-scattering data well because it has the aid of the boundary condition model of the $NN$ interaction [40]. It predicted a 2.7-GeV $NN$ resonance and an experimental signal was claimed.

A direct extension of the naive quark model [16, 17] to baryon–baryon interactions only gives rise to the short-range repulsion of the $NN$ interaction but not the intermediate-range attraction nor the well-established long-range tail (due to one-$\pi$ exchange) in the meson-exchange model. This means that important physics has been missed in the application of the naive quark model to the baryon–baryon interaction. One possibility arises from spontaneously broken chiral symmetry and the related GBE [41]. Obukhovsky and Kusainov [42] applied such a hybrid model, with both gluon and GBE, to the $NN$ interaction. Li *et al.* [43] extended this hybrid model from $SU(2)$ to $SU(3)$ and used it to predict dibaryon resonances. Such a direct extension with a universal $\sigma$–$u$, $d$, $s$ quark coupling will overestimate the $\sigma$-meson contribution in high-strangeness channels and makes the predicted mass of the di-$\Omega$ too low [44].

Kopeliovich *et al.* [45] studied the dibaryon with the Skyrmion model and predicted dibaryons up to high strangeness $-6$. But after taking into account the Casimir energy, these dibaryons become unbound [45].

We have studied the dibaryon candidates in the $u$, $d$, $s$ three-flavor world systematically with both nonrelativistic and relativistic versions of the QDCSM in an adiabatic approximation [46]. This model has been mentioned before. The main new ingredients are as follows: *Baryon distortion or internal excitation in the course of interaction is included by allowing the quark mutual delocalization between two interacting baryons and a variational method is developed to let the interacting baryon choose its own favorable configuration; a new parametrization of color confinement is assumed to take into account the contribution of nontrivial color structures which are not possible in a $q\bar{q}$ meson and $q^3$ baryon.* Only a few states remain after filtration with a more precise dynamic calculation: a nonstrange $IJ^p = 03^+$ $d^*$ ($M = 2165$–$2186$ MeV, $\Gamma = 5.76$–$7.92$ MeV) [47] and a strangeness $-3$ $IJ^P = \frac{1}{2}2^+$ $N\Omega$ ($M = 2549$ MeV, $\Gamma = 12$–$22$ keV). The $H$ and di-$\Omega$ are near-threshold states and might be unbound when the model uncertainty is taken into account [48]. There are few broad $N\Delta$ and $\Delta\Delta$ resonances with widths of 150–250 MeV in the energy range 2.1–2.2 GeV, which might make the identification of the $d^*$ even harder, and all of them might be the origin of the observed broad resonance of the $pp$ and $np$ total cross section in that energy region. The strangeness $-3$ $N\Omega$ state is a very narrow dibaryon resonance and might be detected in $\Omega$-high-productivity reactions such as at RHIC, COMPAS, and the planned JHF in Japan and FAIR in German.

## 6. SUMMARY

Multiquark states have been studied for about 30 years. The $\Theta^+$, if further confirmed, will be the first example. Once the multiquark Pandora's box is opened, the other multiquark states, tetraquark, hexaquark (or dibaryon), etc., cannot be kept inside. One expects that they will be discovered sooner or later. A new landscape of hadron physics will appear and it will not only show new forms of hadronic matter but will also exhibit new features of low-energy QCD.

Nonperturbative and lattice QCD have revealed the color flux tube (or string) structure of the $q\bar{q}$ and $q^3$ states. The multiquark system will have more color structures. How do these color structures interplay within a multiquark state? Nuclear structure seems to be understood in terms of colorless nucleons within a nucleus. We emphasized that other color structures should be studied and the multiquark systems provide a good field to do that. The low mass and narrow width of the $\Theta^+$ might be related to such new structures instead of to residual interactions.

## ACKNOWLEDGMENTS

We apologize to those authors whose contributions in this field have not been mentioned due to the limitation of space.

This work is supported by the NSFC under grants nos. 90103018, 10375030 and in part by the US Department of Energy under contract W-7405-ENG-36.

## REFERENCES

1. Yu. A. Simonov, Phys. Lett. B **107B**, 1 (1981).
2. T. Nakano *et al.* (LEPS Collab.), Phys. Rev. Lett. **91**, 012002 (2003); hep-ex/0301020; V. V. Barmin *et al.* (DIANA Collab.), Yad. Fiz. **66**, 1763 (2003) [Phys. At. Nucl. **66**, 1715 (2003)]; hep-ex/0304040; S. Stepanyan *et al.* (CLAS Collab.), Phys. Rev. Lett. **91**, 252001 (2003); hep-ex/0307018; J. Barth *et al.* (SAPHIR Collab.), Phys. Lett. B **572**, 127 (2003); hep-ex/0307083; A. E. Asratyan *et al.*, hep-ex/0309042; V. Kubarovsky *et al.* (CLAS Collab.), hep-ex/0311046; A. Airapetian *et al.* (HERMES Collab.), hep-ex/0312044; A. Aliev *et al.*

(SVD Collab.), hep-ex/0401024; M. Abdel-Bary *et al.* (COSY-TOF Collab.), hep-ex/0403011; P. Z. Aslanyan *et al.*, hep-ex/0403044; S. Chekanov *et al.* (ZEUS), hep-ex/0403051.
3. C. Alt *et al.* (NA49 Collab.), hep-ex/0310014.
4. A. Aktas *et al.* (HERA-H1 Collab.), hep-ex/0403017.
5. A. R. Dzierba *et al.*, hep-ph/0311125.
6. M. Zavertyaev, hep-ph/0311250.
7. R. A. Arndt, I. I. Strakovsky, and R. L. Workman, nucl-th/0308012.
8. N. G. Kelkar, M. Noakowski, and K. P. Khemchandani, hep-ph/0307134.
9. E. Lesquoy *et al.*, Nucl. Phys. B **99**, 346 (1975).
10. H. G. Fischer and S. Wenig, hep-ex/0401014.
11. K. T. Knopfle, M. Zavertyaev, and T. Zevko (HERA-B Collab.), hep-ex/0403020.
12. J. Z. Bai *et al.* (BES Collab.), hep-ex/0402012.
13. D. Diakonov, V. Petrov, and M. Polyakov, Z. Phys. A **359**, 305 (1997); hep-ph/9703373.
14. R. Jaffe and F. Wilczek, Phys. Rev. Lett. **91**, 232003 (2003); hep-ph/0307341.
15. M. Karliner and M. P. Mattis, Phys. Rev. D **34**, 1991 (1986).
16. A. De Rujula, H. Georgi, and S. L. Glashow, Phys. Rev. D **12**, 147 (1975).
17. N. Isgur and G. Karl, Phys. Rev. D **18**, 4187 (1978); **19**, 2653 (1979); **20**, 1191 (1979).
18. I. M. Narodeskii, Yu. A. Simonov, M. A. Trusov, and A. I. Veselov, Phys. Lett. B **578**, 318 (2004); hep-ph/0310118.
19. J. J. Dudek and F. E. Close, hep-ph/0311258.
20. P. Buccella and P. Sorba, hep-ph/0401083.
21. D. Diakonov and V. Petrov, hep-ph/0310212.
22. F. Stancu and D. O. Riska, hep-ph/0307010; L. Ya. Glozman, Phys. Lett. B **575**, 18 (2003); hep-ph/0308232.
23. L. Ya. Glozman, Nucl. Phys. A **663–664**, 103c (2000); N. Isgur, Phys. Rev. D **62**, 054026 (2000).
24. B. K. Jennings and K. Maltman, hep-ph/0308286.
25. D. Qing, X. S. Chen, and F. Wang, Phys. Rev. C **57**, R31 (1998); Phys. Rev. D **58**, 114032 (1998); Commun. Theor. Phys. **32**, 403 (1999); X. S. Chen, D. Qing, and F. Wang, Chin. Phys. Lett. **16**, 403 (1999).
26. Yu. A. Simonov, Nucl. Phys. B **307**, 512 (1988); D. S. Kuzmenko, V. I. Shevchenko, and Yu. A. Simonov, hep-ph/0310190.
27. F. Csikor *et al.*, hep-lat/0309090; S. Sasaki, hep-lat/0310014.
28. T. Goldman *et al.*, Nucl. Phys. A **481**, 621 (1988); F. Wang *et al.*, Phys. Rev. Lett. **69**, 2901 (1992).
29. G. H. Wu, L. J. Teng, J. L. Ping, *et al.*, Phys. Rev. C **53**, 1161 (1996); G. H. Wu, J. L. Ping, L. J. Teng, *et al.*, Nucl. Phys. A **673**, 279 (2000); J. L. Ping, H. R. Pang, F. Wang, and T. Goldman, Phys. Rev. C **65**, 044003 (2002); X. F. Lu, J. L. Ping, and F. Wang, Chin. Phys. Lett. **20**, 42 (2003).
30. S. L. Zhu, hep-ph/0307345; J. Sugiyama, T. Doi, and M. Oka, hep-ph/0309271.
31. R. L. Jaffe, Phys. Rev. D **15**, 267, 281 (1977).
32. B. Aubert *et al.* (BABAR Collab.), Phys. Rev. Lett. **90**, 242001 (2003).
33. D. Besson *et al.* (CLEO Collab.), Phys. Rev. D **68**, 032002 (2003).
34. S. K. Choi *et al.* (BELLE Collab.), hep-ex/0309032.
35. D. Acosta *et al.* (CDF-II Collab.), hep-ex/0312021.
36. E. Klempt, hep-ph/0404270.
37. R. L. Jaffe, Phys. Rev. Lett. **38**, 195 (1977).
38. Yu. A. Simonov, Nucl. Phys. A **416**, 109c (1984); **463**, 231c (1987).
39. E. L. Lomon, Phys. Rev. D **26**, 576 (1982); P. LaFrance and E. L. Lomon, Phys. Rev. D **34**, 1341 (1986).
40. E. L. Lomon and H. Feshbach, Ann. Phys. (N.Y.) **48**, 94 (1968).
41. A. Manohar and H. Georgi, Nucl. Phys. B **234**, 189 (1984).
42. I. T. Obukhovsky and A. M. Kusainov, Phys. Lett. B **238**, 142 (1990).
43. Q. B. Li *et al.*, Nucl. Phys. A **683**, 487 (2001).
44. H. R. Pang, J. L. Ping, F. Wang, and T. Goldman, Phys. Rev. C **66**, 025201 (2002); Commun. Theor. Phys. **38**, 424 (2002).
45. V. B. Kopeliovich, Phys. Lett. B **259**, 234 (1991).
46. F. Wang, J. L. Ping, G. H. Wu, *et al.*, Phys. Rev. C **51**, 3411 (1995); T. Golman, K. Maltman, G. J. Stephenson, Jr., *et al.*, Mod. Phys. Lett. A **13**, 59 (1998).
47. T. Goldman, K. Maltman, G. J. Stephenson, Jr., *et al.*, Phys. Rev. C **39**, 1889 (1989); J. L. Ping, F. Wang, and T. Goldman, Nucl. Phys. A **688**, 871 (2001).
48. T. Goldman, K. Maltman, and G. J. Stephenson, Jr., Phys. Rev. Lett. **59**, 627 (1987); H. R. Pang, J. L. Ping, F. Wang, *et al.*, nucl-th/0306043; H. R. Pang, J. L. Ping, L. Z. Chen, *et al.*, Phys. Rev. C **70**, 035201 (2004).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 814–827. Translated from Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 847–859.*


# Supernovae and Properties of Matter in the Densest and Most Rarefied States

**S. I. Blinnikov***

*Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117218 Russia and MPA, Garching, D-85741, Germany*

Received June 2, 2004; in final form, November 3, 2004

**Abstract**—An overview of the relationship between the astrophysics of supernovae and fundamental physics is given. It is shown how astronomical observations of supernovae are used to determine the parameters of matter in the most rarefied states ("dark energy"); it is also revealed that the mechanism of supernovae explosion is related to the properties of matter in the densest states. The distinction between thermonuclear and collapsing supernovae is explained. Some problems that arise in the theory of powerful cosmic explositions— supernovae and gamma-ray bursts—and which require new physics for solving them are indicated. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Still being an upperclassman in secondary school, I eagerly awaited the appearance of each successive issue of the "Feynman Lectures on Physics," translated into Russian by a group of physicists, including Yu.A. Simonov. In one of his lectures, R. Feynman said approximately the following. If asked to choose one clause that would explain the maximum number of phenomena in nature, he would say, "Matter consists of atoms." I recall that I was deeply impressed by these words. Since the main subjects of my scientific activities are associated with modern astrophysics, I have been dealing primarily with processes that are explained by the properties of atoms and ions in various states. Modern theoretical physics focuses primarily on more fundamental things—the properties of fields, strings, or some other geometric objects, on the basis of which theorists try to explain the entire micro- and macrocosm. (Yet, it cannot be stated that all of the theoretical problems—especially those that are needed for applications, in atomic physics, for example—have already been solved.)

Astrophysics is flourishing at the present time. Since the end of the twentieth century, astrophysics and classical observational astronomy have provided ever more data that can crucially affect the development of the most fundamental branches of physics. In particular, it has become clear to date that about 95% of the mean density of the observed Universe does not consist of ordinary matter—that is, of atoms. In this review article, I will consider but a small numbers of facts concerning the impact of astronomy on physics—specifically, supernovae, which are exploding objects, will be the focus of attention here.

At the beginning of the article, it is explained how parameters of cosmological models can be determined with the aid of a constant-power source of light, so-called cosmological standard candle. Further, it is shown that some subtle points in employing supernovae as cosmological standard candles are often treated incorrectly by physicists. For example, Dolgov writes, in his very useful review article [1], that type Ia supernovae are cosmological standard candles. But in fact, these supernovae are not standard candles! Nevertheless, the acceleration of the expansion of the Universe was discovered with the aid of precisely those objects.

There is no doubt that a thermonuclear explosion is the mechanism underlying the explosion of type Ia supernovae. This mechanism leads to a complete disruption of a star. Supernovae of other types stem from the collapse of the core of a star. In contrast to thermonuclear supernovae, there are many more unclear points here in the explosion mechanism. The concluding part of this brief review article is devoted mostly to collapsing supernovae.

## 2. REDSHIFT IN COSMOLOGY

In 1929, E. Hubble, who worked at a new 2.4-m telescope of the Mount Wilson Observatory near Los Angeles, was able to estimate the distance to galaxies for the first time. By plotting the redshifts $z$ of the spectral lines of those galaxies against the estimated distances $d$, he constructed a graph that, later on, was

*E-mail: sergei.blinnikov@itep.ru

1063-7788/05/6805-0814$26.00 

called the Hubble diagram. Fitting a straight line to these data, he obtained the dependence

$$z = \frac{\lambda - \lambda_{\text{lab}}}{\lambda_{\text{lab}}} = \frac{H_0 d}{c},$$

which is presently referred to as the Hubble law. Here, $\lambda$ is the observed value of the wavelength for some structure in the spectrum of a specific object, while $\lambda_{\text{lab}}$ is the laboratory value for the same structure. The value of the Hubble constant $H_0$ was overestimated by Hubble himself by almost an order of magnitude in relation to present-day data [about 70 km/(s Mpc)], but the very fact of the growth of the redshift with distance was established correctly. This fact is referred to as the recession of masses of luminous matter (galaxies and their clusters), in which case $z$ is interpreted as a manifestation of the Doppler effect ($z = v/c$ at $v \ll c$ and $v = H_0 d$). I will try to show that, although it is legitimate to interpret unambiguously the Hubble law as a manifestation of the expansion of the observed Universe, the treatment of the redshift as the Doppler effect at large cosmological distances leads to logical difficulties and is, strictly speaking, unacceptable (sometimes, the cosmological redshift is associated in the literature with the Doppler effect by definition [1]).

Theoretically, the expansion of the Universe was discovered by Friedmann [2, 3] long before Hubble's studies. (De Sitter [4] derived his solutions still earlier, but his treatment of these solutions was purely formal; only the studies of Friedmann [2, 3] established the physics pattern of a nonstationary universe—that is, they created the framework within which modern physical cosmology develops.) Friedmann's line of reasoning can be traced in the following way.

A two-dimensional sphere specified by the metric $d\ell^2 = a^2(d\theta^2 + \sin^2\theta d\varphi^2)$ is curved, but, since it is the surface of the ball $x^2 + y^2 + z^2 = a^2$, the sphere is obviously isotropic and uniform, in just the same way as a two-dimensional (2D) plane, whose metric can be represented in the form $d\ell^2 = a^2(d\theta^2 + \theta^2 d\varphi^2)$, where the polar radius is denoted by $\theta$ instead of conventional $r$.

If we now take a flat three-dimensional (3D) space and replace its metric

$$d\ell^2 = da^2 + a^2(d\theta^2 + \sin^2\theta d\varphi^2)$$

by

$$d\ell^2 = a^2[d\chi^2 + \sin^2\chi^2(d\theta^2 + \sin^2\theta d\varphi^2)], \qquad (1)$$

then, by analogy with a 2D sphere, we obtain an isotropic and uniform 3D space in the form of the surface of the 4D ball,

$$w^2 + x^2 + y^2 + z^2 = a^2,$$

where $a$ is the radius of curvature of the 3D space. Writing $ds^2 = c^2dt^2 - d\ell^2$ for four-dimensional spacetime, we will arrive at the metric of the world in the form discovered by Friedmann and used in the cosmological section of the textbook by Landau and Lifshitz [5]. A somewhat different form is obtained upon the substitution $\sin\chi = r$, in which case $d\chi^2 = dr^2/(1 - r^2)$; as a result, the interval can be written in the form

$$ds^2 = c^2dt^2 - a^2(t) \times \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2)\right].$$

(We note that, here, $r$ is a dimensionless quantity!) This is the Friedmann metric in the Robertson–Walker form. At $k = 1$, we have a closed Friedmann universe (where the curvature of 3D space is positive), while, at $k = 0$, there arises a 4D world with a flat 3D space. In the latter case, the quantity $a(t)$ loses the meaning of the radius of curvature; therefore, it would be better to refer to it as a scale factor for all of the versions. It can easily be verified that the case of $k = -1$ corresponds to yet another world, that where 3D space is of negative curvature, featuring Lobachevski geometry [3]. It is obtained from the case of $k = +1$ by means of the substitution $\sin\chi \to \sinh\chi$. In all of these cases, the point $r = 0$ can be chosen arbitrarily in a uniform space, in just the same way as a pole on a uniform 2D sphere. We assume that galaxies are points associated with fixed values of $r$, $\theta$, and $\varphi$ (comoving coordinates) and that the scale factor $a(t)$ determines the expansion of the Universe. According to present-day data, it is better, strictly speaking, to take the centers of mass of galaxy clusters (rather than galaxies) for "fixed" points, since galaxies can move within clusters at peculiar velocities of about 1000 km/s.

Friedmann did much more than just a number of coordinate transformations. He showed (taking into account the possibility that the cosmological $\Lambda$ term in nonzero) that all versions of the metrics that he considered can provide exact solutions to the Einstein equations for a reasonable choice of equation of state for matter. It is only necessary to determine the behavior of the scale factor $a(t)$ for a nonstationary world with allowance for this equation of state.

From the Einstein equations (see, for example, [5]) or directly from the Hilbert variational principle (see [6]), one can easily obtain

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_{\text{N}}}{3c^2}\mathcal{E} - \frac{kc^2}{a^2}, \qquad (2)$$

where $\dot{a} \equiv da/dt$, $G_{\text{N}}$ is the Newtonian gravitational constant and $\mathcal{E}$ is the density of all forms of energy. This relation is referred to as the Friedmann equation.

The physical meaning of this equation can be understood by using prerelativistic physics. If we introduce the mass $M$ within the radius $R = a$,

$$\frac{4\pi\mathcal{E}a^3}{3c^2} = M,$$

the result obtained from the Friedmann equation is identical to that in the Newtonian case under the condition that the whole energy density is due to nonrelativistic baryons—that is, to matter of pressure $P = 0$.

Assuming that the mean matter density $\rho$ is uniform at large scales ($R > 10$ Mpc in the present-day Universe), we find that the mass within the radius $R$ is $M = 4\pi\rho R^3/3$, and Newton's laws lead to the energy-conservation law

$$\frac{u^2}{2} - \frac{G_\mathrm{N}M}{R} = -\mathrm{const},$$

which holds as long as $u \equiv \dot{R} \ll c$, so that

$$\frac{(\dot{R})^2}{2} - \frac{4\pi G_\mathrm{N}\rho R^2}{3} = -\mathrm{const}.$$

This relation is equivalent to the Friedman equation (2), and this is good, since, on small scales, the world (spacetime) is flat, so that the general theory of relativity must reduce to nonrelativistic mechanics according to the correspondence principle.

But in fact, matter in the Universe is likely to involve, in addition to baryons, dark matter, which is cold in all probability—that is, nonrelativistic. Moreover, the results of observations require introducing a new substance, which was referred to, not quite appropriately, as dark energy (DE). This may be either a constant nonzero vacuum energy density or a new field [1]. In order to explain observations, it is only necessary that an equation of state of the form $P = w\mathcal{E}_\mathrm{DE}$ be valid, with the coefficient $w$ being negative (in the first approximation, it is close to $-1$)—that is, in contrast to ordinary gases, which have a positive pressure, this substance must be in a state of tension, as stretched elastic rubber (but stretched isotropically—that is, uniformly in all directions of 3D space!).

In the Friedman equation (2), one can immediately take into account a nonzero vacuum energy density (dark energy) by assuming that the energy density $\mathcal{E}$ involves two components, $\mathcal{E} = \mathcal{E}_m + \mathcal{E}_\mathrm{DE}$, where the first component is the energy density of matter, while the second component is the vacuum (dark) energy density.

In the simplest form, the inclusion of a constant vacuum energy density is equivalent to supplementing the Einstein equations with a cosmological constant $\Lambda$. The vacuum energy behaves as an ideal liquid whose density is given by

$$\mathcal{E}_\mathrm{vac} = \frac{c^4\Lambda}{8\pi G_\mathrm{N}}.$$

We arrive at $P = -\mathcal{E}_\mathrm{vac}$, provided that the thermodynamic identity $d(\mathcal{E}_\mathrm{vac}/\rho) - Pd\rho/\rho^2 = 0$, where $\rho$ is the baryon-number density, holds. Thus, we have $w = -1$. Below, we will not distinguish between $\mathcal{E}_\mathrm{vac}$ and $\mathcal{E}_\mathrm{DE}$, but, in general, $w \neq -1$ for the case of dark energy. Since the matter energy density and the radiation energy density both decrease as the Universe expands, the vacuum energy, if any, may appear to be dominant in the dynamics of expansion.

The velocity of expansion is measured by the Hubble parameter

$$H = \frac{\dot{a}}{a}.$$

The value of the Hubble parameter in the present era is referred to as the Hubble constant $H_0$ (it is constant in space rather than in time in uniform models featuring a synchronized time $t$ [7]). Since $H_0 = (\dot{a}/a)_\mathrm{now}$, we can see that the constant $k$ in the Friedmann equation is positive or negative, depending on the ratio

$$\Omega \equiv \frac{\mathcal{E}}{\rho_c c^2}, \quad \text{where} \quad \rho_c \equiv \frac{3H_0^2}{8\pi G_\mathrm{N}}.$$

If $\Omega > 1$, then $k = 1$. At $\Omega = 1$, 3D space is flat: $k = 0$; for $\Omega < 1$, we have $k = -1$.

For matter (that is, for ordinary and dark matter taken together), we will write $\Omega_m \equiv \mathcal{E}_m/\rho_c c^2$. For the vacuum or dark energy, we introduce the notation $\Omega_\Lambda \equiv \mathcal{E}_\mathrm{DE}/\rho_c c^2$.

Astronomers also introduce the dimensionless deceleration parameter

$$q = -\frac{a\ddot{a}}{\dot{a}^2},$$

which measures the rate at which the velocity of expansion changes. For historical reasons, this parameter was defined with a minus sign, since it was believed that deceleration was natural, but, according to present-day data, $q < 0$; that is, $\ddot{a} > 0$, and the expansion of the Universe is accelerated.

It can be shown that the photon frequency satisfies the relation

$$\omega a(t) = \mathrm{const}. \tag{3}$$

The simplest way to derive this relation is to go over to the conformal time $\eta$ specified by the equation $d\eta = cdt/a(t)$ (see, for example, [5]).

The metric $ds^2 = c^2dt^2 - a(t)^2d\ell^2$ is nonstatic, but, written in terms of the time coordinate $\eta$, it assumes the form $ds^2 = a(\eta)^2(d\eta^2 - d\ell^2)$. The spatial

part $d\ell^2$ is independent of $\eta$, so that zero geodesic lines are $ds = 0$—that is, light rays in the coordinates $\eta$ and $\ell$ are identical for different initial instants $\eta_e$ of the emission of light signals, and all intervals $\Delta\eta$ between the signals are constant. This means that, if the interval of the physical time between the instants of emission of two light signals at one point is $\Delta t_e$, then the interval between the instants of reception, $\Delta t_r$, varies in proportion to $a$, since $dt \propto a(t)d\eta$, whence we obtain Eq. (3).

A photon emitted with a frequency $\omega_1$ will be observed with a lower frequency $\omega_0$ if the scale factor grows:

$$\frac{\omega_0}{\omega_1} = \frac{a_1}{a_0}.$$

In cosmology, the subscript "0" is always reserved for the modern era, so that $a_1$ is an earlier (and smaller) value of the scale factor than $a_0$. The redshift measured by astronomers is then given by

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a_0}{a_1} - 1.$$

It should be emphasized that this redshift does not reduce to the conventional Doppler effect, although relation (3) can sometimes be deduced, under some assumptions, from the Doppler effect for close observers. The relative velocity of close observers (occurring at a small distance of $c\delta t$) is $v = Hc\delta t = \dot{a}c\delta t/a = c\delta a/a$. According to the Doppler effect, we then have

$$\frac{\delta\omega}{\omega} = -\frac{v}{c} = -\frac{\delta a}{a},$$

whence we obtain Eq. (3) upon integration. This derivation is given by Zeldovich and Novikov in [8], but the following comments are in order here: first, this derivation does not work in all of the cases where the above derivation in terms of the conformal time $\eta$ from [5] is valid; second, it is hazardous to apply the Doppler interpretation of the cosmological redshift $z$ at large distances. This issue was discussed in detail by Harrison [7, 9], predominantly from the philosophical point of view. I will give more physical arguments.

Let us consider the closed Friedmann universe ($k = +1$), and let $t_0$ correspond to the instant of largest expansion: $\dot{a}(t_0) = 0$. Suppose that an observer who measures redshifts has the coordinate $r = 0$. In the first order in $r$, neighboring objects are immobile; however, the velocity of recession of objects at a distance $c\delta t = a(t_0)r$ is nonzero in the second order, and calculations reveal that the Doppler redshift is $z_{\mathrm{Dop}} = r^2$. At the same time, the correct result according to Eq. (3) is $z = r^2/2$. This distinction is explained trivially by the difference in the expressions for the velocity $\delta v = \ddot{x}\delta t$ and the coordinate $\delta x = \ddot{x}(\delta t)^2/2$. The correct answer arises only upon supplementing the Doppler effect with the Einstein effect of violet shift, $z_{\mathrm{Ein}} = \delta\phi/c^2 = -r^2/2$ (here, $\delta\phi$ is the change of the gravitational potential between the objects). We then have $z = z_{\mathrm{Dop}} + z_{\mathrm{Ein}} = r^2/2$, in which case the derivation of the cosmological redshift according to [8] is inapplicable, this being indicated there. In the general case of a variable density, the Einstein shift effect cannot be singled out in a pure form, so that the equations of photon propagation must be solved directly [10].

Despite a strong temptation to attribute the cosmological redshift to the relative motion of an observer and an emitter, this is possible at large distances only in a flat world. In a curved space, this can be done only at the instant when an observer and an emitter occur at the same point, moving relative to each other. The problem is that, in a curved spacetime one cannot transport a vector (the velocity of a far object here) to the observation point: the result of a parallel transportation depends on the path along which this is done. By way of illustration, we take an arbitrary vector at the emission time $t_1$ at the point $r_1$ and transport it parallelly toward an observer at $t_0$ at the point $r_0$. In doing this, we fix the angular coordinates $\theta$ and $\varphi$ and compare two possible "natural" paths of the transportation of this vector:

(i) that which first goes from $r_1$ to $r_0$ at constant $t = t_1$ and then from $t = t_1$ to $t_0$ at constant $r = r_0$;

(ii) that which first goes from $t_1$ to $t_0$ at constant $r = r_1$ and then from $r_1$ to $r_0$ at constant $t = t_0$.

It is clear that the results will be different for the different paths in a curved spacetime, the difference being proportional to the area within the contour formed by these two paths (that is, it increases with increasing difference of $t_1$ and $t_0$ and with increasing difference of $r_1$ and $r_0$). It should be emphasized that a nonempty 4D world is curved even if $k = 0$ and 3D space is flat! Therefore, the interpretation of the redshift in terms of the Doppler effect loses physical meaning at cosmological distances, since the concept of a relative velocity becomes meaningless. (Davis and Lineweaver [11] discussed the inapplicability of the formulas of the special theory of relativity in what is concerned with the Doppler effect in cosmology, but they did not even touch upon a greater danger—the fact that the concept of a relative velocity loses physical meaning.) It is safe to say that, owing to the relation $\omega a(t) = \mathrm{const}$, the redshift is a geometric effect of the expansion of the Universe, but it is illegitimate to associate the change in the photon frequency with the relative velocity of the objects being considered.

## 3. PHOTOMETRIC DISTANCE

The redshift can be measured: if we know the laboratory wavelengths of various spectral lines in the spectra of distant objects, we can say how their wavelengths changed from the emission instant $t_1$ to the observation instant $t_0$. From here and from (3), we know the ratio of the scale factors at these two instants:

$$\frac{a(t_0)}{a(t_1)} = 1 + z.$$

How can we measure distances?

We can introduce a formal definition of a proper distance:

$$d(t) = a(t)r; \tag{4}$$

it would be measured (at $k = 0$) at the coordinate-time instant $t$ by a set of observers with rigid rulers between the points having the radial coordinates 0 and $r$ [10, 12]. (Recall that $r$ is dimensionless!)

The definition of a comoving distance, $d_c = a(t)\chi$, where $\chi$ is the Friedmann radial coordinate introduced in Eq. (1) (see, for example, [13]), is of importance in theoretical cosmology.

We cannot measure directly the "distances" $d$ and $d_c$ to far objects in the expanding Universe with the aid of rules or radio detection and ranging. Instead, various definitions are introduced in cosmography that depend on those measurements that can actually be performed (for example, the angular distance determined by measuring the transverse angular size of a standard ruler). Details are nicely explained by Weinberg [12]. I also follow the text of Carroll [14], but I would like to emphasize that it is not easy to find, in standard textbooks, a complete derivation of a formula that would relate distances to cosmological parameters. Therefore, I give such a derivation in the present article.

The so-called photometric distance

$$d_L = \left(\frac{L}{4\pi S}\right)^{1/2},$$

where $L$ is the absolute luminosity (light power) of a source and $S$ is the flux measured by an observer (energy arriving within a unit time per unit area of the receiver), is the most valuable for us. This definition corresponds to the statement that, in flat space, the flux at distance $d$ from the source is $S = L/(4\pi d^2)$. In the Friedmann universe, however, it would be incorrect to substitute here naively $d = a_0 r$ from (4), where $a_0$ is the scale factor at the instant when photons were observed at the comoving coordinate $r$ measured from the source.

The point is that the flux decreases because of two effects: individual photons undergo a phase shift by the factor $(1 + z)$, and the frequency of photon arrival also decreases by the factor $(1 + z)$. Therefore, we have

$$S = \frac{L}{4\pi a_0^2 r^2 (1+z)^2}$$

or

$$d_L = a_0 r(1 + z) = d(t_0)(1 + z). \tag{5}$$

The photometric distance $d_L$ is a quantity accessible to measurement if we have at our disposal an astrophysical source whose absolute luminosity $L$ is known (cosmological standard candle). But $r$ is not observable, so that it is necessary to get rid of it. On the zero geodesic line where $d\theta = d\varphi = 0$, we have

$$0 = ds^2 = c^2 dt^2 - \frac{a^2}{1 - kr_1^2} dr_1^2$$

or

$$\int_{t_1}^{t_0} \frac{cdt}{a(t)} = \int_0^r \frac{dr_1}{(1 - kr_1^2)^{1/2}}.$$

We now make the transformations

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{a_1}^{a_0} \frac{dt}{da}\frac{da}{a} = -\int_{a_0/a_1}^{1} \frac{a}{a_0}\frac{dt}{da} d\left(\frac{a_0}{a}\right)$$

$$= \int_1^{a_0/a_1} \frac{a}{a_0}\frac{dt}{da} d\left(\frac{a_0}{a}\right),$$

whence we obtain

$$a_0 \int_0^r \frac{dr_1}{(1 - kr_1^2)^{1/2}} = c\int_1^{z+1} \frac{d(z_1 + 1)}{H} = c\int_0^z \frac{dz_1}{H},$$

so that everything has reduced to observables like $1/H$ (equal $a/\dot{a}$), $z$ (with the aid of the relation $a_0/a_1 = z + 1$), and so on. In this way, we can get rid of $r$ in the expression for $d_L$ by expressing $r$ in terms of the elementary integral $\int_0^r (1 - kr_1^2)^{-1/2} dr_1$ and $\int_0^z dz_1/H$.

In order to do this, we make use of the Friedman equation, taking simultaneously into account the possibility of a nonzero vacuum energy.

We write the Friedmann equation (2) in the form

$$H^2 = \frac{8\pi G_N \mathcal{E}}{3c^2} - \frac{kc^2}{a^2},$$

which is equivalent to

$$H^2 = H_0^2[\Omega_m(1 + z)^3 + \Omega_\Lambda + (1 - \Omega_m - \Omega_\Lambda)(1 + z)^2],$$

where $\Omega_m = \mathcal{E}_m/\rho_c c^2$ and $\Omega_\Lambda = \mathcal{E}_{\rm DE}/\rho_c c^2$ are the parameters that were introduced above to characterize, respectively, the density of nonrelativistic matter [its density $\mathcal{E}_m$ varies in proportion to $(1+z)^3$ in accordance with the variation of the comoving volume] and the vacuum-energy density (this density is constant in the simplest case of the $\Lambda$ term). Substituting $H$ into the expression $\int_0^z dz_1/H$ and expressing $r$ in terms of $\int_0^r (1-kr_1^2)^{-1/2} dr_1$ only for the case of $k=-1$ (the cases of other $k$ values are obtained automatically by means of an analytic continuation), we arrive at the required formula for the photometric distance:

$$d_L = \frac{c}{H_0}(1+z)\frac{1}{\sqrt{\Omega_k}}\sinh\left\{\sqrt{\Omega_k}\int\limits_0^z \left[\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2\right]^{-1/2} dz\right\}.$$

Here, $\Omega_k \equiv 1-\Omega_m-\Omega_\Lambda$; for $\Omega_k < 0$, the hyperbolic sine (sinh) goes over to the trigonometric sine (sin), while $\sqrt{\Omega_k}$ goes over to $\sqrt{|\Omega_k|}$. For $\Omega_k \to 0$, the limit can easily be evaluated, sinh disappearing from the expression for $d_L$, where there only remains the integral $\int_0^z [\ldots]^{-1/2} dz$.

The dependence $d_L(z)$ can now be expressed in terms of cosmological parameters. We can see from Eq. (5) that $d_L$ is very simply related both to $d$ from Eq. (4) and to the Friedmann radial coordinate $\chi$ from Eq. (1).

For some important particular cases, we can write simple analytic expressions for the proper distance $d = d(z)$, which, according to Eq. (5), differs from $d_L$ only by the factor $(1+z)$. By way of example, we indicate that, for the vacuum-dominated de Sitter universe ($\Omega_\Lambda = 1$, $\Omega_m = 0$), the result is

$$H_0 d = cz.$$

For an empty universe ($\Omega_\Lambda = 0$, $\Omega_m = 0$), we have

$$H_0 d = \frac{c}{2}\left[(1+z) - \frac{1}{1+z}\right].$$

In the case of a flat (parabolic) Friedmann universe ($\Omega_\Lambda = 0$), $\Omega_m = 1$:

$$H_0 d = 2c\left(1 - \frac{1}{\sqrt{1+z}}\right).$$

In all of the above cases, these formulas reduce to the Hubble law $H_0 d = cz$ at small $z$, but they differ significantly at large value of the redshift $z$.

Intermediate values of $\Omega_\Lambda \approx 0.7$ and $\Omega_m \approx 0.3$ correspond to the modern standard model of the Universe ("concordance model"). These values were first obtained from observations of distant supernovae. We will now proceed to consider these extremely interesting objects.

## 4. SUPERNOVAE OF VARIOUS TYPES

Supernovae (Supernova = SN, Supernovae = SNe) are explosive bursts of stars whose luminosity (that is, radiation power) is $L \sim 10^{10} L_\odot$ or higher for a few weeks. Here, $L_\odot \approx 4\times 10^{33}$ erg/s is the luminosity of the Sun—that is, one supernova develops, within some time interval, the same light power as a mean galaxy consisting of billions of stars. It is this power that makes it possible to use supernovae in cosmography. A larger part of the star mass is disintegrated and ejected. Supernovae are among the strongest explosions in the Universe: the ejected kinetic energy is $E \sim 10^{51}$ erg.

The energy of an explosion is estimated as follows. From the widths of spectral lines, it can be found that the speed in the atmosphere is $v \sim 10^4$ km/s or higher. On the basis of a simulation of the light curves—that is, the dependences $L(t)$—the mass $M_{\rm ej}$ is estimated at about $1M_\odot$ to a few tens of $M_\odot$.

From here, one obtains estimates for the kinetic energy of supernova ejection for all types of supernova outbursts [the I$a$, I$b/c$, and II types are introduced in accordance with the special features of respective spectra (for details, see below)]: on average, $E \sim 10^{51}$ erg $\equiv$ 1FOE [10 raised to the power of 51 or Fifty-One Ergs (FOE)], although there occur deviations of an order of magnitude above and below this value. Within a few ten thousand years, this energy dissipates in the circumstellar medium, heating it and generating x rays and cosmic rays—that is, it generates a gaseous remnant of a supernova.

The aforementioned ejecta enrich the medium in heavy elements. Shock waves gather the circumstellar medium into dense clouds, and this leads to the formation of young stars. The history of investigations devoted to supernovae is outlined in the monograph of Shklovsky [15] and in the review articles [16, 17].

Owing to the use of state-of-the-art astrophysical methods, a vast body of observational data on supernovae has been obtained over all ranges of electromagnetic radiation—from the radio-wave to the x-ray range. Also, the first nonelectromagnetic signals—neutrinos from the SN 1987$A$—have been recorded. Despite all this, the mechanisms of supernova explosions have yet to be clarified conclusively. In recent years, there have appeared indications that some cosmic gamma-ray bursts are related to supernovae. Possibly, the origin of gamma-ray bursts is also related to the origin of supernova explosions.

Cosmic gamma-ray bursts are irregular pulses of photons whose energies range between about 0.1 and 1 MeV or take even higher values, the pulse duration being a few tenths of a second to a few minutes. The exposure corresponding to the weakest of such pulses is $S \sim 10^{-7}$ erg/cm$^2$ $\sim$1 photon/cm$^2$, but many of them are much stronger. Gamma-ray bursts are surveyed in [18, 19]. However, any surveys are always behind the latest advances in the vigorously developing science that explores powerful explosions in the Universe, this especially concerning the most recent discoveries in the realms of gamma-ray bursts.

Theory proposes various explanations for the origin of the supernova energy and seeks mechanisms underlying supernova explosions. In thermonuclear scenarios, an explosion begins because of the development of a thermal instability in the degenerate stellar core upon the ignition a carbon–oxygen mixture (helium in some scenarios).

The gravitational collapse of a star into a neutron star or a black hole—this seems a more efficient mechanism of a star explosion—was described long ago. Theoretically, the energy released in this case may be an order of magnitude higher than the thermonuclear energy—specifically, it must be about 10% of the star-core mass. For a core of mass $1M_\odot$, this energy is about $10^{53}$ erg; for a more massive core, it is naturally still higher. But in the case of supernova explosions, the bulk of energy is carried away by the neutrino flux. In gamma-ray bursts, gamma-ray photons alone carry up to about $10^{54}$ erg! This value is obtained from the observed flux $S$ by the formula $E_{\text{gamma-ray burst}} = 4\pi S d_L^2$, where $d_L$ is the photometric distance to the gamma-ray burst under study. It is of course clear that, in this formula, the flux $S$ must be multiplied by some small solid angle rather than by $4\pi$. The energy will then be lower, but it is necessary to reveal concurrently the reasons behind the release of energy in the form of a narrow beam or jet within this small angle. Possibly, the asymmetry of explosions is related to some phenomena in supernovae.

The traditional astronomical classification partitions supernovae into two classes: type-I (SN I, whose spectrum does not contain hydrogen lines in the vicinity of the maximum light) and type-II (SN II, whose spectrum features hydrogen lines there) supernovae. Later, this classification was refined. The first class was divided into subclasses: SN I$a$ and SN I$b/c$. The spectra of SN I$b/c$ in the vicinity of the maximum light do not exhibit a silicon line, which is pronounced in SN I$a$. The spectra of SN I$a$ and SN I$b/c$ show the most glaring distinctions within a late era, $t \gtrsim 250$ days after the explosion, where the spectra of SN I$a$ are formed largely by the lines of ionized iron, while the spectra of SN I$b/c$ are dominated by an extremely powerful emission from oxygen. Nonthermal radiation from SN I$b/c$ was discovered, and these objects are likely to be correlated with the regions of active star formation. In all probability, these supernovae explode through the collapse of cores of massive stars (in just the same way as SN II), whereas SN I$a$ are thermonuclear explosions of white dwarfs in binary systems that have lost hydrogen by the instant of the explosion (no neutron stars or black holes arise in this case). Thus, the classical astronomical classification of supernovae does not take fully into account the special features of the mechanism governing the explosion, which occurs in the interior of the star—it is more adequate to the structure of the outer layers of a supernova.

The SN II phenomenon can arise at the end of the lifetime of a single massive star that preserved hydrogen in its envelope. The outbursts of SN I$b/c$ can occur in the collapse of the core of a single massive star that lost hydrogen. If the SN I$c$ subtype can actually be distinguished from SN I$b$, this means that SN I$c$ massive presupernovae lost not only hydrogen but also helium. In SN I$a$, approximately half the ejected mass is due to elements of the iron peak, while, in SN I$b/c$, the bulk of these elements went into the collapse. Therefore, SN I$b/c$ ejection is dominated by elements like oxygen, whereby the distinction between the spectra of supernovae belonging to different types is explained.

The affiliation of a star with a binary system plays a crucial role in the evolution of type I$a$ presupernovae. Binarity effects seem responsible for the properties of some peculiar type II supernovae as well. The possibility of a gamma-ray burst in a binary system leads to interesting effects—this is one of the possible models for the afterglow of a gamma-ray burst [20]. Moreover, the idea that gamma-ray bursts can be generated at cosmological distances in the merger of a neutron-star binary was put forth long ago [21].

## 5. TYPE I$a$ SUPERNOVAE: STANDARDIZATION OF A CANDLE

Owing to a number of factors, type I$a$ supernovae (SN I$a$) are convenient for measuring distances and for determining the geometry of the Universe. First, these are very bright objects, so that we can obtain rich information about them even if they explode in very distant galaxies characterized by large redshifts $z$. Second, the spectra of SN I$a$ and the shapes of their light curves seem to suggest, at first glance, that they form quite a uniform class. Formerly, it was assumed that they are cosmological standard candles in the sense that the maxima of the absolute luminosity are identical for different SN I$a$.

However, this is not so! A closer inspection of SN I*a* revealed distinctions within this class of objects.

Long ago, Pskovskiĭ showed [22] that the flux maxima in the SN I spectra are not identical, and his colleagues revealed clear-cut distinctions even within the subclass of thermonuclear SN I*a* [23].

Pskovskiĭ [22] also found the interplay between the maximum luminosity of SN I*a* and the rate of the subsequent weakening of flux. The flux of more powerful bursts decreases more slowly than the flux of their less powerful counterparts. Later on, this dependence was vigorously studied by many astronomers interested in SN I*a*—especially meticulous studies on the subject were performed by Phillips and his colleagues [24, 25] on the basis of observations of close supernovae characterized by moderately small $z$.

When astronomers discover a supernova characterized by a large redshift $z$, they determine the rate of the decrease in its flux after the maximum; only after the application of the Pskovskiĭ–Phillips dependence, which provides the only way to perform the "standardization of a candle", can one estimate the luminosity of the supernova and, hence, the photometric distance $d_L$ to it. However, the Pskovskiĭ–Phillips dependence is of a correlation rather than a functional character; therefore, each individual measurement may involve large errors.

An unexpected result was obtained in the studies of two groups [26, 27]: from observational data on distant supernovae, it follows quite reliably that $\Omega_\Lambda > 0$; that is, the expansion of the Universe accelerates.

A vast body of observational data has been accumulated since then. The table presents the cosmological parameters obtained in one of the recent studies [28] on SN I*a*. The authors of [28] show that the observed distribution of $d_L(z)$ can be explained only by assuming a nonzero vacuum energy or a rather artificial systematic effect: the emergence of dust in a recent era.

It should be noted that, in all of the studies devoted to high-$z$ supernovae, use was made of relations of the Pskovskiĭ–Phillips type (*maximum luminosity–rate of the decrease in flux* relation), which were obtained from an analysis of close objects. But even for close SN I*a*, deviations from such dependences for individual objects cannot be explained by the errors of the observations exclusively.

From the theoretical point of view, the flux decreases more slowly with increasing maximum luminosity because both these quantities are controlled primarily by the amount of $^{56}$Ni formed in the explosion. The maximum of the luminosity of SN I*a* is determined by the amount of $^{56}$Ni since the light curve is formed predominantly by the contribution of its radioactive decay. On the other hand, a large amount of nickel is expected to increase the nontransparency of matter greatly. The diffusion of radiation through a stellar medium takes a longer time, and the light curve becomes more gently sloping. However, the decrease on the light curve depends not only on the amount of nickel but also on its distribution (and on the distribution of other heavy elements as well) within the expanding star and on the velocity of the expansion of matter. In turn, this distribution and this velocity depend on how burning propagated through the star.

Values of $\chi^2$ in comparing data on 157 SN I*a* with various models

| Model | $\chi^2$ |
|---|---|
| $\Omega_m = 0.27, \Omega_\Lambda = 0.73$ | 171 |
| $\Omega_m = 1.00, \Omega_\Lambda = 0.00$ | 497 |
| $\Omega_m = 0.00, \Omega_\Lambda = 0.00$ | 196 |
| Gray dust (at $\Omega_m = 1.00, \Omega_\Lambda = 0.00$) | 293 |
| Recent dust (at $\Omega_m = 1.00, \Omega_\Lambda = 0.00$) | 168 |
| Weakening in proportion to $z$ (at $\Omega_m = 1.00, \Omega_\Lambda = 0.00$) | 241 |

The theory of burning in supernovae has actively developed since the studies of Arnett [29], Ivanova *et al.* [30], and Nomoto *et al.* [31], but many problems in it have yet to be solved (see, for example, [32, 33]).

A great number of models have been proposed to describe the explosion of SN I*a* characterized by various masses, various regimes of burning (detonation, deflagration, and various combinations of these mechanisms), various energies of the explosion, and various velocities of matter expansion. In these theoretical models, chemical elements originate from burning in markedly different proportions, their distributions over the star also being different. This leads to different theoretical light curves. By comparing the resulting light curves with their observed counterparts, one can find out which explosion models are more realistic.

By considering the possibility of employing SN I*a* in cosmology, it was concluded in [34] that, at the present time, the statistics of distant supernovae are insufficient for drawing definitive conclusions on the geometry of the Universe. Terrestrial experiments revealed that the regime of burning in an explosion cannot always be predicted in advance. For supernovae, the situation is similar: it is quite feasible that the distinction between initial conditions only changes

the probabilities of various scenarios of burning, but that this does not pinpoint a specific scenario. Since the regime of burning affects strongly the shape of the light curve, the rate of the decrease cannot be reliably predicted if only initial conditions are known. The probability of one value of the rate of the decrease in flux or another—recall that this quantity plays a significant role in determining cosmological parameters—can be established only upon collecting sufficiently vast observational statistics of SN I$a$ at various values of $z$.

Upon an increase in statistics, it would be possible to reveal subtler effects—in particular, to answer the question of whether the dark-energy density is constant and to establish the relevant equation of state. For example, there are even presently attempts [35] at extracting, from observations of SN I$a$, the $z$ dependence of the coefficient $w$ in the equation $P = w\mathcal{E}_{\rm DE}$. These attempts are as yet premature, since they take no account of the fact that the properties of supernovae themselves or the regimes of burning in them and their light curves can evolve with the age of the Universe. Moreover, there are also problems in the very procedure for extracting the dependence $w = w(z)$ from observations (see, for example, the article of Jonsson *et al.* [36], who criticize the results reported in [35]). In addition, it should be noted that no significant evolution of dark energy can be revealed [37] by combining data on supernovae with data on cosmic microwave background radiation and x-ray radiation from galaxy clusters.

## 6. CORE-COLLAPSE SUPERNOVAE

Let us now briefly touch upon the as-yet-unresolved problem of explaining the explosion in the core collapse and indicate what may be here in common with the problem of gamma-ray bursts.

At the end of the lifetime of massive stars, a core collapse must develop upon the depletion of the nuclear fuel.

As far back as the 1930s, Baade and Zwicky [38] put forth the idea that there is a relationship between supernovae outbursts and the formation of neutron stars; however, the quantitative theory of the explosion mechanism in the collapse is still far from completion. From a simple estimate of the gravitational energy $\mathcal{E}_g = -G_{\rm N}\int mdm/r \sim -G_{\rm N}M^2/R$, it follows that an energy of about $|\mathcal{E}_g| \sim 10^{53}$ erg is released in the formation of a neutron star of mass $M \approx 1M_\odot$ and radius $R \approx 10^6$ cm. However, this energy is released predominantly in the form of neutrinos rather than in the form of cosmic rays or photons, as was hypothesized by Baade and Zwicky. It is not easy to estimate the energy that is transferred to the envelope around the emerging neutron star and which is responsible for a supernova outburst. Even detailed computer calculations yield contradictory results because of uncertainties in the equation of state for superdense matter, in the rates of weak-interaction reactions, and in the fundamental properties of the neutrinos (for example, their oscillations), as well as because of difficulties in describing neutrino transport and because of the emergence of convection.

If, in the main sequence, a star had a mass in the range $8M_\odot \lesssim M \lesssim 20M_\odot$, then, at the end of its evolution, there arises a partly degenerate core of mass close to the Chandrasekhar limit. At the same time, the density becomes so high ($10^9$–$10^{10}$ g/cm$^3$) that, owing to a large chemical potential (Fermi energy) of electrons, the neutronization reactions

$$e^- + (A, Z) \to (A, Z-1) + \nu_e \qquad (6)$$

begin to proceed actively even at zero temperature. As a matter of fact, the temperature at these stages reaches of few tens of kiloelectronvolts, this enhancing electron-capture reactions. Since electrons are relativistic at such densities, the adiabatic exponent is close to the critical value of $4/3$. With increasing density, the number of electrons per baryon, $Y_e$, decreases, and the pressure begins, at some instant, growing more slowly than in proportion to $\rho^{4/3}$; this means that the gravitational force grows faster than the pressure force, with the result that a catastrophic compression (collapse) develops [16]. At the initial mass of a star in the region $M > 20M_\odot$, the mass and temperature are substantially higher, and a collapse begins owing to the photon-induced splitting of nuclei. At a still higher mass, $M \gtrsim 60M_\odot$, the production of $e^+e^-$ pairs begins contributing to the reduction of the elasticity of matter and to the loss of stability. It should be borne in mind that the above mass values are very rough because the modern theory takes very roughly into account a number of important phenomena, including a continuous loss of star mass, the rotation of stars, and the fact that they form binary systems.

As the collapse reaches the dynamical stage, the central regions of the star are compressed, within a hydrodynamic time of $t_{\rm hyd} \sim (G_{\rm N}\rho)^{-1/2}$, which is a few tenths of a second, to nuclear-matter densities. Within so short a time, photon diffusion and the electron thermal conductivity are unable to remove heat efficiently; therefore, temperature grows almost adiabatically at first. The majority of the nucleons remain bound in nuclei up to densities at which nuclei begin touching one another. Only at such densities does the elasticity of matter increase sharply, and there can occur the termination of the collapse if the mass does not exceed some limit. The reverse motion (bounce) of matter generates a shock wave at a distance of

about 50 km from the center, and this shock wave heats matter strongly. There then appear many free nucleons (because of the disintegration of nuclei). As a result, processes like

$$e^- + p \to n + \nu_e, \tag{7}$$

$$e^+ + n \to p + \bar{\nu}_e, \tag{8}$$

and the annihilation of electron–positron pairs into neutrionos,

$$e^- + e^+ \to \nu + \bar{\nu} \tag{9}$$

(this is one the most important processes at later stages of the evolution of massive stars), come into play.

An order of magnitude estimate of weak-interaction cross sections is $\sigma \sim \widetilde{G}_\mathrm{F}^2 E^2$, where $E$ is the characteristic energy of a given process and $\widetilde{G}_\mathrm{F} = G_\mathrm{F}/(\hbar c)^3$, with $G_\mathrm{F}$ being the Fermi constant. If one measures the particle energy in megaelectronvolts, it is convenient to write $\widetilde{G}_\mathrm{F}^2 = 5.3 \times 10^{-44}\ \mathrm{cm}^2/\mathrm{MeV}^2$. At temperatures of a few tens of MeV, which are attained in the case of collapse, an estimate of the cross section shows that neutrinos are vigorously produced, and it seems that they can readily transfer energy to the envelope. In the formation of a neutron star, neutrinos carry away more than $10^{53}$ erg—that is, about 10% of the mass of the Sun! If one percent of this energy were captured by the envelope of a star, the problem of the core-collapse mechanism of supernova explosion would be solved.

From the estimate of $\sigma$, it can be seen that, at densities above a value of about $10^{12}\ \mathrm{g/cm}^3$, the neutrino mean free path is indeed short—it may become five to six orders of magnitude smaller than the dimensions of a hot neutron star. In deep layers, the mean free path is determined primarily by the reactions that are inverse to the processes in (7) and (8). In the vicinity of the neutrinosphere and above it, coherent neutrino scattering on surviving nuclei is more important. Because of short mean free paths, neutrinos diffuse slowly, lose energy, and cannot eject its envelope.

For heating and ejecting outer layers of a collapsing stellar core, the process

$$\nu + \bar{\nu} \to e^- + e^+ \to \gamma,$$

which is inverse to that in (9), may also be of importance. Neutrino–antineutrino pairs of all neutrino flavors must be copiously produced in the collapse. The detailed neutrino spectra were first calculated by Nadyozhin [39]. Unfortunately, the neutrinos are overly soft for this process to be of importance for supernovae. But if hard neutrinos escape into a vacuum, it can produce many photons there! In fact, the process $\nu\bar{\nu} \to e^- + e^+$ was proposed by Berezinskii and Prilutskii [40] for explaining gamma-ray bursts before the commencement of its applications in the physics of supernovae.

## 7. ASYMMETRY OF THE EXPLOSION

Since spherically symmetric model calculations of collapsing presupernovae have not yet provided a successful pattern of explosions, it is necessary to seek mechanisms that do not feature symmetry. These mechanisms may be operative in gamma-ray bursts as well. If they produce a radiation flux into a solid angle $\Omega$, the requirements on the energy of gamma-ray bursts are relaxed by the factor $\Omega/(4\pi)$. Observations have given many indications that supernova explosions are asymmetric:

(i) Radiation from collapsing supernovae is polarized to a considerable extent. The degree of this polarization grows with decreasing mass of the hydrogen envelope, reaching a maximum for SN I$b/c$, which are deprived of hydrogen. A spectacular example is provided by a record polarization of the type I$c$ SN 1997$X$ (in all probability, such supernovae are deprived of not only the hydrogen but also the helium envelope, and this means that the ejected mass must be especially small and that the asymmetry of the explosion must be the most pronounced in such objects).

(ii) In many cases (maybe even always), the explosion of a collapsing supernova is followed by the formation of a neutron star (known examples are provided by pulsars in the Crab nebula and in the Vela remnant). Many radio pulsars are observed to have velocities of up to 1000 km/s. A high momentum corresponding to such velocities is possibly associated with an asymmetry of the respective explosion.

(iii) Observations of SN 1987$A$ revealed that

(a) in the course of the explosion, radioactive matter was very fast transferred to outer layers (also, a considerable mixing of $^{56}$Ni is required for explaining the SN 1987$A$ light curves);

(b) the infrared lines of oxygen, iron, nickel, and hydrogen exhibit a considerable asymmetry of the profiles;

(c) light was polarized;

(d) photographs from the Hubble cosmic telescope show a manifest asymmetry of ejecta, and the Chandra x-ray observatory recorded jets.

(iv) In the vicinity of the young remnant of the supernova ~1680 Cassiopeia $A$ (Cas $A$), there are quickly moving lumps of matter rich in oxygen beyond the main envelope of the remnant (maybe, there are also two oppositely directed jets).

Three-dimensional images of the Cas $A$ remnant show that the flocculent distribution of calcium, sulfur, and oxygen is not symmetric in the direction toward an observer. No simple spherical envelopes can be seen. This and other remnants have a systematic velocity of up to 900 km/s with respect to a local circumstellar medium. All of these asymmetries are expected to be associated with asymmetric flows of type I$b/c$ presupernovae leading to explosions that produce remnants like Cas $A$—that is, stars of the Wolf–Rayet type.

The latest x-ray observations of Cas A from the Chandra satellite show that ejected lumps rich in iron are farther from the center than layers rich in silicon.

(v) X-ray observations of ROSAT revealed lumps ("bullets") beyond the main envelope of the Vela remnant, radio-wave-emitting shock waves associated with them being indicative of a high speed of ejection of these lumps in a supernove explosion.

### *7.1. Mechanisms of the Asymmetry of a Collapse*

Searches for the mechanism of a supernova explosion in the collapse of the stellar core have been a challenge for the theory for several decades. I would like to indicate three possible ways of an explosion:

(i) explosion under the effect of a neutrino flux,

(ii) magnetorotational mechanism of a supernova outburst (see [41] and § 36 in [42]; see also [43] for the latest results on the subject),

(iii) merger and explosions of neutron stars.

All of these mechanisms involve asymmetry in some degree and have some bearing on the generation of gamma-ray bursts. We will briefly dwell upon the last idea exclusively.

### *7.2. Exploding Neutron Stars in Binary Systems*

For a neutron-star binary (NS + NS), an evolution scenario that leads to the explosion of one of the components and to a possible gamma-ray burst was proposed in [21]. The evolution of such a binary system is determined by gravitational radiation, which leads to the merger of the components. A similar process of the merger of white dwarfs may be one of the possible ways toward the explosion of SN I$a$. It is then natural to address the question of how frequently such events may occur in the Milky Way Galaxy. This question was explored in [44], and it was shown there that the frequency of the (NS + NS) mergers of neutron-star binaries, $R_{\rm NS}$, is approximately equal to unity per about 3000 years if there is no recoil in the formation of neutron stars. This frequency falls to unity per about 10 000 years at a recoil velocity of 400 km/s.

The evolution of a neutron-star binary has not yet been calculated in detail. In just the same way as in some models of gamma-ray bursts, there can occur a direct merger involving the formation of a black hole and jets induced by the accretion disk. An alternative possibility was considered in [21]. When the major half-axis $a$ of the orbit of a binary star becomes significantly smaller than its initial value, the less massive component (whose radius is larger) will fill its Roche lobe. This may lead to a significant flow to a massive satellite [21]. A neutron star of mass satisfying the condition $M < M_{\rm cr} \simeq 0.1M_\odot$ is dynamically unstable. Therefore, the low-mass satellite must explode at some stage. A numerical simulation revealed that the explosion results in an energy release of $E_{\rm kin} \simeq 8.8 \times 10^{50}$ erg ($\sim$4.8 MeV/nucleon). Upon taking into account more accurately physical processes accompanying the explosion, this value becomes somewhat smaller—neutrino carry away a considerable part of the energy.

If, at the center of a quickly rotating collapsing presupernova, a neutron-star binary is formed (owing to the disintegration of the core), this must lead to an asymmetric explosion, which may serve as a trigger for a full-scale supernova explosion and to a strong mixing. This scenario was proposed by Imshennik [45]. It should be noted that, in contrast to what we have in the magnetorotational mechanism [42, 43], the magnetic field does not play a crucial role in the scenario considered in [45]. If the stability of flow is lost prior to reaching the minimum mass, there occurs the merger of neutron stars, in which case the energy is released predominantly in the form of neutrinos. Concurrently, there can occur jet formation [46]. In any case, the explosion is asymmetric.

### *7.3. Prospects of Verifying the Mechanisms of the Explosion of Collapsing Supernovae and Possible Relation to Gamma-Ray Bursts*

Per average galaxy, the outbursts of core-collapse supernovae are severalfold more frequent than the outbursts of thermonuclear supernovae; however, the understanding of the former has not yet reached the level of understanding of the latter. The reasons for this are the following: the physics of matter at densities in excess of the nuclear-matter density is more uncertain, it is necessary to take into account all interactions (including relativistic gravity) at such densities, hydrodynamic flows are multidimensional in a collapse, and so on.

We will briefly indicate the main difficulties for employing the aforementioned promising mechanisms of the explosion of core-collapse supernovae.

(i) An explosion under the effect of neutrino radiation in a collapse requires developing an elaborate formalism that would describe neutrino transport in a three-dimensional convective flow for all neutrino flavors. It is necessary to take into account rotation and magnetic fields; possibly, it is also necessary to allow for the emission of gravitational waves. For an overview of the current state of this sector of supernovae theory, the interested reader is referred to [47], for example.

(ii) At the present time, the magnetorotational mechanism of a supernova outburst [41, 43] is the most successful. Here, the theory takes fully into account rotation and magnetic fields, which form the basis of the mechanism; however, other physics—for example, neutrino transport—has so far been described quite roughly. An insufficiently fast rotation of the cores of the majority of stars may be one of the difficulties in this model [48]—there is still no clarity in this issue.

(iii) Imshennik's mechanism [45, 49] also depends on the value of the moment of momentum of central star regions before the collapse, and the emerging neutron star cannot disintegrate if this value is overly small. Here, there has remained one more as-yet-unexplored issue. The merger of a neutron-star binary [21] must inevitably proceed under the effect of gravitational radiation (we know this from observations of binary pulsars in the Milky Way Galaxy), but it is unclear at the present time whether the same radiation of gravitational waves (which efficiently carries away the moment of momentum) may in principle cause the disintegration of a hot neutron protostar [50].

Per average galaxy, the frequency of gamma-ray bursts is two or three orders of magnitude less than the frequency of supernova outbursts; however, we cannot rule out the possibility that gamma-ray bursts accompany the collapse of massive stars if some special, quite exotic, conditions hold simultaneously (see, for example, [51, 52]). Since the mechanism of supernova explosion in the collapse has yet to be clarified, the theory of gamma-ray burst has to overcome more significant difficulties. We cannot rule out the possibility that the exoticism in a collapse does not concern the prevalent conditions exclusively: possibly, unknown exotic particles emerge under special conditions of a collapse, and it is these exotic particles that generate bursts [19]. If, for example, axion-like particles whose decay involves photon emission off the stellar core were formed in a collapse, then we would observe a supernova outburst in the presence of a massive envelope [53] or a powerful gamma-ray burst in the rare case of the absence of an envelope [54]. Such particles cannot be "conventional" hypothetical axions, since their properties must be substantially different from the properties of the latter; however, the constraints that were obtained from astrophysical data [55] should be respected. There is yet another exotic possibility, that which is associated with the process involving the formation of quark cores of neutron stars in a collapse. This process was proposed long ago in [56] (for relevant references and new ideas, see [57]).

In the near future, there will arise the possibility of recording the spectra of neutrinos and gravitational waves from star collapses. Together with terrestrial astronomical observations of supernovae and gamma-ray bursts and their observations performed beyond the atmosphere over all ranges of the electromagnetic spectrum, this will contribute to solving the most challenging problems of modern astrophysics and the physics of fundamental interactions.

## 8. CONCLUSION

I have tried to show how supernovae aid fundamental physics. On one hand, they demonstrate that the properties of a cosmological vacuum are nontrivial. On the other hand, problems associated with explaining the mechanism of the explosion of core-collapse supernovae require performing a thorough analysis of the properties of matter at supernuclear densities and taking into account effects of all known interactions. Possibly, resort to new exotic particles is also necessary here.

## ACKNOWLEDGMENTS

The idea of writing this review article resulted from my participation in the remarkable seminar organized by Yu.A. Simonov at the Institute of Theoretical and Experimental Physics for discussing articles of topical interest from the literature. I am grateful to him and to M.I. Vysotsky, V.S. Imshennik, D.K. Nadyozhin, L.B. Okun, I.V. Panov, K.A. Postnov, P.V. Sasorov, E.I. Sorokina, V.P. Utrobin, N.N. Chugai, and other colleagues for numerous discussions and cooperation. The article was written predominantly in Garching (Germany) under support of the MPA (Max-Planck-Institut für Astrophysik) guest program. I am indebted to W. Hillebrandt and R.A. Sunyaev for the kind hospitality extended to me there.

This work was supported in part by the Russian Foundation for Basic Research (project nos. 05-02-17480, 04-02-16793).

## REFERENCES


1. A. D. Dolgov, in *18th Les Rencontre de Physique de la Valle d'Aosta. La Truile, Aosta Valley, Italy, 2004*; hep-ph/0405089.
2. A. Friedmann, Z. Phys. **10**, 377 (1922).
3. A. Friedmann, Z. Phys. **12**, 326 (1924).
4. W. de Sitter, Mon. Not. R. Astron. Soc. **78**, 3 (1917).
5. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1967; Pergamon, Oxford, 1975).
6. S. Blinnikov, *Astrophysics of Exploding Objects. Lecture Notes* (Institute for Laser Engineering, Osaka University, 2000).
7. E. R. Harrison, Astrophys. J. **403**, 28 (1993).
8. Ya. B. Zel'dovich and I. D. Novikov, *Structure and Evolution of the Universe* (Nauka, Moscow, 1975) [in Russian].
9. E. R. Harrison, *Cosmology. The Science of the Universe* (Cambridge Univ. Press, Cambridge, 1981).
10. R. Kayser, P. Helbig, and T. Schramm, Astron. Astrophys. **318**, 680 (1997).
11. T. M. Davis and C. H. Lineweaver, Proc. Astron. Soc. Aust. **21**, 97 (2004).
12. S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972; Mir, Moscow, 1975).
13. D. W. Hogg, astro-ph/9905116.
14. S. M. Carroll, gr-qc/9712019.
15. I. S. Shklovskiĭ, *Supernovae and Problems Related to Them* (Nauka, Moscow, 1976) [in Russian].
16. V. S. Imshennik and D. K. Nadezhin, Itogi Nauki Tekh., Ser.: Astron. **21**, 63 (1982); Usp. Fiz. Nauk **156**, 561 (1988) [Sov. Phys. Usp. **31**, 461 (1988)].
17. S. I. Blinnikov, T. A. Lozinskaya, and N. N. Chugaĭ, Itogi Nauki Tekh., Ser.: Astron. **32**, 142 (1987).
18. K. A. Postnov, Usp. Fiz. Nauk **169**, 545 (1999) [Phys. Usp. **42**, 469 (1999)].
19. S. I. Blinnikov, Surveys High Energ. Phys. **15**, 37 (2000).
20. S. I. Blinnikov and K. A. Postnov, Mon. Not. R. Astron. Soc. **293**, L29 (1998).
21. S. I. Blinnikov, I. D. Novikov, T. V. Perevodchikova, and A. G. Polnarev, Pis'ma Astron. Zh. **10**, 422 (1984) [Sov. Astron. Lett. **10**, 177 (1984)].
22. Yu. P. Pskovskiĭ, Astron. Zh. **54**, 1188 (1977) [Sov. Astron. **21**, 675 (1977)].
23. O. S. Bartunov and D. Yu. Tsvetkov, Astrophys. Space Sci. **122**, 343 (1986).
24. M. M. Phillips, Astrophys. J. **413**, L105 (1993).
25. M. Hamuy, M. M. Phillips, N. B. Suntzeff, *et al.*, Astron. J. **112**, 2391 (1996).
26. B. P. Schmidt *et al.*, Astrophys. J. **507**, 46 (1998); A. G. Riess *et al.*, Astron. J. **116**, 1009 (1998).
27. S. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999).
28. A. G. Riess, L.-G. Strolger, J. Tonry, *et al.*, Astrophys. J. **607**, 665 (2004).
29. W. D. Arnett, Astrophys. Space Sci. **5**, 180 (1969).
30. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Astrophys. Space Sci. **31**, 497 (1974).
31. K. Nomoto, D. Sugimoto, and S. Neo, Astrophys. Space Sci. **39**, L37 (1976).
32. J. C. Niemeyer and S. E. Woosley, Astrophys. J. **475**, 740 (1997).
33. M. Reinecke, W. Hillebrandt, and J. C. Niemeyer, Astron. Astrophys. **386**, 936 (2002).
34. E. I. Sorokina, S. I. Blinnikov, and O. S. Bartunov, Pis'ma Astron. Zh. **26**, 90 (2000) [Astron. Lett. **26**, 67 (2000)].
35. U. Alam, V. Sahni, T. Deep Saini, and A. A. Starobinsky, Mon. Not. R. Astron. Soc. **344**, 1057 (2003); **354**, 275 (2004).
36. J. Jonsson, A. Goobar, R. Amanullah, and L. Bergstrom, JCAP **9**, 007 (2004).
37. D. Rapetti, S. W. Allen, and J. Weller, astro-ph/0409574, submitted for publication in Mon. Not. R. Astron. Soc.
38. W. Baade and F. Zwicky, Proc. Natl. Acad. Sci. USA **20**, 254 (1934).
39. D. K. Nadyozhin, Astrophys. Space Sci. **49**, 399 (1977); **51**, 283 (1977); **53**, 131 (1978).
40. V. S. Berezinskii and O. F. Prilutskii, Astron. Astrophys. **175**, 309 (1987).
41. G. S. Bisnovatyĭ-Kogan, Astron. Zh. **47**, 813 (1970) [Sov. Astron. **14**, 652 (1970)].
42. G. S. Bisnovatyĭ-Kogan, *Physical Questions of the Stallar Evolution Theory* (Nauka, Moscow, 1989) [in Russian].
43. N. V. Ardeljan, G. S. Bisnovatyi-Kogan, K. V. Kosmachevskii, and S. G. Moiseenko, Astrophys. **47**, 37 (2004); S. G. Moiseenko, G. S. Bisnovatyi-Kogan, and N. V. Ardeljan, in *"Supernovae as Cosmological Lighthouses", Padua, Italy, 2004*; astro-ph/0410330.
44. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Mon. Not. R. Astron. Soc. **288**, 245 (1997).
45. V. S. Imshennik, Pis'ma Astron. Zh. **18**, 489 (1992) [Sov. Astron. Lett. **18**, 194 (1992)]; Space Sci. Rev. **74**, 325 (1995).
46. H.-T. Janka, M. Ruffert, and T. Eberl, in *Nuclei in the Cosmos V* (Frontières, Paris, 1999), p. 325; astro-ph/9810057.
47. H.-Th. Janka, R. Buras, *et al.*, in *Proceedings of the 12th Workshop on Nuclear Astrophysics, Ringberg Castle, 2004*; astro-ph/0405289; A. Mezzacappa, in *"Supernovae as Cosmological Lighthouses", Padua, 2004*; astro-ph/0410085.
48. H. C. Spruit and E. S. Phinney, astro-ph/9803201; H. C. Spruit, Astron. Astrophys. **349**, 189 (1999).
49. V. S. Imshennik and O. G. Ryazhskaya, Pis'ma Astron. Zh. **30**, 17 (2004) [Astron. Lett. **30**, 14 (2004)].

50. A. G. Aksenov and V. S. Imshennik, Pis'ma Astron. Zh. **20**, 32 (1994) [Astron. Lett. **20**, 24 (1994)]; A. G. Aksenov, Pis'ma Astron. Zh. **25**, 163 (1999) [Astron. Lett. **25**, 127 (1999)].
51. S. E. Woosley, Astrophys. J. **405**, 273 (1993); A. I. MacFadyen and S. E. Woosley, Astrophys. J. **524**, 262 (1999).
52. S. S. Gershteĭn, Pis'ma Astron. Zh. **26**, 848 (2000) [Astron. Lett. **26**, 730 (2000)].
53. D. N. Schramm and J. R. Wilson, Astrophys. J. **260**, 868 (1982).
54. Z. Berezhiani and A. Drago, Phys. Lett. B **473**, 281 (2000).
55. G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
56. D. D. Ivanenko and D. F. Kurdgelaidze, Astrofiz. **1**, 479 (1965); Lett. Nuovo Cimento **2**, 13 (1969); A. R. Bodmer, Phys. Rev. D **4**, 1601 (1971); E. Witten, Phys. Rev. D **30**, 272 (1984).
57. Z. Berezhiani, I. Bombaci, A. Drago, *et al.*, Nucl. Phys. B (Proc. Suppl.) **113**, 268 (2002); Astrophys. J. **586**, 1250 (2003).

*Translated by A. Isaakyan*

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 828–832. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 860–864.*


# Cosmological Model with Dynamical Cancellation of Vacuum Energy and Dark Energy*

**A. D. Dolgov[1),2)] and M. Kawasaki[2)]**

Received May 21, 2004

**Abstract**—We propose a model with a compensating scalar field whose back reaction to the cosmological curvature cancels possible vacuum energy density down to the terms of the order of the time-dependent critical energy density. Thus, the model simultaneously solves the mystery of the compensation of vacuum energy with an accuracy of 120 orders of magnitude and explains the existence of the observed dark energy. At an early stage, the suggested cosmological model might experience exponential expansion without an additional inflaton field. However, the solution found is unstable with respect to small perturbations. The stability can be ensured by introducing nonanalytical terms depending upon the absolute value of the curvature scalar $R$. Unfortunately, stable solutions do not describe realistic cosmology at the matter-dominated stage. © *2005 Pleiades Publishing, Inc.*

The problem of vacuum energy (or what is the same, cosmological constant) seems to be the most serious one in contemporary fundamental physics. Any reasonable estimate of the magnitude of this energy gives a nonzero result which is 50–100 orders of magnitude larger than the value allowed by astronomy and just by our existence. The potential importance of this problem was indicated in the 1930s by several people after formulation of quantum field theory, but a more serious attitude was stimulated much later by papers [1]. Several mechanisms have since been discussed to explain this mysterious and extremely precise cancellation, but none has yet been satisfactory. A quite promising one is based on the idea [2] that there may exist a compensating field whose back reaction to the curvature of spacetime induced by vacuum energy would cancel the latter and change the expansion regime from the de Sitter to the Friedmann one. A generic feature of scenarios based on this idea is that vacuum energy is never compensated down to zero, but a noncompensated remnant is of the order of $m_{\rm Pl}^2/t^2$, where $m_{\rm Pl}$ is the Planck mass and $t$ is the cosmological time. Such models predicted that there should exist in the universe a new form of matter with an energy density close to the critical one $\rho_c \sim m_{\rm Pl}^2/t^2$ with possibly a quite unusual equation of state [2]. In a sense, it is good old "news" because it was found later that there indeed exists some dark energy in the universe which induces accelerated expansion [3] and contributes about 70% to the total energy density [4]. However, though several models with different concrete mechanisms of the compensation have been explored [5–14], none of them could satisfactorily describe realistic cosmology. In particular, the universe expansion regime was not related to the matter content of the universe. The list of papers quoted above is probably incomplete and more references and discussions can be found in reviews [15].

In this work, we take a scalar field $\phi$ as a compensating agent, though higher spin fields are also possible [5, 10, 11]. For simplicity, we confine our model to the action which depends only on the curvature scalar $R$, but in higher orders in curvature, the action may depend upon the Ricci, $R_{\mu\nu}$, and Riemann (Riemann–Christoffel), $R_{\mu\alpha\nu\beta}$, tensors. The general action containing only first derivatives of $\phi$ and the curvature scalar $R$ can be written as

$$A = \int d^4x\sqrt{g}\Big[ -\frac{1}{2}(R+2\Lambda) + F_1(R) \qquad (1)$$
$$+\frac{1}{2}F_2(R)D_\mu\phi D^\mu\phi + F_3(R)D_\mu\phi D^\mu R - U(\phi,R)\Big],$$

where the metric has the signature $(+,-,-,-)$; $D_\mu$ is the covariant derivative in this metric; $g = -\det[g_{\mu\nu}]$; $F_j$ are some functions of the curvature scalar—in particular, the Hilbert–Einstein action term (the first one in this expression) is separated out and the function $F_1(R)$ contains possible terms nonlinear in curvature; and the function $U$ is a generalization of the $\phi$ potential, which may also depend upon the curvature.

*This article was submitted by the authors in English.

[1)]INFN, Sezione di Ferrara, Italy; ICTP, Trieste, Italy; ITEP, Moscow, Russia.

[2)]Research Center for the Early Universe, Graduate School of Science, University of Tokyo, Tokyo, Japan.

1063-7788/05/6805-0828$26.00 

We took the units such that $m_{\rm Pl}^2/8\pi = 1$. The cosmological constant $\Lambda$ is expressed through the vacuum energy as $\rho_{\rm vac} = \Lambda m_{\rm Pl}^2/8\pi = \Lambda$. Redefining the field by $\phi \to K(R)\phi$, one can always bring its kinetic term to the canonical form $D^\mu\phi D^\mu\phi/2$ or annihilate the term proportional to $D_\mu R$.

In what follows, we will not use the general form of the action but will take a simplified version with $F_3 = 0$, $U = U(\phi)$, $F_2 = (1/R)^2$, and $F_1 = C_1R^2$. With this choice of $F_1$, the terms proportional to $C_1$ do not enter the trace equation for constant $R$ [see below Eq. (5)]. In this form, the action is similar to that considered in [14] with the difference that, in the quoted paper, $F_2 = (1/R)^{2n}$ with $n > 3/2$.[3)] According to the results of [14], the vacuum energy could indeed be compensated by the field $\phi$, but the compensation is very slow, so that the expansion regime changes from the exponential to an almost exponential one, $a(t) \sim \exp(\beta t^\kappa)$, with $2/3 \le \kappa < 1$. In this regime, the energy density of the usual matter is quickly inflated away and we arrive at an empty universe not only without vacuum energy but also devoid of the usual matter. A more detailed discussion of the properties of such a model is presented in [16]. As we see in what follows, a smaller $n$ drastically changes the behavior of possible solutions [17].

The equation of motion for the field $\phi$, as follows from the action (1), with $F_1 = C_1R^2$, $F_2 = (1/R)^2$, and $F_3 = 0$, can be written as

$$D_\mu\left[D^\mu\phi\left(\frac{1}{R}\right)^2\right] + U'(\phi) = 0. \qquad (2)$$

In the cosmological Friedmann–Robertson–Walker (FRW) background, this equation takes the form

$$\ddot\phi + 3H\dot\phi - 2\frac{\dot R}{R}\dot\phi + R^2U'(\phi) = 0. \qquad (3)$$

The Einstein equations are modified as

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - C_1\left(4R_{\mu\nu}R - g_{\mu\nu}R^2\right) - \left(\frac{1}{R}\right)^2 \qquad (4)$$
$$\times D_\mu\phi D_\nu\phi + \frac{1}{2}\left(\frac{1}{R}\right)^2(D_\alpha\phi)^2\left(g_{\mu\nu} + \frac{4R_{\mu\nu}}{R}\right)$$
$$- g_{\mu\nu}\left[U(\phi) + \rho_{\rm vac}\right] - 2\left(g_{\mu\nu}D^2 - D_\mu D_\nu\right)$$
$$\times\left[2C_1R - \left(\frac{1}{R}\right)^2\frac{(D_\alpha\phi)^2}{R}\right] = T_{\mu\nu},$$

where $T_{\mu\nu}$ is the energy–momentum tensor of the usual matter and $(D_\alpha\phi)^2 \equiv D_\alpha\phi D^\alpha\phi$.

[3)]We thank Andrei Linde for informing us about this work.

Taking the trace over $(\mu - \nu)$ in this equation, we obtain

$$-R + 3\left(\frac{1}{R}\right)^2(D_\alpha\phi)^2 - 4\left[U(\phi) + \rho_{\rm vac}\right] \qquad (5)$$
$$- 6D^2\left[2C_1R - \left(\frac{1}{R}\right)^2\frac{(D_\alpha\phi)^2}{R}\right] = T^\mu_\mu.$$

If the usual matter is absent, i.e., $T_{\mu\nu} = 0$, then Eqs. (3) and (5) together with the relation

$$R = -6\left(2H^2 + \dot H\right) \qquad (6)$$

form a complete system which allows the solution $H \sim h/t$ and $R \sim 1/t^2$. That is, instead of or after some de Sitter exponential regime of expansion, we arrive at the Friedmann one. To see this, let us look at the equation of motion of $\phi$ (3). This is the equation of Newtonian dynamics with the force proportional to $R^2U'$ and "liquid" friction term $(3H - 2\dot R/R)$. The friction force would explode at $R = 0$ if $\dot R \ne 0$ or, more generally, if $\dot R$ vanished more slowly than $R$. This is the case if $R$ reaches zero in finite time. It implies that the field $\phi$ would stick at the stable point of this equation at $R^2U' = 0$ (or, better to say, $\phi$ would asymptotically approach the stable point). So, in equilibrium, either $U' = 0$ or $R = 0$. If one may neglect the time derivative terms in Eq. (5), then

$$R = -4\left[\rho_{\rm vac} + U(\phi)\right]. \qquad (7)$$

So at least when the motion is slow and with a proper initial value of $\phi$, the system could evolve toward $R = 0$ and hence to a vanishing effective vacuum energy $\rho_{\rm vac}^{\rm (eff)} = \rho_{\rm vac} + U(\phi_0) = 0$, where $\phi_0$ is the value of $\phi$ for which the condition of vanishing of $\rho_{\rm vac}^{\rm (eff)}$ is realized.

One can check that the approach to the equilibrium point asymptotically for $t \gg 1$ and $|\phi - \phi_0| \ll |U'(\phi_0)/U''(\phi_0)| \ll 1$ can be described by the law

$$\phi - \phi_0 \approx a/t^2, \quad R \approx r/t^2, \quad H \approx h/t, \qquad (8)$$

and the constant coefficients satisfy the equations

$$k(h) \equiv \frac{h(2h-1)(6h-1)}{(1+3h)^2} = \frac{1}{6}\left(\frac{1}{U'(\phi_0)}\right)^2. \qquad (9)$$

The other coefficients are $a = 18U'h^2(2h-1)^2/(1+3h)$ and $r = -6h(2h-1)$. We excluded the evident solution $a = r = h = 0$, which corresponds to $\phi = \phi_0$ when the system stays at the equilibrium point.

The function $k(h)$ on the right-hand side of Eq. (9) is positive if $0 < h < 1/6$ or $h > 1/2$. (We do not consider here the contraction regime when $h < 0$.) So possible interesting solutions lie in one of these two regions, depending upon the initial conditions. The maximum value of $k(h)$ in the interval $0 < h < 1/6$ is

approximately 0.0243. So the solution in this interval exists if $|U'(\phi_0)| > 2.6$.

There could also be another regime of approach to equilibrium which, at least asymptotically, agrees with equations of motion (3), (5), and (6). In this regime, the last term on the right-hand side of Eq. (5) proportional to $D^2$ becomes nonnegligible and the solution takes the form $\dot\phi \approx -U'(\phi_0)R^2t$. In this case, both $R$ and $\phi - \phi_0$ tend to zero faster than $1/t^2$, $R \sim 1/t^{2+\sigma}$ and $\phi - \phi_0 \sim 1/t^{2+2\sigma}$, and the Hubble parameter decreases as $1/t^{1+\sigma}$ with $\sigma$ satisfying $\sigma(\sigma + 1) = 1/(6(U')^2)$. If such a regime were realized, the universe would reach a stationary state with a constant scale factor. We assume that such a pathological situation does not exist.

One can check that there are several branches of the solution. For example, the determination of $R$ from Eq. (5) through $\dot\phi$ and $(U + \rho_{\rm vac})$, in the case that the last term on the right-hand side of (5) may be neglected, can be done by solution of a cubic algebraic equation which has three possible roots. The choice of a certain root is dictated by initial conditions. We have found numerically that there exists a solution for which $Ht \to \text{const} \approx 0.5$ and $R$ tends to zero somewhat faster than $1/t^2$. So the system approaches the expansion regime typical for relativistic matter. A detailed numerical study of different solutions will be presented elsewhere.

The cosmological expansion regime is determined by the value of the constant $h$ according to $a(t) \sim t^h$. The evolution of the energy density of the usual matter is governed by the covariant conservation of its energy–momentum tensor:

$$\dot\rho_m = -3H(\rho_m + p_m), \tag{10}$$

where $\rho_m$ and $p_m$ are, respectively, energy and pressure densities of matter. So, as is well known, the energy density of relativistic matter decreases as $\rho_{\rm rel} \sim 1/a^4$, while that of nonrelativistic evolves as $\rho_{\rm nr} \sim 1/a^3$. Thus, if initially the energy density of matter were subdominant with respect to the vacuum energy and $h < 1/6$, the solution presented above should change its character because at a certain moment the usual matter would start to dominate. Probably, the energy density of matter and not compensated remnant of vacuum energy would manifest tracking behavior as suggested in [18]. One more comment is worth making here. For relativistic matter, $T^\mu_\mu = 0$ and the solution found above should be valid in the presence of such matter. However, even a small admixture of matter violating the condition of vanishing of $T^\mu_\mu$, either by the existence of massive particles, even with $m < T$, or by a quantum trace anomaly, could strongly change the behavior of the vacuum solution if $\rho_m$ drops slower than $1/t^2$, i.e., for $h < 2/3$.

If, at some stage, the universe were dominated by relativistic matter and the vacuum energy were vanishingly small, then, according to Eqs. (3) and (1), $\phi$ would be zero or a constant and the cosmology would be the usual Friedmann one. It is well known, however, that, in the course of expansion and cooling down, several phase transitions could take place in cosmological plasma. Typically, in the course of phase transition, vacuum energy changes, and if it were zero initially, it would become nonvanishing and, if $\phi$ remained constant, the newly created $\rho_{\rm vac}$ would ultimately dominate the total energy density, because $\rho_{\rm vac}$ does not decrease in the course of expansion. However, when $\rho_{\rm vac}$ becomes cosmologically essential, the compensation mechanism described above should be switched on and the solution (8) would be approached. However, in this case, the initial conditions would probably demand $h > 1/2$. A more detailed investigation is in progress.

The simple solution (8) does not describe the observed universe acceleration, but it may be because it is not precise, and a more accurate treatment could reveal the accelerated behavior. The "optimistic" picture presented above is based on the properties of a specific power law solution (8) to the equation of motion. For such a solution, the highest derivative term in Eq. (5) is subdominant and can indeed be neglected. However, the neglect of the highest derivative could be justified if the corresponding solution is stable. Unfortunately, as we have checked both numerically and analytically, this is not so and the account of $D^2$ in Eq. (5) leads to an essential instability of the solution. We have found that the scalar curvature $R$ changes sign, becoming positive, and the Hubble parameter $H$ tends to infinity in finite time (see Figs. 1 and 2).

To avoid such an unpleasant cosmological singularity, we have suggested a toy model with the substitution [19]

$$\frac{(D\phi)^2}{R^2} \to -\frac{(D\phi)^2}{R|R|}. \tag{11}$$

In this case, the solution drastically changes its behavior. In particular, $H$ does not explode to infinity but, instead, tends to radiation dominance value $H \sim 1/(2t)$, while $R$ quickly oscillates around zero with gradually decreasing amplitude (see Fig. 3). So the cosmology looks similar to the radiation dominated one, though not exactly the same. In particular, an addition of nonrelativistic matter does not essentially change the expansion regime. A more detailed discussion of the stability problem and the suggested cure is presented in our paper [19].

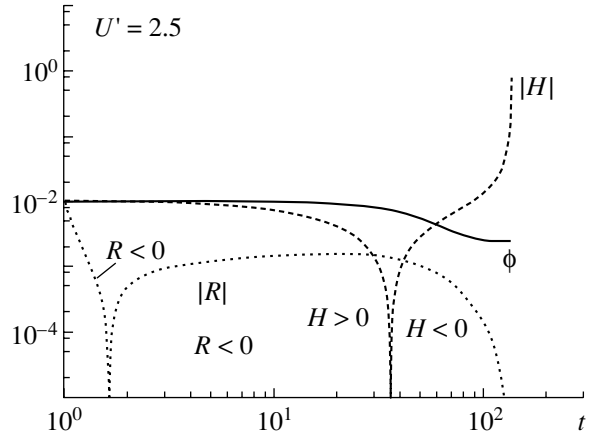


**Fig. 1.** Evolution of the scalar $\phi$, Hubble $H$, and the curvature $R$ for the potential $U' = 2.5$ and $\rho_m = 0$. For initial conditions, we take $\phi(0) = 0.01$, $H(0) = 0.01$, $R(0) = -0.01$, $\phi'(0) = 10^{-5}$, and $R'(0) = 0$.

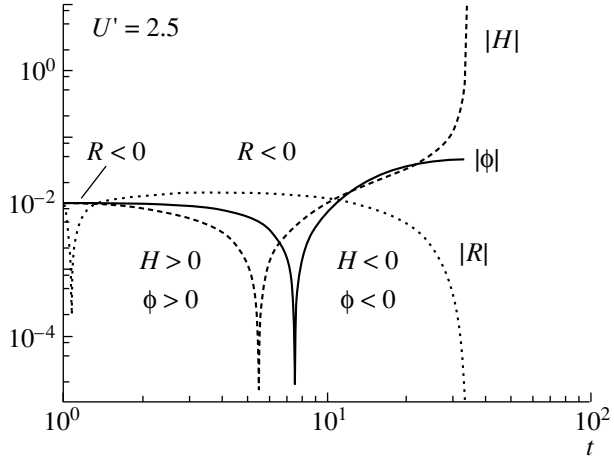


**Fig. 2.** Same as Fig. 1 except for $\rho_m(0) = 1$.

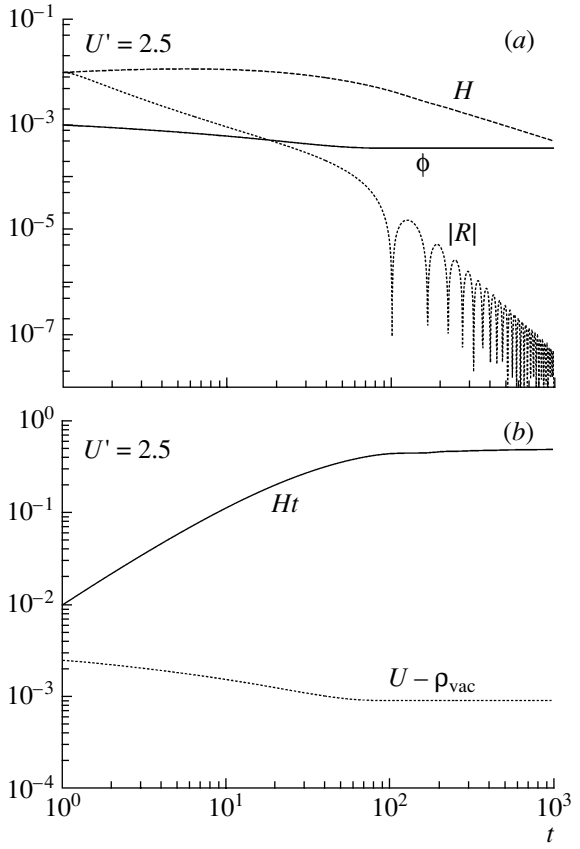


**Fig. 3.** ($a$) Evolution of the scalar $\phi$, Hubble $H$, and the curvature $R$ for the modified kinetic term Eq. (11). We take $U' = 2.5$ and $\rho_m = 0$. The initial values are $\phi(0) = 0.001$, $H(0) = 0.01$, $R(0) = -0.01$, $\phi'(0) = 10^{-4}$, and $R'(0) = 0$. ($b$) Evolution of $h = Ht$ and $U - \rho_{\rm vac}$.

The model presented here may be compatible with the inflationary scenario without an additional inflaton field. If initially $\rho_{\rm vac} \neq 0$ and the feedback (or, better to say, "kill-back") effect of $\phi$ were sufficiently slow (this could be achieved by initial conditions and the shape of $U(\phi)$), there would be a period of superluminous (inflationary) expansion. However, again the inflationary solution suffers from instability. Moreover, the modification of the model according to Eq. (11) does not enforce stability to the inflationary regime. It can be shown that the exponential expansion $a \sim \exp(H_I t)$, with $H_I \approx$ const, is a solution to the equations of motion of the model if $U'$ is sufficiently small but there is an unstable rising mode, $\delta \sim \exp(cH_I t)$ with $c > 1$ ($c \approx 2$). Thus, to realize sufficiently long inflation, this rising mode should be eliminated by initial conditions with an exponential accuracy. In this sense, inflation is unnatural, demanding quite strong fine-tuning. On the other hand, since the regions with an unsuppressed unstable mode would remain cosmologically (and maybe even microscopically) small, the anthropic principle might help. The exit from inflation could be triggered by the unstable mode and the matter might be created by the gravity itself. As was shown many years ago [20], particle production by the Planck scale gravitational field could be quite significant. Moreover, in the model suggested here, the particle production would be much more efficient because the curvature scalar oscillates with the Planck frequency.

We do not intend here to present a detailed discussion of cosmology, including big bang nucleosynthesis, large scale structure formation, etc., in this short paper. It should be a subject of a much longer work. We only want to demonstrate that, in the considered scenario, there exists a solution (maybe unstable) for which the vacuum energy is automatically canceled out with an accuracy of the order of $\rho_c \sim m_{\rm Pl}^2/t^2$ and the usual matter could survive in the course of

cosmological evolution (it was a strong drawback of previously considered models). Moreover, the noncompensated remnant of the vacuum energy could be an excellent candidate for the observed dark energy today. The model may encounter serious problems with quantization, but we think that, at this stage, it is too early to go to quantum physics; even demonstration of the existence of a classical model with automatic cancellation of vacuum energy and emergence of dark energy in the same frameworks would be an interesting result.

## ACKNOWLEDGMENTS

A.D. Dolgov is grateful to the Research Center for the Early Universe of the University of Tokyo for the hospitality during the time when this work was done.

## REFERENCES

1. Y. B. Zeldovich, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 833 (1967) [JETP Lett. **6**, 316 (1967)]; Usp. Fiz. Nauk **95**, 209 (1968) [Sov. Phys. Usp. **11**, 381 (1969)].
2. A. D. Dolgov, in *The Very Early Universe*, Ed. by G. Gibbons, S. W. Hawking, and S. T. Tiklos (Cambridge Univ. Press, Cambridge, 1982), p. 449.
3. A. G. Riess *et al.* (Supernova Search Team Collab.), Astron. J. **116**, 1009 (1998); S. Perlmutter *et al.* (Supernova Cosmology Project Collab.), Astrophys. J. **517**, 565 (1999); J. L. Tonry *et al.*, astro-ph/0305008.
4. A. T. Lee *et al.*, Astrophys. J. **561**, L1 (2001); C. B. Netterfield *et al.*, Astrophys. J. **571**, 604 (2002); N. W. Halverson *et al.*, Astrophys. J. **568**, 38 (2002); D. N. Spergel *et al.*, astro-ph/0302209.
5. A. D. Dolgov, Pis'ma Zh. Éksp. Teor. Fiz. **41**, 280 (1985) [JETP Lett. **41**, 345 (1985)].
6. R. D. Peccei, J. Sola, and C. Wetterich, Phys. Lett. B **195**, 18 (1987).
7. L. H. Ford, Phys. Rev. D **35**, 2339 (1987).
8. S. M. Barr and D. Hochberg, Phys. Lett. B **211**, 49 (1988).
9. Y. Fujii and T. Nishioka, Phys. Rev. D **42**, 361 (1990); Phys. Lett. B **254**, 347 (1991), and references therein.
10. A. D. Dolgov, Yukawa Int. Preprint YITP/K-924 (1991) (unpublished).
11. A. D. Dolgov, Phys. Rev. D **55**, 5881 (1997).
12. V. A. Rubakov, Phys. Rev. D **61**, 061501 (2000).
13. A. Hebecker and C. Wetterich, Phys. Rev. Lett. **85**, 3339 (2000).
14. S. Mukohyama and L. Randall, hep-th/0306108.
15. S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989); A. D. Dolgov, in *Proceedings of the XXIV Rencontre de Moriond; Series: Moriond Astrophysics Meetings*, Ed. by J. Adouse and J. Tran Thanh Van (Les Arcs, France, 1989), p. 227; astro-ph/9708045; P. Binetruy, Int. J. Theor. Phys. **39**, 1859 (2000); V. Sahni and A. Starobinsky, Int. J. Mod. Phys. D **9**, 373 (2000); A. D. Dolgov, hep-ph/0203245; N. Straumann, astro-ph/0203330; T. Padmanabhan, Phys. Rep. **380**, 235 (2003); J. Yokoyama, gr-qc/0305068.
16. S. Mukohyama, hep-th/0306208.
17. A. D. Dolgov and M. Kawasaki, astro-ph/0307442.
18. R. R. Caldwell, R. Davis, and P. J. Steinhardt, Phys. Rev. Lett. **80**, 1582 (1998).
19. A. D. Dolgov and M. Kawasaki, astro-ph/0310822.
20. L. Parker, Phys. Rev. **183**, 1057 (1969); A. A. Grib and S. G. Mamaev, Yad. Fiz. **10**, 1276 (1969) [Sov. J. Nucl. Phys. **10**, 722 (1969)]; Y. B. Zeldovich, Pis'ma Zh. Éksp. Teor. Fiz. **12**, 443 (1970) [JETP Lett. **12**, 307 (1970)]; Y. B. Zeldovich and A. A. Starobinsky, Zh. Éksp. Teor. Fiz. **61**, 2162 (1971) [Sov. Phys. JETP **34**, 1159 (1971)]; N. D. Birrel and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge Univ. Press, Cambridge, 1982; Mir, Moscow, 1984); A. A. Grib, S. G. Mamaev, and V. M. Mostepanenko, *Vacuum Quantum Effects in Strong Fields* (Friedmann Laboratory Publ., St. Petersburg, 1994).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 833–841. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 865–873.*


# On the Short-Distance Potential in Confining Theories*

**V. I. Shevchenko****

*Institute of Theoretical and Experimental Physics,*
*Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received July 15, 2004; in final form, November 1, 2004

**Abstract**—The structure of leading nonperturbative corrections to the static Coulomb potential at small distances in Abelian and non-Abelian theories is analyzed. Related problems of validity of the Dirac quantization condition for running charges in Abelian theory and significance of the quantity $\langle A^2\rangle$ are briefly discussed. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Much attention has been given recently to the question about next-to-leading terms for the static potential in confining theory such as QCD at short distances (see [1–6] and references therein). This problem is closely related to the structure of OPE in confining theories, despite the fact that, to define the potential, one has to go to the large-time limit and therefore leave the region of applicability of the standard OPE [7]. In particular, the appearance of new terms in the so-called SVZ condensate expansion [8] is discussed in [2]. Phenomenologically, the static potential between charge and anticharge is given at small distances as the following expansion:

$$V(r) = \frac{c_{-1}}{r} + c_0 + c_1 r + c_2 r^2 + \mathcal{O}(r^3), \quad (1)$$

where the coefficients $c_i$ may also depend on $r$ (but only logarithmically in four-dimensional renormalizable theories). We discuss some typical patterns below.[1)]

The first example is given by the Abelian charges $(e; -e)$ in infinite space beyond tree level:

$$c_{-1} = -\frac{e^2(r)}{4\pi}; \quad c_1 = c_2 = \ldots = 0. \quad (2)$$

If the Abelian charges $(e; -e)$ are put into a cavity of size $L$, one gets at the tree level [9]

$$c_{-1} = -\frac{e^2}{4\pi}; \quad c_1 = 0; \quad (3)$$
$$c_i = \gamma_i e^2 L^{-(i+1)} \quad \text{for} \quad i \geq 2.$$

For non-Abelian charges $(gT_D^a; -gT_D^a)$ in the representation $D$ of $SU(N)$ interacting with soft nonperturbative fields, the result is given by [10, 11]

$$c_{-1} = -\tilde{\alpha}_s = -C_D(g^2/4\pi); \quad (4)$$
$$c_1 = 0; \quad c_2 = \gamma\tilde{\alpha}_s\langle G^a_{\mu\nu}G^a_{\mu\nu}\rangle T_g; \ldots,$$

where $C_D$ is an eigenvalue of quadratic Casimir operator, $g$ is strong interaction constant, $\langle G^a_{\mu\nu}G^a_{\mu\nu}\rangle$ is nonperturbative gluon condensate [8], and nonperturbative correlation length $T_g$ characterizes the falloff of the gauge-invariant two-point correlator of the field-strength tensors normalized to the gluon condensate at the origin (see review [12] and references therein). The eigenvalue of quadratic Casimir operator reads

$$C_D = \mathrm{Tr}_D T^a_D T^a_D; \quad \mathrm{Tr}_D \mathbf{1}_D = 1. \quad (5)$$

In the expressions above, $\gamma$ remain for some dimensionless numerical factors, whose values are of no relevance for our discussion. What attracts attention in all mentioned cases is the absence of a linear correction proportional to $c_1$. This is also true for nother models [13] which we have not included in the list.

However, one cannot propose strong theoretical arguments why $c_1$ should vanish in the context of QCD or another theory with complicated vacuum structure. Moreover, there are at least two independent sets of lattice data indicating that $c_1 \neq 0$ [14, 15]. Different models explaining small-distance linear correction have been proposed (see [1–6] and references therein). They can be divided into three groups. In the first group, there are models [2, 5, 6] which are based on different corrections to the perturbative two-point correlation function either in the form of gluon propagator $\langle A^a_\mu(x)A^b_\nu(y)\rangle$ or in the form of gauge-invariant path-dependent field strength correlator $\langle \mathrm{Tr}\, G_{\mu\nu}(x)\Phi(x,y)G_{\rho\sigma}(y)\Phi(y,x)\rangle$. The second group is represented by the "dynamical bag" model [1],

*This article was submitted by the author in English.

** e-mail: shevchen@itep.ru

1) In principle, one could consider a situation with noninteger powers of distance $r$ entering (1), but this case is beyond our analysis.

which essentially explores (3) but with $L = L(r)$ instead of a constant. Finally, studies of the static potential at small distances in Abelian confining theories [3] form the third group. We briefly comment on the latter one at the end of the paper, while our attention is focused on the first and second groups. One of the main points of this paper is that, despite many common points, it is possible to make a distinction on the lattice between models of the first and of the second group by studying the small-distance static potential for charges in higher representations of the gauge group.

The paper consists of two relatively independent parts. In the first part (Sections 2 and 3), we discuss some elements of short-distance physics in Abelian theories. In particular, we briefly discuss the following question: does the Dirac quantization condition [16] for bare charges $eg = 2\pi$ stay intact when loop effects are included? There has long been discussion of this question in the literature starting from [17] (see reviews [18, 19] and references therein). An affirmative answer is given and the physics behind it is explained. We will also point the reader's attention in Section 3 to the fact that the local average of the square of the photon field $\langle A_\mu(x)A_\mu(x)\rangle$ can be defined in a covariant gauge in the context of theory with boundaries.

The second part of the paper (Sections 4 and 5) is concentrated on the non-Abelian theory. We demonstrate the difference between dynamical bag models incorporating the Meissner effect and modified propagator models in Section 4, which makes it possible to distinguish these scenarios. Section 5 presents our conclusions.

In the rest of the paper, unless explicitly stated otherwise, we work in Euclidean space with the notation for the four-vectors $k = (k_1, k_2, k_3, k_4)$ and scalar product $kp = k_\mu p^\nu \delta^\mu_\nu$. The three-vectors are denoted as $\mathbf{k} = (k_1, k_2, k_3)$ and the Wick rotation corresponds to the replacement $k_4 \to ik_0$. As usual, $\hbar = c = 1$.

## 2. ELEMENTS OF SHORT-DISTANCE PHYSICS IN ABELIAN THEORIES WITH MONOPOLES

Since we are interested in the physics of interactions at small distances, as a warm-up example, we consider here the theory of Abelian vector field $A_\mu$ interacting with the massive electrically charged matter field and external monopole currents. An integral part of this section is of a review type, since most of the discussed results can be found in the literature. We believe, however, that a consistent presentation of them can be of some use.

After integration over the matter fields, the partition function of effective low-energy theory is given by

$$Z[j^e_\mu] = \int \mathcal{D}A_\mu \times \exp\left(-\frac{1}{2}\int d^dp A_\mu(p)[D^{-1}]^\mu_\nu(p)A^\nu(-p) - S_{\text{eff}}[A] + i\int d^dp j^e_\mu(p)A^\mu(-p)\right), \quad (6)$$

where $S_{\text{eff}}[A]$ contains interaction terms for the field $A_\mu$. The function $[D^{-1}]^\mu_\nu(p)$ can be decomposed as

$$[D^{-1}]^\mu_\nu(p) = \delta^\mu_\nu d_0(p^2) + p^\mu p_\nu d_1(p^2), \quad (7)$$

where the function $d_1(p)$ is gauge-dependent if the theory in question is a gauge theory. In the latter case, the theory can be rewritten in the field-strength formulation [20] as

$$Z[j^e_{\mu\nu}] = \int \mathcal{D}F_{\mu\nu}\delta(\partial_\nu\tilde{F}_{\mu\nu}) \times \exp\left(-\frac{1}{8}\int d^dp F^{\mu\nu}(p)\Delta^{\rho\sigma}_{\mu\nu}(p)F_{\rho\sigma}(-p)\right) \times \exp\left(-S_{\text{eff}}[F] + i\int d^dp j^e_{\mu\nu}(p)F^{\mu\nu}(-p)\right), \quad (8)$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}$ and

$$\Delta^{\rho\sigma}_{\mu\nu}(p) = (\delta^\rho_\mu\delta^\sigma_\nu - \delta^\sigma_\mu\delta^\rho_\nu)\Delta_0(p^2) + (p_\mu p^\rho\delta^\sigma_\nu - p_\mu p^\sigma\delta^\rho_\nu + p_\nu p^\sigma\delta^\rho_\mu - p_\nu p^\rho\delta^\sigma_\mu)\Delta_1(p^2). \quad (9)$$

The relation

$$d_0(p^2) = p^2(\Delta_0(p^2) + p^2\Delta_1(p^2)) \quad (10)$$

provides the correlator's matching between the two formulations.

The field-strength formulation (8) allows one to introduce monopoles into the theory, performing the shift

$$\delta(\partial^\nu\tilde{F}_{\mu\nu}) \to \delta(\partial^\nu\tilde{F}_{\mu\nu} - j^m_\mu). \quad (11)$$

If one may neglect $S_{\text{eff}}[F]$ (see discussion below), straightforward integration gives the resulting monopole partition function

$$Z[j^m_\mu, j^e_{\mu\nu} = 0] = \int \mathcal{D}B_\mu \times \exp\left(-\frac{1}{2}\int d^dp B_\mu(p)\frac{\delta^\mu_\nu p^2 - p^\mu p_\nu}{\Delta_0(p^2)}B^\nu(-p) + i\int d^dp j^m_\mu(p)B^\mu(-p)\right). \quad (12)$$

For the free theory, $d_0(p^2) = p^2$, $\Delta_0(p^2) = 1$, and $\Delta_1(p^2) = 0$, which means that (6) coincides with (12)

up to the interchange of electric and magnetic currents (charges); in other words, pure photodynamics is exactly self-dual. If we switch on the interactions, the inverse propagator gets renormalized and runs with $p^2$ as $d_0(p^2) = p^2(1 + \Pi(p^2))$, where $\Pi(p^2)$ is the corresponding polarization operator. Let us assume for the time being that the dynamics keeps the gauge invariance of the theory intact, which means, in particular, that $\Pi(0) = 0$. When computing the static potential in $d = 3 + 1$ dimensions, one gets formally in the electric case

$$V_{e\bar{e}}(r) = -e^2 \int \frac{d\mathbf{p}}{(2\pi)^3} \frac{1}{d_0(\mathbf{p}^2)} \exp(i\mathbf{p}\cdot\mathbf{r}), \quad (13)$$

while the corresponding expression in the magnetic case reads

$$V_{m\bar{m}}(r) = -g^2 \int \frac{d\mathbf{p}}{(2\pi)^3} \frac{d_0(\mathbf{p}^2)}{\mathbf{p}^4} \exp(i\mathbf{p}\cdot\mathbf{r}), \quad (14)$$

where the charges $e, g$ are physical renormalized (but not running) charges and $r = |\mathbf{r}|$.

It is seen from (13), (14) that the leading corrections to the Coulomb potential have different signs in the electric and magnetic cases:

$$\frac{1}{e^2}\delta V_{e\bar{e}}(r) = -\frac{1}{g^2}\delta V_{m\bar{m}}(r) \quad (15)$$
$$= \int \frac{d\mathbf{p}}{(2\pi)^3} \frac{\Pi(\mathbf{p}^2)}{\mathbf{p}^2} \exp(i\mathbf{p}\cdot\mathbf{r}).$$

Sometimes, this is considered as a proof[2)] of the fact that Dirac quantization condition $eg = 2\pi n$ holds also for running charges (see discussion and references in [19]) if it does for bare charges:

$$e_0 g_0 = eg = e(r)g(r) = 2\pi n, \quad (16)$$

where $e_0, e$, and $e(r)$ denote the bare, the renormalized, and the running charge, respectively. The situation is much more subtle, however. The problem is that, introducing monopoles by (11), one sets the scale of magnetic fields by the large coupling $g \sim 1/e$, since $j^m_\mu$ contains the factor $g \gg 1$. As an immediate result, all higher order irreducible correlators of magnetic fields interacting with the electrically charged matter field described by $S_{\text{eff}}$ can play a role. One should distinguish three different kinematical regions. For large distances $rm \gg 1$, where $m$ is electron mass, the corrections to the static potential are exponentially damped with $r$ for both $V_{e\bar{e}}(r)$ and $V_{m\bar{m}}(r)$. Of main interest is the small-distance region $rm \ll 1$. The leading term in the $k$-point irreducible vertex in $S_{\text{eff}}$, contributing to the monopole potential (14) at small distances, contains the factor $g^k e^k = (2\pi n)^k$, while the analogous contribution to (13) is suppressed as $e^k e^k = e^{2k} \ll 1$. On the other hand, the $k > 4$ correlators, describing the processes of multiphoton scattering, do not have[3)] contributions logarithmically rising with $p^2$. It means that, for large $p^2$ and not too large a value of the product $eg$, one can indeed take the one-loop result (15) as a correct first approximation. With the product $eg$ getting large, however, higher order terms in the non-Gaussian part $S_{\text{eff}}$ become more and more important and their effect might overcome the leading kinematical one-loop logarithm and spoil (16). However, this is not what happens. Since the effective Lagrangian is formed in this case at distances $l \sim r/\sqrt{eg}$, which are for $eg \gg 1$ smaller than the typical scale of the field change $r$ (see, e.g., [21, 22]), the constant field approximation can be used. The Lagrangian in a constant magnetic field is given (we use Minkowski metric conventions here) at the first order by the following expression [23]:

$$L = -\frac{\mathbf{H}^2}{2}\left(1 - \frac{\alpha}{3\pi}\log\frac{e\mathrm{H}}{m^2}\right), \quad (17)$$

where $4\pi\alpha = e^2$ and $\mathrm{H} = |\mathbf{H}|$. Since the pointlike magnetic charge is defined microscopically via divergence of $\mathbf{H}$, $\operatorname{div}\mathbf{H}(\mathbf{r}) = g\delta(\mathbf{r})$, one gets from (17) for the running magnetic charge with logarithmic accuracy

$$g(r) = g\left[1 - \frac{\alpha}{3\pi}\log\frac{1}{(mr)^2}\right]. \quad (18)$$

On the other hand, pointlike electric charge is defined via Maxwell's equation $\operatorname{div}\mathbf{D}(\mathbf{r}) = e\delta(\mathbf{r})$, where $\mathbf{D} = \partial L/\partial\mathbf{E}$, which leads to the standard Uehling and Serber result [24] (again with logarithmic accuracy)

$$e(r) = e\left[1 + \frac{\alpha}{3\pi}\log\frac{1}{(mr)^2}\right]. \quad (19)$$

This is precisely the same answer one can get from (15) with the standard one-loop expression for $\Pi(\mathbf{p}^2)$.

At first glance, the coincidence between (15) and (18), (19) seems rather surprising, since the latter two expressions were obtained from the exact strong-field Lagrangian (17), which sums up an infinite number of graphs, while the former one is a trivial outcome of Gaussian integrations with a one-loop two-point function. The reason for this result is the remarkable correspondence between QED at small distances and

[2)] Actually, only at the next-to-leading order in electric coupling $e^2$, which is assumed to be a small parameter. Notice that $\Pi(p^2)$ is $\mathcal{O}(e^2)$.

[3)] As is well known, at $k = 4$, the amplitude formally diverges logarithmically, but this divergence is exactly cancelled in the sum of all diagrams.

QED in strong fields, studied in the series of papers [22]. It follows directly from the minimal coupling principle for the gauge and charged matter fields: $p_\mu \to P_\mu = p_\mu - eA_\mu$, which gives a hint that large $p_\mu$ and large $eA_\mu$ can describe one and the same physics. The above example shows a rather nontrivial manifestation of this correspondence. It is worth noticing that no topological objects like Dirac strings have been involved in our analysis.

If $\tilde{\Pi}(p^2) \equiv p^2\Pi(p^2)$ does not vanish when $p^2 \to 0$, the theory leaves the Coulomb phase and one cannot introduce external monopoles simply by (11). A well-known example of such a theory is given by the Abelian Higgs model, where the electrically charged field is condensed:

$$L = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}|D_\mu\phi|^2 + \lambda(|\phi|^2 - \eta^2)^2. \quad (20)$$

Attempts to write down an exact expression for the confining potential in this theory encounter a serious problem. The reason for that is a physical one: since the confining string is created between the particles (external monopole–antimonopole pair), its quantum dynamics should be properly taken into account. The confining interaction cannot therefore be described in terms of particle exchanges. Moreover, even in the effective Abelian theory framework, there exists no consistent procedure to perform corresponding calculations analytically in terms of strings at the moment. The best one can do is to compute the Wilson loop for a particular choice of the confining string world-sheet geometry. The conventional choice is the minimal surface $S$ for the given contour $C$, the flat one if the contour is rectangular (as it is for the static potential). In the London limit, the Higgs boson is much heavier than the vector boson, $m_H \gg m = e\eta$, and such a defined potential reads in this limit (see, e.g., [25])

$$V_{m\bar{m}}^{[S]}(r) = g^2 \lim_{T\to\infty}\frac{1}{T} \quad (21)$$
$$\times \int d^4x \int d^4y \left(\frac{1}{4}\Sigma_{\mu\nu}(x)\Delta_m(x-y)\Sigma_{\mu\nu}(y)\right)$$
$$- g^2 \int \frac{d\mathbf{p}}{(2\pi)^3}\frac{\exp(i\mathbf{p}\cdot\mathbf{r})}{\mathbf{p}^2+m^2},$$

where the surface current $\Sigma_{\mu\nu}(x)$ and the kernel $\Delta_m(x-y)$ are given by the expressions

$$\Sigma_{\mu\nu}(x) = \int\limits_S d\sigma_{\mu\nu}(\xi)\delta^{(4)}(x(\xi)-x); \quad (22)$$
$$\Delta_m(x) = \int \frac{d^4p}{(2\pi)^4}\frac{m^2}{p^2+m^2}\exp(ipx).$$

We have indicated by superscript $[S]$ that the potential depends on the profile of the string chosen in (22). At large distances $rm \gg 1$, the potential (21) has linear asymptotics and describes confinement.

We are to address the question of applicability of (21) at small distances. Obviously, expression (21) breaks down at $r \lesssim m_H^{-1}$ since at such small distances the scalar field condensate gets excited and virtual loops of scalar particles contribute to the polarization of the vacuum. If one takes (20) as some low-energy effective theory, the ultraviolet cutoff implicitly enters through higher dimensional operators. Moreover, if the Higgs boson in this effective theory is composite in terms of underlying microscopic theory degrees of freedom, the vector bosons with energy higher than the corresponding binding energy start to resolve the constituents. In either case, the potential (21) gets modified. Let us consider the simplest possible modification, which is to replace $m^2$ in (21), (22) by vector-boson self-energy $\tilde{\Pi}(p^2)$ such that $\tilde{\Pi}(0) = m^2$. Such replacements are in line with (13), (14) and can be justified[4)] in terms of perturbation theory in the electric coupling $e$. Needless to say, the high-momentum asymptotics of $\tilde{\Pi}(p^2)$ will be different for different microscopic scenarios outlined above.

The leading short-distance contribution to (21) is an attractive Coulomb potential, while there are two kinds of subleading corrections:

$$V_{m\bar{m}}^{[r\times T]}(r) = -\frac{g^2}{4\pi r} + \delta V_{m\bar{m}}^{(1)}(r) \quad (23)$$
$$+ \delta V_{m\bar{m}}^{(2)}(r) + \mathcal{O}(g^2e^4).$$

The second correction term comes from the second Yukawa-like term in (21) and formally reads

$$\frac{1}{g^2}\delta V_{m\bar{m}}^{(2)}(r) = \int \frac{d\mathbf{p}}{(2\pi)^3}\frac{\tilde{\Pi}(\mathbf{p}^2)}{\mathbf{p}^4}\exp(i\mathbf{p}\cdot\mathbf{r}). \quad (24)$$

The low-$\mathbf{p}^2$ divergence is artificial and can be easily regularized. The term with the double integral in (21) produces the following correction:

$$\frac{1}{g^2}\delta V_{m\bar{m}}^{(1)}(r) = r\int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{\tilde{\Pi}\left(\mathbf{k}^2/r^2\right)}{\mathbf{k}^2} \quad (25)$$
$$\times \left(\frac{4\sin^2(k_3/2)}{(k_3)^2}\right),$$

where for convenience we put the Wilson contour in the $(3,4)$ plane. It is seen that the dependence of the correction (25) on $r$ is completely determined by the behavior of $\tilde{\Pi}(p^2)$ as a function of $p^2$. In particular, if $\tilde{\Pi}(p^2)$ is such that the $r$ dependence of the integral is weak enough, the correction to the potential will

[4)] Of course, $\tilde{\Pi}(0) = 0$ in the perturbative QED context.

be approximately linear. If $\tilde{\Pi}(p^2) \equiv m^2$, the Yukawa-like correction is linear in $r$ and negative: $\delta V^{(2)}_{m\bar{m}}(r) = -(g^2m^2r)/(8\pi)$ (we have omitted the irrelevant constant term). This is the only correction (up to the change $g \to e$) one would get for the static electric charges without taking into account the string dynamics. The "confining" correction (25) is positive in this case, but logarithmically divergent in the ultraviolet region. This divergence will be cured if the correct form of $\tilde{\Pi}(p^2)$ in the high-momentum regime is taken. Presumably, the relative minus sign between $\delta V^{(1)}_{m\bar{m}}(r)$ and $\delta V^{(2)}_{m\bar{m}}(r)$ persists in this case as well.

It is worth putting an emphasis on the fact that the physics behind $\delta V^{(1)}_{m\bar{m}}(r)$ and $\delta V^{(2)}_{m\bar{m}}(r)$ is very different. In a sense, the former one describes the process of string tension formation, which starts essentially at some ultraviolet scale in this model and goes all the way up to the distance scale given by $m^{-1}$. Since the corresponding formation speed is only logarithmic, one can indeed have an approximately linear potential at small distances—some kind of "ultraviolet confinement," a phenomenon which apparently goes beyond the conventional OPE.

## 3. A COMMENT ON THE QUANTITY $\langle A^2_\mu \rangle$

The linear correction to the potential implies the existence of a local $d = 2$ parameter in the theory, proportional to $c_1$. It cannot be a local polynomial function of original fields of the theory since the only candidate, the average of vector potential squared $\langle \mathrm{Tr} A_\mu(x)^2 \rangle$, is not gauge-invariant. Moreover, there is a gauge defined by the infinitesimal transformation $U(x) = 1 + igA_\mu(x_0)(x - x_0)^\mu + \ldots$ such that ${}^U A^a_\mu(x_0) = 0$; i.e., one always has locally

$$\left\langle \min_{U\in G} \left({}^U A^a_\mu(x)\right)^2 \right\rangle = 0, \tag{26}$$

where $G$ is the gauge group, for Abelian or non-Abelian theory. In the Abelian case, however, the minimum of the integral of $\langle A_\mu(x)^2 \rangle$ over the whole space can be rewritten in terms of gauge-invariant quantities [26]. The physical relevance of such a non-local object for the non-Abelian case is under debate [27, 28].

We would like to point the reader's attention to the fact that, though it seems to be impossible to assign any gauge-invariant meaning to the local quantity $\langle A^2 \rangle$, one can have in some cases a $\xi$-independent definition of $\langle A^2 \rangle$, where $\xi$ is the covariant gauge-fixing parameter. It happens when one studies the theories which contain some external parameter of the dimension of mass/length and if the part of the effective action depending on this parameter does not depend on $\xi$. By way of example, let us study photodynamics with the Casimir-type static boundary conditions (see review [29] and references therein):

$$Z[j^e_\mu] = \int \mathcal{D}A_\mu \delta[BC] \tag{27}$$
$$\times \exp\left( - \int d^dx \left( \frac{1}{4} F_{\mu\nu}(x) F^{\mu\nu}(x) \right.\right.$$
$$\left.\left. + \frac{1}{2\xi} \left(\partial_\mu A_\nu(x)\right)^2 + i j^e_\mu(x) A^\mu(x) \right)\right),$$

where the boundary conditions read

$$\delta[BC] = \delta(n_\mu \tilde{F}_{\mu\nu}(x_1, x_2, x_3 = a_1, x_4))$$
$$\times \delta(n_\mu \tilde{F}_{\mu\nu}(x_1, x_2, x_3 = a_2, x_4)).$$

Here, $n_\mu = (0, 0, 1, 0)$ is a unit vector in the $x_3$ direction and $a_1$, $a_2$ mark the $x_3$ positions of parallel infinitely thin ideally conducting plates. The distance between plates is given by $a = |a_2 - a_1|$.

The gauge field propagator can be readily obtained from (27). It consists of two parts:

$$\langle A_\mu(x) A_\nu(y) \rangle = D_{\mu\nu}(x, y) = D^{(0)}_{\mu\nu}(x - y; \xi)$$
$$+ \bar{D}_{\mu\nu}(x, y; a_1, a_2),$$

where the term $D^{(0)}_{\mu\nu}(x - y; \xi)$ is the standard gauge-dependent tree-level photon propagator, while the term $\bar{D}_{\mu\nu}(x, y; a_1, a_2)$ encodes the information about the boundaries. The exact form of the latter was first obtained in [30]. The function $D^{(0)}_{\mu\nu}(x - y; \xi)$ is translation-invariant but gauge-dependent, while the function $\bar{D}_{\mu\nu}(x, y; a_1, a_2)$ depends on $x_3$ and $y_3$ separately, but is $\xi$-independent (not to be confused with the gauge-invariance!). Moreover, it has a finite limit when $x$ approaches $y$. Therefore, one can address the issue of the $\langle A_\mu(x) A^\mu(x) \rangle$ condensate in the Casimir vacuum exactly in the same way as one computes the energy density in this problem, namely, subtracting the boundary-independent part:

$$\langle A^2(x) \rangle \stackrel{\text{def}}{=} \lim_{y\to x} \Big( \langle A_\mu(x) A^\mu(y) \rangle \tag{28}$$
$$- \langle A_\mu(x) A^\mu(y) \rangle^{(0)} \Big) = \bar{D}^\mu_\mu(x, x; a_1, a_2).$$

It is convenient to introduce the notation $z = a_1 + a_2 - 2x_3$. Then, using the exact expression for $\bar{D}_{\mu\nu}(x, y; a_1, a_2)$, one gets

$$\langle A^2(z) \rangle = \frac{1}{12a^2} \left[ 1 + \frac{3}{2} \tan^2\left( \frac{\pi z}{2a} \right) \right]. \tag{29}$$

Such a defined local two-dimensional "condensate" is $\xi$-independent, strictly positive, and diverges on the boundaries. In a completely analogous way, one can define the $\langle A^2 \rangle$ condensate at finite temperature.

In the context of the Casimir problem, one is usually interested in studying the changes in physical quantities like components of the energy–momentum tensor with respect to their vacuum values. The above example demonstrates that a "nonphysical" quantity like $\langle A^2\rangle$ is also nontrivially modified. It is worth noting that (29) is not related to any linear correction to the potential, since the potential for a dipole between the mirrors is of the form (3) with $L = a$.

## 4. DYNAMICAL BAG VERSUS MODIFIED PROPAGATOR

One of the main physical assumptions of the SVZ framework can be formulated as follows: the nonperturbative vacuum of the theory is not perturbed by the external sources. In the case of large virtualities, it is sometimes justified using the language of condensed matter physics: since the time scale of the hard process we wish to study is much smaller than typical relaxation times characterizing such a medium as a nonperturbative QCD vacuum, it is reasonable to suggest that the vacuum state remains unchanged.

It is of interest to relax this assumption and to study the corresponding physics. An example of the situation when one has to do it is provided by the monopole–antimonopole pair in the superconductor. It is well known that the condensate must be broken along some line connecting the particles, no matter how small their charge is (see review [19] and references therein). The reason for that, eventually, is the nonperturbative nature of the interaction between magnetic and electric particles.

First, we rederive the results of the dynamical bag, or dynamical cavity, model [1] in a more quantitative way. After that, we will come to the modified propagator model [2]. Since we consider gluodynamics in this section, we switch to dual terminology, so that particles which are confined (quarks) carry electric and not magnetic charge. The microscopic description of confinement as the dual Meissner effect via monopole condensation refers to the Abelian projection procedure [31]. We adopt the physical picture of confinement as the dual Meissner effect but will make no use of the Abelian Higgs model language in this section. As is well known, the energy density of nonperturbative fields is given by the energy–momentum tensor trace anomaly [32]:

$$\epsilon = \frac{1}{4}\langle\theta_{\mu\mu}\rangle = \frac{\beta(\alpha_s)}{16\alpha_s}\langle G^a_{\mu\nu}(0)G^a_{\mu\nu}(0)\rangle, \qquad (30)$$

where

$$\beta(\alpha_s) = \mu^2\frac{d\alpha_s(\mu^2)}{d\mu^2} \qquad (31)$$

$$= -\left(\frac{11}{3}N - \frac{2}{3}N_f\right)\frac{\alpha_s^2}{2\pi} + \mathcal{O}(\alpha_s^3).$$

In principle, the average (30) depends on the actual state of the system; for example, one can study its density or temperature dependence. In terms of invariant functions $D(z^2)$ and $D_1(z^2)$, the vacuum gluon condensate is given as [10]

$$\langle g^2 \mathrm{Tr} G_{\mu\nu}(0)G_{\mu\nu}(0)\rangle = 24(D(0) + D_1(0)), \qquad (32)$$

where only the function $D(z^2)$ is responsible for confinement. In particular, the deconfinement transition is characterized by the condition $D(z^2) \equiv 0$, with the smooth behavior of $D_1(z)$ over the phase transition [33].

One can raise a question about the status of (30) in a color field of an external source. Owing to $SU(3)$ and $O(4)$ invariance, such an average can be defined as

$$G_2(\mathbf{r}) = \left\langle \frac{\alpha_s}{\pi} G^a_{\mu\nu}(\mathbf{r})G^a_{\mu\nu}(\mathbf{r})\right\rangle \qquad (33)$$

$$= \frac{g^2}{2\pi^2}\frac{\left\langle \mathrm{Tr}\,(G_{\mu\nu}(\mathbf{r})G_{\mu\nu}(\mathbf{r}))\cdot \mathrm{Tr\,P}\exp i\int_C A_\rho dz_\rho\right\rangle}{\left\langle \mathrm{Tr\,P}\exp i\int_C A_\rho dz_\rho\right\rangle},$$

where $C$ stays for the (closed) world line of the source. If the point $\mathbf{r}$ is far enough from $C$, the average reaches the vacuum value: $G_2(\mathbf{r}) \to G_2$. Correspondingly, one defines $\epsilon(\mathbf{r}) = \epsilon(\mathbf{E}(\mathbf{r}))$ via (30). To the best of the author's knowledge, no systematic lattice analysis of the behavior of $\epsilon(\mathbf{r})$ in a strong external field has been performed. This problem is of direct relevance for our discussion. Namely, let us put a single static charge in the nonperturbative QCD vacuum. Due to the confinement property, there should be an anticharge somewhere in the space to end the confining string, and therefore the field distribution will not be spherically symmetric. Suppose, however, that we look at the gluon fields in a thin spherical layer between $r$ and $r + \Delta r$, where $r$ is smaller than the typical nonperturbative nonlocality scale given by $T_g$ (it is worth mentioning that the same quantity $T_g$ defines the width of the confining string). Then it is natural to assume that the Faraday flux lines of the charge are affected only weakly by the opposite charge sitting on the other end of the string. On the other hand, they are affected by the nonperturbative vacuum and act back on it. Correspondingly, we take the energy of the gluon fields in the volume $V$ as a sum of perturbative and nonperturbative parts

$$\mathcal{E} = \mathcal{E}_p + \mathcal{E}_{np} = \frac{1}{2}\int\limits_V d\mathbf{r}\mathbf{E}^a(\mathbf{r})\cdot\mathbf{E}^a(\mathbf{r}) + \int\limits_V d\mathbf{r}\epsilon(\mathbf{r}), \qquad (34)$$

where $\mathbf{E}^a(\mathbf{r})$ is the perturbative electric field of the charge. In principle, the coordinate dependence of

$\mathbf{E}^a(\mathbf{r})$ and $\epsilon(\mathbf{r}) = \epsilon(\mathbf{E}(\mathbf{r}))$ on $\mathbf{r}$ can be very complex, resembling the formation of an intermediate state in the vicinity of the charge, as happens in ordinary superconductors. It can be energetically advantageous to squeeze slightly perturbative flux lines in one part of the volume[5)] and to rarefy them in the rest of it [this increases the first term in (34)], but compensate the excess of energy by the nonperturbative energy gain from the rest of the volume, which can be possible if $\epsilon(\mathbf{r})$ decreases towards its vacuum value when $\mathbf{E}^a(\mathbf{r})$ decreases in this part of the volume. This is the essence of the Meissner effect. Let us simplify the matter further by approximating (34) as

$$\Delta\mathcal{E} = \frac{1}{2}\left[C_D\left(\frac{g}{4\pi r^2}\right)^2\right]\left(\frac{4\pi r^2}{S_1(r)}\right)^2 \tag{35}$$
$$\times\, S_1(r)\Delta r + \epsilon_1 S_1(r)\Delta r + \epsilon_2 S_2(r)\Delta r + \mathcal{O}(\Delta r^2).$$

We have taken $\epsilon(\mathbf{r})$ as constants $\epsilon_1$ and $\epsilon_2$ in the parts of the volume $S_1\Delta r$ and $S_2\Delta r$, respectively. Notice that $S_1(r) + S_2(r) = 4\pi r^2$. Expression (35) assumes that the parts of the space occupied by the nonperturbative fields with the (vacuum) energy density $\epsilon_2$ completely expel the perturbative field flux lines into the regions where the energy density of nonperturbative fields is given by $\epsilon_1$. The rescaling of the electric perturbative field of the point charge reflects the flux conservation.

Since $S_1(r)$ can be never greater than $4\pi r^2$, there exists a critical radius [corresponding to the stationary point of (35)]

$$r_c^4 = \frac{\tilde{\alpha}_s}{8\pi(\epsilon_1 - \epsilon_2)}. \tag{36}$$

When $r$ reaches $r_c$ from above, the region of unperturbed vacuum $S_2$ shrinks to zero. Physically, the model describes the formation of a bag of the radius $r_c$, where perturbative fields are strong enough to change the nonperturbative vacuum state significantly. We are not addressing an interesting question at the moment of whether there is a local deconfinement transition (in any sense of the word) in this volume (which is natural to think of as being based on an Abelian Meissner effect analogy). Let us estimate $r_c$ numerically. We use different sets of data [34] and take for the gluon condensate

$$G_2 = \left\langle \frac{\alpha_s}{\pi} G^a_{\mu\nu} G^a_{\mu\nu} \right\rangle = 0.014{-}0.026\ \mathrm{GeV}^4. \tag{37}$$

Hence, for the nonperturbative vacuum energy in one loop choosing $N = 3$, $N_f = 2$, we obtain (remember that $\epsilon_2$ coincides with the vacuum energy density $\epsilon$ in our approach)

$$\epsilon_1 - \epsilon_2 = \kappa\epsilon - \epsilon = (1-\kappa)\,(4{-}8) \times 10^{-3}\ \mathrm{GeV}^4. \tag{38}$$

The value of the positive factor $\kappa = \epsilon_1/\epsilon_2 < 1$ is unknown and enters as a free parameter. Taking $\tilde{\alpha}_s = 0.47$ (corresponding to the charges in the fundamental representation and $\alpha_s(M_\tau)$), we finally get

$$r_c = (1-\kappa)^{-1/4}(0.2{-}0.3)\ \mathrm{fm}. \tag{39}$$

Unless $\kappa$ is unnaturally close to unity, this rather small value of $r_c$ is compatible with the discussed picture—were it much larger than $T_g$, the analysis would be meaningless.

Let us, following [1], consider now a small color dipole of the size $R$. The expression for the energy (34) stays intact, but the physics of critical distance $r_c$ is different. Since there is no flux of a perturbative field through the surface of the volume with the dipole inside, it is advantageous to confine flux lines inside some finite volume. Suppose that we have chosen the volume of some fixed size $L \gtrsim R$ around the dipole such that vacuum energy density $\epsilon_2 = \epsilon$ outside is reduced to $\epsilon_1 > \epsilon_2$ inside it. The perturbative electric field felt on the boundary of this volume is proportional to $R/L^3$. It is important that it go to zero when $R$ decreases. It means that, if we make $R$ small enough, the total energy can also be lowered by appropriate decreasing of $L$, i.e., $L = L(R)$. Indeed, when we go from $L$ to $L - \Delta L$, the perturbative energy increase due to the rearrangement of the perturbative field is given by

$$\Delta\mathcal{E}_p = \frac{1}{2}C_D\left(\frac{g}{4\pi}\right)^2 \tag{40}$$
$$\times \int\limits_{L-\Delta L}^{L} d\mathbf{r}\left(\frac{\mathbf{R}^2}{\mathbf{r}^6} + 3\frac{(\mathbf{R}\cdot\mathbf{r})^2}{\mathbf{r}^8}\right)$$
$$= \tilde{\alpha}_s\frac{R^2}{L^4}\Delta L + \mathcal{O}(\Delta L^2),$$

while the nonperturbative energy gain is

$$\Delta\mathcal{E}_{np} = 4\pi L^2(\epsilon_1 - \epsilon_2)\Delta L + \mathcal{O}(\Delta L^2). \tag{41}$$

This defines the stationary point

$$L_c = \left(\frac{\tilde{\alpha}_s R^2}{4\pi(\epsilon_1 - \epsilon_2)}\right)^{1/6}. \tag{42}$$

We can now estimate the nonperturbative correction to the energy of the dipole as

$$\delta\mathcal{E}(R) \approx \tilde{\alpha}_s\frac{R^2}{3L_c^3} + (\epsilon_1 - \epsilon_2)\frac{4\pi}{3}L_c^3 \tag{43}$$

[5)] Since we are working at small distances, the total perturbative flux (i.e., the number of the flux lines) is conserved. More accurately, because of asymptotic freedom it is not, but its violation due to asymptotic freedom is a weak logarithmic effect.

$$= \frac{4}{3}\sqrt{\pi\tilde{\alpha}_s(\epsilon_1-\epsilon_2)}R.$$

This correction is linear in $R$, as one can alternatively see, inserting (42) into (3). Numerically, we get

$$\frac{4}{3}\sqrt{\pi\tilde{\alpha}_s(\epsilon_1-\epsilon_2)} = (1-\kappa)^{1/2}(0.10{-}0.14)\ \mathrm{GeV}^2. \quad (44)$$

The discussed picture reminds one of the old MIT bag model [35]. This similarity is formal, in some sense. While in the MIT bag model the bag has the typical hadron size, the dynamical cavity of the size $L_c(R)$ given by (42) is an object which has physical meaning only at very small distances. Since $L_c \sim R^{1/3}$, at small enough $R$, we are always in the $T_g \gtrsim L_c \gtrsim R$ regime, where all the picture is meaningful. With $R$ and $L_c$ rising and reaching $T_g$, the confining string formation process starts and we leave the domain of qualitative applicability of the model.

We can now come to the modified propagator models [2, 5, 6]. Actually, their analysis is simpler, and we take as an example the model of [2]. It is suggested that one modify the one-gluon exchange propagator by adding a "tachyon"-mass term to the gluon, i.e. (in Minkowski metric and Feynman gauge),

$$\langle A^a_\mu(k)A^b_\nu(-k)\rangle = \eta_{\mu\nu}\delta^{ab}\frac{1}{k^2} \to \eta_{\mu\nu}\delta^{ab}\left(\frac{1}{k^2}+\frac{\lambda^2}{k^4}\right). \quad (45)$$

Notice that (45) is a large-momentum expansion, so one actually never gets closer to the artificial tachyon pole. The analysis of the existing phenomenology leads to an estimate $\lambda^2 \approx -0.5\ \mathrm{GeV}^2$ for the "tachyon" mass [2]. However, the physical origin of the $\lambda^2$ term in the context of QCD has not been clarified. Alternative scenarios suggest a $\mu^2/z^2$ contribution[6] to the function $D_1(z^2)$ [5], strong coupling constant freezing [6], and infrared renormalons [1]. It is easy to see that all these proposals lead to the Casimir scaling law for the corresponding correction; for example, with expression (45), one gets for the static potential between charges in the representation $D$ of the gauge group

$$V_D(r) + \mathrm{const} = -\tilde{\alpha}_s\left(\frac{1}{r}+\frac{\lambda^2 r}{2}\right) + \mathcal{O}(r^2), \quad (46)$$

where $\tilde{\alpha}_s = C_D\alpha_s = C_D(g^2/4\pi)$; i.e., $V_D(r)$ demonstrates Casimir scaling.

The magnitudes of corrections are close to each other in both models; for example, taking value (44), one gets in terms of (45) $\lambda^2 = -(1-\kappa)^{1/2}(0.4{-}0.6)\ \mathrm{GeV}^2$, which is surprisingly close to the value found in [2]. The important difference, however, is a different $C_D$ dependence. The remarkable property of (43) is its square-root dependence on the eigenvalue of Casimir operator $C_D$. This result is known in the context of the MIT bag model [36], where it takes place for the slope of the confining linear potential. As such, the square-root law was definitely ruled out by recent precise studies of the static potential at large distances in different representations of the gauge group [37, 38]. Instead, the Casimir scaling phenomenon (proportionality of $V_D(r)$ to $C_D$) was confirmed (see [39] for a review). However, there exist no calculations of static potential for charges in higher representations at small distances. Such a lattice simulation will confront in a nontrivial way (43) and (46).

From the general point of view, the Casimir scaling and $\sqrt{C_D}$ law for the leading small-distance correction would correspond to rather different physical pictures. Both cases physically describe perturbative–nonperturbative interference and require modification of the standard condensate ideology in QCD. However, in the latter case, this modification must be by far more radical. Roughly speaking, the former case corresponds to a new kind of nonperturbative corrections to the perturbative propagator, which are "harder" than those studied before, while the latter case indicates that the nonperturbative vacuum structure itself is strongly affected by the perturbative field, when it reaches some critical value.

To be self-contained, let us briefly mention the analysis of the classical Abelian Higgs model on the shortest distances performed in [3]. The small-distance correction to the potential was found numerically and fitted by a linear function with a good accuracy. This behavior was linked to the Dirac veto, i.e., the condition for the charged condensate $\phi$ to vanish along some line (the Dirac string) between the dual charges. The problem mentioned in Section 2 still persists in this case, because the position of the Dirac string can be chosen in an arbitrary way and the answer for the short-distance potential depends on this choice. Moreover, it is not quite clear what the Dirac string of effective Abelian theory corresponds to in terms of the original non-Abelian theory. In the dynamical cavity model, the linear behavior (43) also results from the Meissner effect but without any special role of singularities like Dirac strings. Roughly speaking, the effect in [3] comes from the infinitely thin line between the charges, while in the discussed case all the volume of the size $L_c$ contributes to it.

[6] It can be shown that the $1/z^2$ term in the function $D(z^2)$ would produce ultraviolet divergence on the world sheet of the confining string, while such a term in $D_1(z^2)$ is safe.

## 5. CONCLUSION

There are a few models in the literature that aim to reproduce the linear dependence on distance of the leading nonperturbative correction to the static potential at small distances. Comparing different models, we call the reader's attention to the fact that a possible way to test them is to check the dependence of this correction on the eigenvalue of the quadratic Casimir operator for higher representations of the charges. Such a computation can shed new light on strong interaction physics and the problem of confinement.

## ACKNOWLEDGMENTS

It is a pleasure for me to express my deep respect to Prof. Yurii Antonovich Simonov, the physicist and the man whose quest for the truth in science and life constantly inspires people around him.

Partial support from the Federal Program of the Russian Ministry of Industry, Science, and Technology 40.052.1.1.1112 and from the grant for Scientific Schools NS-1774.2003.2 is acknowledged. I am grateful to the nonprofit Dynasty Foundation and ICFPM for financial support.

## REFERENCES

1. R. Akhoury and V. I. Zakharov, Phys. Lett. B **438**, 165 (1998).
2. K. G. Chetyrkin, S. Narison, and V. I. Zakharov, Nucl. Phys. B **550**, 353 (1999).
3. M. N. Chernodub, F. V. Gubarev, M. I. Polikarpov, and V. I. Zakharov, Phys. Lett. B **475**, 303 (2000); F. V. Gubarev, M. I. Polikarpov, and V. I. Zakharov, Nucl. Phys. B (Proc. Suppl.) **86**, 437 (2000).
4. F. V. Gubarev, M. I. Polikarpov, and V. I. Zakharov, hep-th/9812030.
5. Yu. A. Simonov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 471 (1999) [JETP Lett. **69**, 505 (1999)]; Phys. Rep. **320**, 265 (1999).
6. A. M. Badalian and D. S. Kuzmenko, Phys. Rev. D **65**, 016004 (2002).
7. K. G. Wilson, Phys. Rev. **179**, 1499 (1969); Phys. Rev. D **3**, 1818 (1971); K. G. Wilson and J. Kogut, Phys. Rep. **12**, 75 (1974).
8. M. Shifman, A. Vainshtein, and V. Zakharov, Nucl. Phys. B **147**, 385, 448 (1979).
9. H. B. G. Casimir and D. Polder, Phys. Rev. **73**, 360 (1948).
10. H. G. Dosch, Phys. Lett. B **190**, 177 (1987); H. G. Dosch and Yu. A. Simonov, Phys. Lett. B **205**, 339 (1988); Yu. A. Simonov, Nucl. Phys. B **307**, 512 (1988).
11. Yu. A. Simonov, S. Titard, and F. J. Yndurain, Phys. Lett. B **354**, 435 (1995).
12. A. Di Giacomo, H. G. Dosch, V. I. Shevchenko, and Yu. A. Simonov, Phys. Rep. **372**, 319 (2002).
13. M. B. Voloshin, Nucl. Phys. B **154**, 365 (1979); H. Leutwyler, Phys. Lett. B **98B**, 447 (1981); J. S. Bell and R. A. Bertlmann, Nucl. Phys. B **187**, 285 (1981); I. I. Balitsky, Nucl. Phys. B **254**, 166 (1985).
14. G. Burgio, F. Di Renzo, G. Marchesini, and E. Onofri, Phys. Lett. B **422**, 219 (1998).
15. G. S. Bali, Phys. Lett. B **460**, 170 (1999).
16. P. A. M. Dirac, Phys. Rev. **74**, 817 (1948).
17. J. Schwinger, Phys. Rev. **144**, 1087 (1966); **173**, 1536 (1968).
18. M. Blagojevic and P. Senjanovic, Phys. Rep. **157**, 233 (1988).
19. M. N. Chernodub, F. V. Gubarev, M. I. Polikarpov, and V. I. Zakharov, Yad. Fiz. **64**, 615 (2001) [Phys. At. Nucl. **64**, 561 (2001)].
20. M. Halpern, Phys. Rev. D **19**, 517 (1979).
21. G. J. Goebel and M. T. Thomas, Phys. Rev. D **30**, 823 (1984).
22. V. I. Ritus, Zh. Éksp. Teor. Fiz. **69**, 1517 (1975) [Sov. Phys. JETP **42**, 774 (1975)]; **73**, 807 (1977) [**46**, 423 (1977)]; hep-th/9812124.
23. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936).
24. R. Serber, Phys. Rev. **48**, 49 (1935); E. Uehling, Phys. Rev. **48**, 55 (1935).
25. E. Akhmedov, M. Chernodub, M. Polikarpov, and M. Zubkov, Phys. Rev. D **53**, 2087 (1996).
26. F. V. Gubarev, L. Stodolsky, and V. I. Zakharov, Phys. Rev. Lett. **86**, 2220 (2001); F. V. Gubarev and V. I. Zakharov, Phys. Lett. B **501**, 28 (2001).
27. M. J. Lavelle and M. Schaden, Phys. Lett. B **208**, 297 (1988).
28. P. Boucaud *et al.*, Phys. Lett. B **493**, 315 (2000); L. Stodolsky, P. van Baal, and V. I. Zakharov, Phys. Lett. B **552**, 214 (2003).
29. M. Bordag, U. Mohideen, and V. M. Mostepanenko, Phys. Rep. **353**, 1 (2001).
30. M. Bordag, D. Robaschik, and E. Wieczorek, Ann. Phys. (N.Y.) **165**, 192 (1985).
31. G. 't Hooft, Nucl. Phys. B **138**, 1 (1978).
32. J. Collins, A. Duncan, and S. Joglekar, Phys. Rev. D **16**, 438 (1977).
33. A. Di Giacomo, M. D'Elia, H. Panagopoulos, and E. Meggiolaro, hep-lat/9808056.
34. M. D'Elia, A. Di Giacomo, and E. Meggiolaro, Phys. Lett. B **408**, 315 (1997); S. Narison, Phys. Lett. B **387**, 162 (1996).
35. A. Chodos *et al.*, Phys. Rev. D **9**, 3471 (1974).
36. K. Johnson and C. B. Thorn, Phys. Rev. D **13**, 1934 (1976); T. H. Hansson, Phys. Lett. B **166B**, 343 (1986).
37. G. S. Bali, Nucl. Phys. B (Proc. Suppl.) **83**, 422 (2000); Phys. Rev. D **62**, 114503 (2000).
38. S. Deldar, Phys. Rev. D **62**, 034509 (2000); Nucl. Phys. B (Proc. Suppl.) **73**, 587 (1999).
39. V. I. Shevchenko and Yu. A. Simonov, Int. J. Mod. Phys. A **18**, 127 (2003).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 842–860. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 874–893.*


# Nonperturbative Dynamics of Scalar Field Theories through the Feynman–Schwinger Representation*

**C. Savkli[1], F. Gross[2], and J. Tjon[2,3]**

Received May 13, 2004; in final form, October 29, 2004

**Abstract**—We present a summary of results obtained for scalar field theories using the Feynman–Schwinger (FSR) approach. Specifically, scalar QED and $\chi^2\phi$ theories are considered. The motivation behind the applications discussed in this paper is to use the FSR method as a rigorous tool for testing the quality of commonly used approximations in field theory. Exact calculations in a quenched theory are presented for one-, two-, and three-body bound states. Results obtained indicate that some of the commonly used approximations, such as Bethe–Salpeter ladder summation for bound states and the rainbow summation for one-body problems, produce significantly different results from those obtained from the FSR approach. We find that more accurate results can be obtained using other, simpler, approximation schemes. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

In the study of hadronic physics, one has to face the problem of determining the quantum-dynamical properties of physical systems in which the interaction between the constituents is of a nonperturbative nature. In particular, such systems support bound states and clearly nonperturbative methods are needed to describe their properties. Assuming that such systems can be described by a field theory, one has to rely on some approximation scheme. One common approximation is known as perturbation theory. Perturbation theory involves making an expansion in the coupling strength of the interaction. The Green's function in field theory can be expanded in powers of the coupling strength. In order to be able to obtain a bound-state result, one must sum the interactions to all orders. Most practical calculations to date have been done within the Bethe–Salpeter framework, where the resulting kernel is perturbatively truncated. In this paper, we will discuss another method, one which is based on the path integral formulation of Feynman and Schwinger [1, 2]. It was initiated by Simonov and collaborators [3–9] in their study of quantum chromodynamics (QCD).

With the discovery of QCD, nonperturbative calculations in field theory have become even more essential. It is known that the building blocks of matter, quarks and gluons, only exist in bound states. Therefore, any reaction that involves quarks will necessarily involve bound states in the initial and/or final states. This implies that, even at high-momentum transfers, where QCD is perturbative, formation of quarks into a bound state necessitates a nonperturbative treatment. Therefore, it is essential to develop new methods for doing nonperturbative calculations in field theory.

The plan of this article is as follows. In the following section, a brief review of the particle trajectory method in field theory will be given. The Feynman–Schwinger representation will be introduced through applications to scalar fields. In particular, the emphasis will be on comparing various nonperturbative results obtained by different methods. It will be shown with examples that nonperturbative calculations are interesting and exact nonperturbative results could significantly differ from those obtained by approximate nonperturbative methods.

Results for scalar quantum electrodynamics are discussed in Section 3. In particular, a cancellation is found between the vertex corrections and self-energy contributions, so that the exact result can be described by essentially the sum of only the generalized exchange ladders. In Section 4, the bound states are obtained for a $\chi^2\phi$ theory in a quenched approximation. Exact results are presented for the binding energies of two- and three-body bound states and compared with relativistic quasi-potential predictions. Section 5 deals with the stability of the $\chi^2\phi$ theory. It is argued that the quenched theory does not suffer from the well-known instability of a $\phi^3$ theory. The paper closes with some concluding remarks.

*This article was submitted by the authors in English.

[1]Lockheed Martin Space Operations, Greenbelt, USA.
[2]Jefferson Lab, Newport News, USA.
[3]Department of Physics, University of Maryland, College Park, USA.

1063-7788/05/6805-0842$26.00 

## 2. FEYNMAN–SCHWINGER REPRESENTATION

Nonperturbative calculations can be divided into two general categories: (i) integral equations and (ii) path integrals. Integral equations have been used for a long time to sum interactions to all orders with various approximations [10–13]. In general, a complete solution of field theory to all orders can be provided by an infinite set of integral equations relating vertices and propagators of the theory to each other. However, solving an infinite set of equations is beyond our reach and usually integral equations are truncated by various assumptions about the interaction kernels and vertices. The most commonly used integral equations are those that deal with few-body problems. The Bethe–Salpeter equations [10] are the starting point of those investigations. Approximation schemes have extensively been studied, where in addition to solutions of the Bethe–Salpeter equation [14–18] also three-dimensional reductions [19–24] have been explored for the $N$-particle free particle Green's function. In most calculations, the ladder approximation has been used for the kernel of the resulting equations. Issues of convergence of these schemes remain. Therefore it is important to have exact solutions of field-theory models available to test these approximations. One promising way to reconstruct exact solutions of field theory is the path-integral method.

Path integrals provide a systematic method for summing interactions to all orders. The Green's function in field theory is given by the path-integral expression

$$\langle 0|T[\phi(x_1)\phi(x_2)\cdots\phi(x_n)]|0\rangle \qquad (1)$$
$$= \frac{\int [\mathcal{D}\phi]\phi(x_1)\phi(x_2)\cdots\phi(x_n)\exp\left[i\int d^4x L(x)\right]}{\int [\mathcal{D}\phi]\exp\left[i\int d^4x L(x)\right]}.$$

While path integrals provide a compact expression for the exact nonperturbative result for propagators, evaluation of the path integral is a nontrivial task.

In general, field-theoretical path integrals must be evaluated by numerical integration methods, such as Monte Carlo integration. The best known numerical integration method is lattice gauge theory. Lattice gauge theory involves a discretization of space–time and numerical integrations over field configurations are carried out using Monte Carlo techniques. A more efficient method of performing path integrals in field theory has been proposed and it consists of explicitly integrating out the fields. It has been demonstrated to be highly successful for the case of simple scalar interactions [6, 25–33]. This method is known as the Feynman–Schwinger representation (FSR). Through applications of the FSR, the importance of *exact* nonperturbative calculations will be shown with explicit examples.

The basic idea behind the FSR approach is to transform the field-theoretical path integral (1) into a quantum-mechanical path integral over *particle trajectories*. When written in terms of trajectories, the exact results decompose into separate parts and permit us to study the individual role and numerical size of exchange, self-energy, and vertex corrections. This, in turn, alows us to study different approximations to field theory and, in some cases, prove new results.

To illustrate these ideas, we consider the application of the FSR technique to scalar QED. The Minkowski metric expression for the scalar QED Lagrangian in Stueckelberg form is given by

$$\mathcal{L}_{\mathrm{SQED}} = -m^2\chi^2 - \frac{1}{4}F^2 + \frac{1}{2}\mu^2A^2 \qquad (2)$$
$$+ (\partial_\mu - ieA_\mu)\chi^*(\partial^\mu + ieA^\mu)\chi - \lambda\frac{1}{2}(\partial\cdot A)^2,$$

where $A^\mu$ is the gauge field of mass $\mu$, $\chi$ is the charged field of mass $m$, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the gauge field tensor, and, for example, $A^2 = A_\mu A^\mu$. The presence of a mass term for the exchange field breaks gauge invariance and was introduced in order to avoid infrared singularities that arise when the theory is applied in $0+1$ dimensions. For dimensions larger than $n = 2$, the infrared singularity does not exist and therefore the limit $\mu \to 0$ can be safely taken to insure gauge invariance.

The path integral is to be performed in Euclidean metric. Therefore, we perform a Wick rotation:

$$\exp\left[i\int d^4x\mathcal{L}_{\mathrm{M}}\right] \longrightarrow \exp\left[-\int d^4x\mathcal{L}_{\mathrm{E}}\right]. \qquad (3)$$

The Wick rotation for coordinates is obtained by

$$x_0 \to -ix_0, \quad \partial_0 = \frac{\partial}{\partial x^0} \to i\partial_0. \qquad (4)$$

The transformation of field $A$ under Wick rotation is found by noting that, under a gauge transformation,

$$A_\mu \to A_\mu + \partial_\mu\Lambda. \qquad (5)$$

Then, under a Wick rotation,

$$A_0 \to iA_0, \qquad (6)$$

and the Wick-rotated Lagrangian for SQED becomes

$$\mathcal{L}_{\mathrm{SQED}} = \chi^*[m^2 - \partial^2 - 2ieA\cdot\partial \qquad (7)$$
$$- ie\partial\cdot A + e^2A^2]\chi + \mathcal{L}_A.$$

The exchange field part of the Lagrangian is given by

$$\mathcal{L}_A \equiv \frac{1}{2}A_\mu(\mu^2 g_{\mu\nu} - \lambda\partial_\mu\partial_\nu)A_\nu + \frac{1}{4}F^2 \qquad (8)$$

$$= \frac{1}{2} A_\mu \Big[ (\mu^2 - \Box) g_{\mu\nu} + (1-\lambda)\partial_\mu \partial_\nu \Big] A_\nu.$$

We employ the Feynman gauge $\lambda = 1$, which yields

$$\mathcal{L}_A = \frac{1}{2} A_\nu (\mu^2 - \Box) A_\nu. \tag{9}$$

The two-body Green's function for the transition from an initial state $\Phi_i(x, \bar{x})$ to final state $\Phi_f(y, \bar{y})$ is given by

$$G(y, \bar{y}|x, \bar{x}) \tag{10}$$
$$= N \int \mathcal{D}\chi^* \int \mathcal{D}\chi \int \mathcal{D}A \Phi_f^* \Phi_i e^{-S_{\rm E}},$$

where

$$S_{\rm E} = \int d^4x \mathcal{L}_{\rm SQED}, \tag{11}$$

and a gauge-invariant two-body state $\Phi$ is defined by

$$\Phi(x, \bar{x}) = \chi^*(x) U(x, \bar{x}) \chi(\bar{x}). \tag{12}$$

The gauge link $U(x, y)$ which insures gauge invariance of bilinear product of fields is defined by

$$U(x, y) \equiv \exp\left[ -ie \int_x^y dz A(z) \right]. \tag{13}$$

One can easily see that, under a local gauge transformation

$$\chi(x) \to e^{ie\Lambda(x)} \chi(x), \tag{14}$$
$$A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x),$$

$\Phi_i(x, \bar{x})$ remains gauge invariant:

$$\Phi(x, \bar{x}) \tag{15}$$
$$\to \exp\Bigg[ \underbrace{-ie\Lambda(x) + ie\Lambda(\bar{x}) - ie \int_x^{\bar{x}} dz_\mu \partial_\mu \Lambda}_{0} \Bigg] \chi^*(x)$$
$$\times U(x, \bar{x}) \chi(\bar{x}) = \Phi(x, \bar{x}).$$

Performing path integrals over $\chi$ and $\chi^*$ fields in Eq. (10), one finds

$$G(y, \bar{y}|x, \bar{x}) = N \int \mathcal{D}A (\det S) U(x, \bar{x}) \tag{16}$$
$$\times U^*(y, \bar{y}) S(x, y) S(\bar{x}, \bar{y}) e^{-S[A]},$$

where the interacting one-body propagator $S(x, y)$ is defined by

$$S(x, y) \equiv \left\langle y \left| \frac{1}{m^2 + H(\hat{z}, \hat{p})} \right| x \right\rangle \tag{17}$$

with

$$H(\hat{z}, \hat{p}) \equiv (\hat{p} + ieA(\hat{z}))^2. \tag{18}$$

The Green's function (16) includes contributions coming from all possible interactions. The determinant in Eq. (16) accounts for all matter ($\chi\bar{\chi}$) loops. Setting this determinant equal to unity ($\det S \to 1$, referred to as the quenched approximation) eliminates all contributions from these loops (illustrated in Fig. 1) and greatly simplifies the calculation.

Analytical calculation of the path integral over the gauge field $A$ in Eq. (16) seems difficult due to the nontrivial $A$ dependence in $S(x, y)$. In more complicated theories, such as QCD, integration of the gauge field, as far as we know, cannot be done analytically. Therefore, in QCD, the only option is to do the gauge-field path integral by using a brute force method. Here, one usually introduces a discrete space–time lattice and integrates over the values of the field components at each lattice site. However, for the simple scalar QED interaction under consideration, it is in fact possible to go further and eliminate the path integral over the field $A$. In order to be able to carry out the remaining path integral over the exchange field $A$, it is desirable to represent the interacting propagator in the form of an exponential. This can be achieved by using a Feynman representation for the interacting propagator. The first step involves the exponentiation of the denominator in Eq. (17):

$$S(x, y) = \int_0^\infty ds e^{-sm^2} \langle y | \exp[-sH] | x \rangle. \tag{19}$$

This expression is similar to a quantum-mechanical propagator with $s = it$ and a Hamiltonian $H$ which is a covariant function of 4-vector momenta and coordinates. It is known how to represent a quantum-mechanical propagator as a path integral. The representation is in terms of the Lagrangian, and a covariant Lagrangian can easily be obtained from the Hamiltonian (18)

$$H(\hat{z}, \hat{p}) = (\hat{p} + ieA(\hat{z}))^2 \Longrightarrow L(z, \dot{z}) \tag{20}$$
$$= \frac{\dot{z}^2}{4} - ie\dot{z} \cdot A(z).$$

Using this Lagrangian, the path-integral representation for the interacting propagator becomes

$$S(x, y) = \int_0^\infty ds \int (\mathcal{D}z)_{xy} \tag{21}$$
$$\times \exp\left[ -sm^2 - \frac{1}{4} \int_0^s d\tau_s \dot{z}^2(\tau_s) \right.$$
$$\left. - ie \int_0^s d\tau_s \dot{z} A(z(\tau_s)) \right],$$

where $z_i(\tau_s)$ is a particle trajectory which is a parametric function of the parameter $\tau_s$, with $s \geq \tau_s \geq 0$ and endpoints $z_i(0) = x_i$, $z_i(s) = y_i$, and $i = 1$ to 4. This representation allows one to perform the remaining path integral over the exchange field $A$. The final result for the two-body propagator involves a quantum-mechanical path integral that sums up contributions coming from all possible trajectories of the two charged particles:

$$G = -\int_0^\infty ds \int_0^\infty d\bar{s} \int (\mathcal{D}z)_{xy} \tag{22}$$
$$\times \int (\mathcal{D}\bar{z})_{\bar{x}\bar{y}} e^{-K[z,s]-K[\bar{z},\bar{s}]} \langle W(C)\rangle,$$

where the parameter $\tau_s$ is rescaled, so that $\tau_s = s\tau$; the kinetic term $K$ is defined by

$$K[z,s] = m^2 s + \frac{1}{4s}\int_0^1 d\tau \dot{z}^2(\tau); \tag{23}$$

and the Wilson loop average $\langle W(C)\rangle$ is given by

$$\langle W(C)\rangle \equiv \int \mathcal{D}A \exp\left[ -ie \oint_C dz A(z) \right. \tag{24}$$
$$\left. - \frac{1}{2}\int d^4 z A(z)(\mu^2 - \partial^2)A(z)\right],$$

where the contour of integration $C$ (shown in Fig. 2) follows a clockwise trajectory $x \to y \to \bar{y} \to \bar{x} \to x$ as parameters $\tau$ and $\bar{\tau}$ are varied from 0 to 1. The $A$ integration in the definition of the Wilson loop average is of standard Gaussian form and can be easily performed to obtain

$$\langle W(C)\rangle \tag{25}$$
$$= \exp\left[ -\frac{e^2}{2}\int_C dz_\mu \int_C d\bar{z}_\nu \Delta_{\mu\nu}(z - \bar{z}, \mu)\right],$$

$$\Delta_{\mu\nu}(x,\mu) = g_{\mu\nu}\int \frac{d^4 p}{(2\pi)^4}\frac{e^{ipx}}{p^2 + \mu^2}. \tag{26}$$

When it is necessary to regulate the ultraviolet singularities in (26), a double Pauli–Villars subtraction will be used, so that (26) will be replaced by

$$\Delta_{\mu\nu}(x,\mu) \tag{27}$$
$$= g_{\mu\nu}\int \frac{d^4 p}{(2\pi)^4}\frac{e^{ipx}(\Lambda_1^2 - \mu^2)(\Lambda_2^2 - \mu^2)}{(p^2 + \mu^2)(p^2 + \Lambda_1^2)(p^2 + \Lambda_2^2)}.$$

Through the results given in Eqs. (22), (25) and either (26) or (27), the path integration expression involving fields has been transformed into a path integral representation involving trajectories of particles.

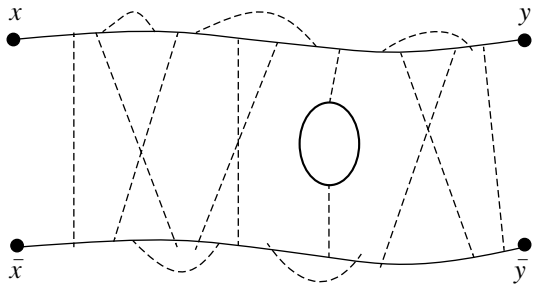


**Fig. 1.** The dashed curves represent exchanges of the gauge field with mass $\mu$ and the solid curves the propagation of the matter fields with mass $m$. Note the matter loop in one of the middle exchanges. All loops of this kind are neglected in the quenched approximation (when $\det S \to 1$).

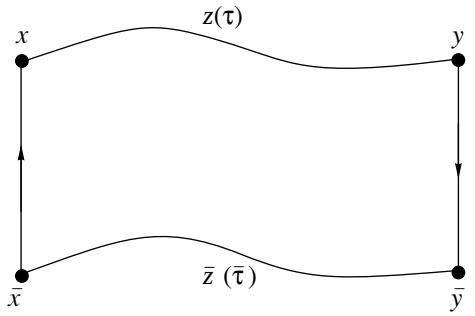


**Fig. 2.** The contour $C$, known as a Wilson loop, that arises in Eq. (24).

Equation (22) has a very nice physical interpretation. The term $\Delta_{\mu\nu}(z_a - z_b, \mu)$ describes the propagation of gauge field interactions between any two points on the particle trajectories, and the appearance of these interaction terms in the exponent means that the interactions are summed to all orders with arbitrary ordering of the points on the trajectories. Self-interactions come from terms with the two points $z_a$ and $z_b$ on the same trajectory, generalized ladder exchanges arise if the two points are on different trajectories, and vertex corrections arise from a combination of the two. Because the particles forming the two-body bound state carry opposite charges, it follows that the self-energy and exchange contributions have different signs.

The bound-state spectrum can be determined from the spectral decomposition of the two-body Green's function

$$G(T) = \sum_{n=0}^{\infty} c_n e^{-m_n T}, \tag{28}$$

where $T$ is defined as the average time between the

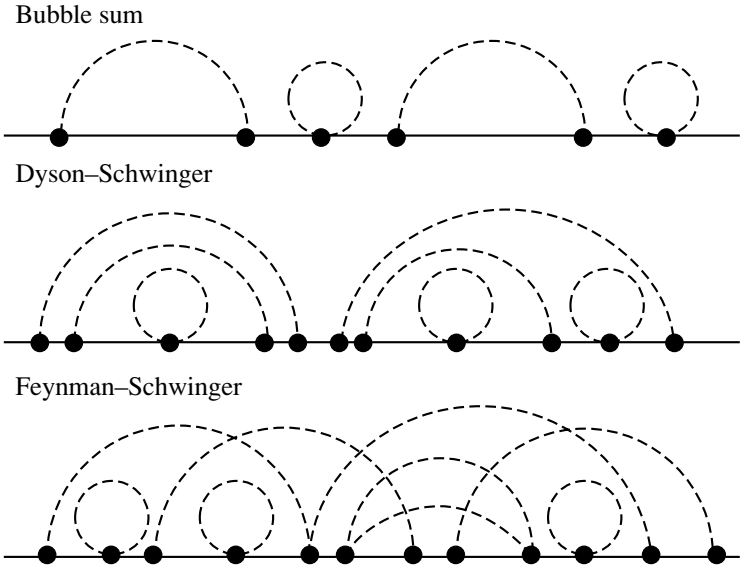


**Fig. 3.** Various interactions included in each approach are shown.

initial and final states,

$$T \equiv \frac{1}{2}(y_4 + \bar{y}_4 - x_4 - \bar{x}_4). \tag{29}$$

In the limit of large $T$, the ground-state mass is given by

$$m_0 = -\lim_{T\to\infty} \frac{d}{dT}\ln[G(T)] = -\frac{\int \mathcal{D}Z S'[Z] e^{-S[Z]}}{\int \mathcal{D}Z e^{-S[Z]}}. \tag{30}$$

## 3. SCALAR QUANTUM ELECTRODYNAMICS

In this section, we take a closer look at one-body mass pole calculations for the case of SQED. Two popular methods frequently used to find the dressed mass of a particle are to do a simple bubble summation or to solve the one-body Dyson–Schwinger equation in the rainbow approximation. It is interesting to compare results given by the bubble summation and the Dyson–Schwinger with the exact FSR result. Below, we first give a quick overview of how dressed masses can be obtained in bubble summation and the Dyson–Schwinger equation approaches. For technical simplicity, the one-body discussion will be limited to $0+1$ dimension.

The simple bubble summation involves a summation of all bubble diagrams to all orders. The dressed propagator is given by

$$\Delta_d(p) = \frac{1}{p^2 + m^2 + \Sigma(p)}. \tag{31}$$

The dressed mass $M$ is determined from the self-energy using

$$M = \sqrt{m^2 + \Sigma(iM)}. \tag{32}$$

The self-energy for the simple bubble sum (in $0+1$ dimension) is given by

$$\Sigma(p) = -e^2 \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{1}{(k^2+\mu^2)} \times \left\{ \frac{(2p-k)^2}{[(p-k)^2+m^2]} - 1 \right\}. \tag{33}$$

The self-energy integral in this case is trivial and can be performed analytically, and the dressed mass is determined from Eq. (32).

The rainbow Dyson–Schwinger equation sums more diagrams than the simple bubble summation (Fig. 3). The self-energy of the rainbow Dyson–Schwinger equation involves a momentum-dependent mass:

$$\Sigma(p) = -e^2 \int_{-\infty}^{\infty} \frac{dk}{2\pi} \frac{1}{(k^2+\mu^2)} \times \left\{ \frac{(2p-k)^2}{[(p-k)^2 + \underbrace{m^2 + \Sigma(k)}]} - 1 \right\}. \tag{34}$$

In this case, the self-energy is nontrivial and it must be determined by a numerical solution of Eq. (34). The dressed mass is determined by the logarithmic derivative of the dressed propagator in coordinate space

$$M = -\lim_{T\to\infty} \frac{d}{dT}\log[\Delta_d(t)]. \tag{35}$$

The type of diagrams summed by each method is shown in Fig. 3. Note that the matter loops do not

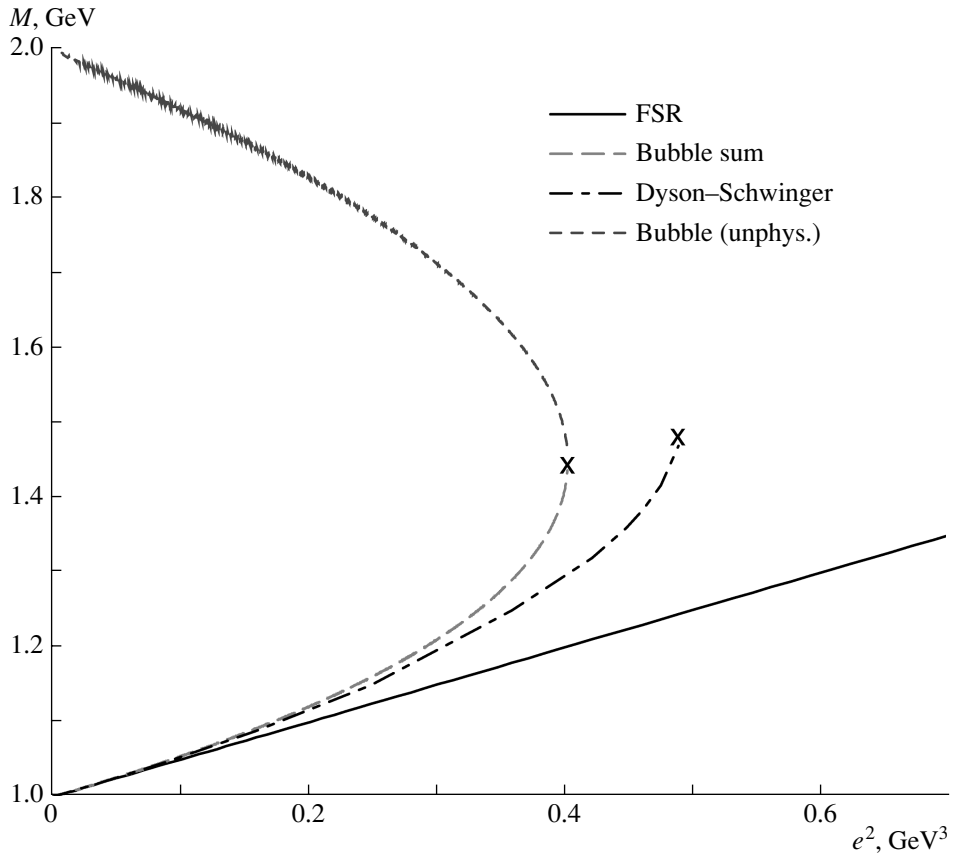


**Fig. 4.** The function $M(e^2)$, calculated in $0+1$ dimension with values of $m = \mu = 1$ GeV, using the FSR approach, the Dyson–Schwinger equation in the rainbow approximation, and the bubble summation. While the exact result is always real, the rainbow Dyson–Schwinger and the bubble summation results become complex beyond a critical coupling.

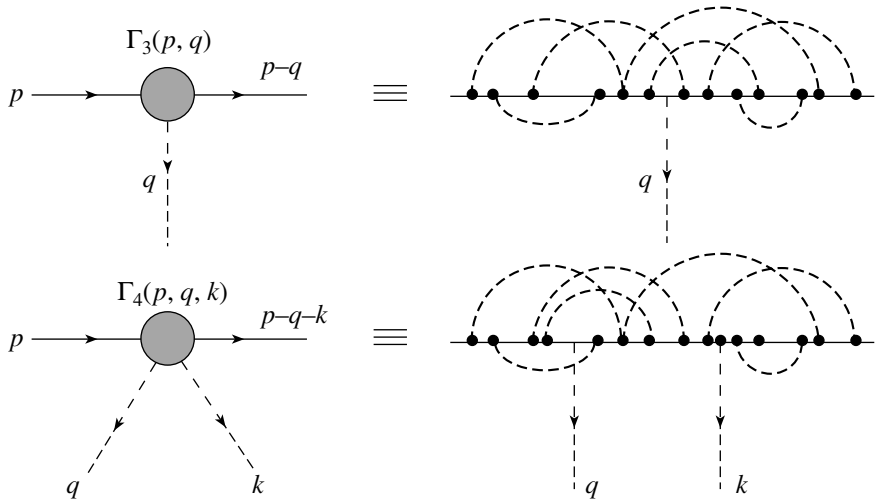


**Fig. 5.** Because of the interactions, the one-particle irreducible vertex functions $\Gamma_n$ $(n = 3, 4, \dots)$ depend on the external momenta.

give any contribution, as explained earlier. Results obtained by these three methods are shown in Fig. 4. It is interesting to note that the simple bubble summation and the rainbow Dyson–Schwinger results display similar behavior. While the exact result provided by the FSR linearly increases for all coupling strengths, both the simple bubble summation and the rainbow Dyson–Schwinger results come to a critical

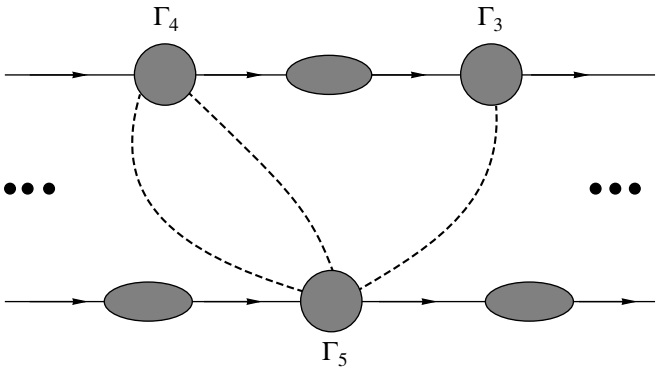


**Fig. 6.** Exact computation of the two-body bound-state propagator requires the summation of all particle self-energies, vertex corrections, and ladder and crossed ladder exchanges.

point beyond which solutions for the dressed masses become complex. This example very clearly shows that conclusions about the mass poles of propagators based on approximate methods such as the rainbow Dyson–Schwinger equation can be misleading.

In general, a consistent treatment of any nonperturbative calculation must involve summation of all possible vertex corrections. Vertex corrections are those irreducible diagrams that surround an interaction vertex. The elementary vertex is the three-point vertex, $\Gamma_3$, but the particle interactions will lead to the appearance of $n$th-order irreducible vertices, $\Gamma_n$, as illustrated in Fig. 5. The propagation of a bound state therefore involves a summation of all diagrams with the inclusion of higher order vertices (Fig. 6). A rigorous determination of all of these vertices is not feasible. In the literature on bound states, $\Gamma_{n>3}$ interaction vertices are usually completely ignored. The three-point vertex $\Gamma_3$ can be approximately calculated in the ladder approximation [34]. However, a rigorous determination of the exact form of the three-point vertex is not possible, for this requires the knowledge of even higher order vertices.

In order to be able to make a connection between the exact theory and predictions based on approximate bound-state equations, it is essential that the role of interaction vertices be understood. The FSR is a useful technique for this purpose.

We now present the interesting outcome that the full bound-state result dictated by a Lagrangian can be obtained by summing only generalized ladder diagrams ("generalized" ladders include both crossed ladders and, in theories with an elementary four-point interaction, both overlapping and nonoverlapping "triangle" and "bubble" diagrams).

We adopt the following procedure for determining the contribution of vertex corrections in $3+1$ dimensions. We start with an initial bare mass $m$ and calculate the full two-body bound-state result with the inclusion of all interactions: generalized ladders, self-energies, and vertex corrections. Let us denote the result for the exact two-body bound-state mass by $M_2^{\rm tot}(e^2, m)$, since it will be a function of the coupling strength $e$ and the bare input mass $m$, and the superscript "tot" implies that all interactions are summed. Next, we calculate the dressed one-body mass $M_1(e^2, m)$. Then, using the dressed mass value $M_1(e^2, m)$, we calculate the bound-state mass $M_2^{\rm exch}(e^2, M_1)$ by summing only the generalized exchange-interaction contributions. In this last calculation, we sum only exchange interactions (generalized ladders), but the self-energy is approximately taken into account since we use the (constant) dressed one-body mass as input. However, the vertex corrections and wave-function renormalization are completely left out since we use the original vertex provided by the Lagrangian. In order to compare the full result, where all interactions have been summed, with the result obtained by two dressed particles interacting only by generalized ladder exchanges, we plot the bound-state masses obtained by these methods. Numerical results are presented in Fig. 7. This result is qualitatively similar to that obtained analytically for SQED in $0+1$ dimension [33].

The numerical results presented here yield the following prescription for bound-state calculations: In order to get the full result for bound states, it is a good approximation to first solve for dressed one-body masses exactly (summing all generalized rainbow diagrams) and then use these dressed masses and the bare interaction vertex provided by the Lagrangian to calculate the bound-state mass by summing only generalized ladder interactions (leaving out vertex corrections). In terms of Feynman graphs, this prescription can be expressed as in Fig. 8.

The significance of the results presented above rests in the fact that the problem of calculating exact results for bound-state masses in SQED has been reduced to that of calculating only generalized ladders. Summation of generalized ladders can be addressed within the context of bound-state equations [22, 23, 35]. In this paper, we have shown the connection between the full prediction of a Lagrangian and the summation of generalized ladder diagrams. Our results are rigorous for SQED, but are only suggestive for more general theories with spin or internal symmetries. Since we have neglected charged particle loops (our results are in the quenched approximation) and the current is conserved in SQED, it is perhaps not surprising that the bare coupling is not renormalized, but the fact that the momentum dependence of the dressed mass and vertex corrections seem to

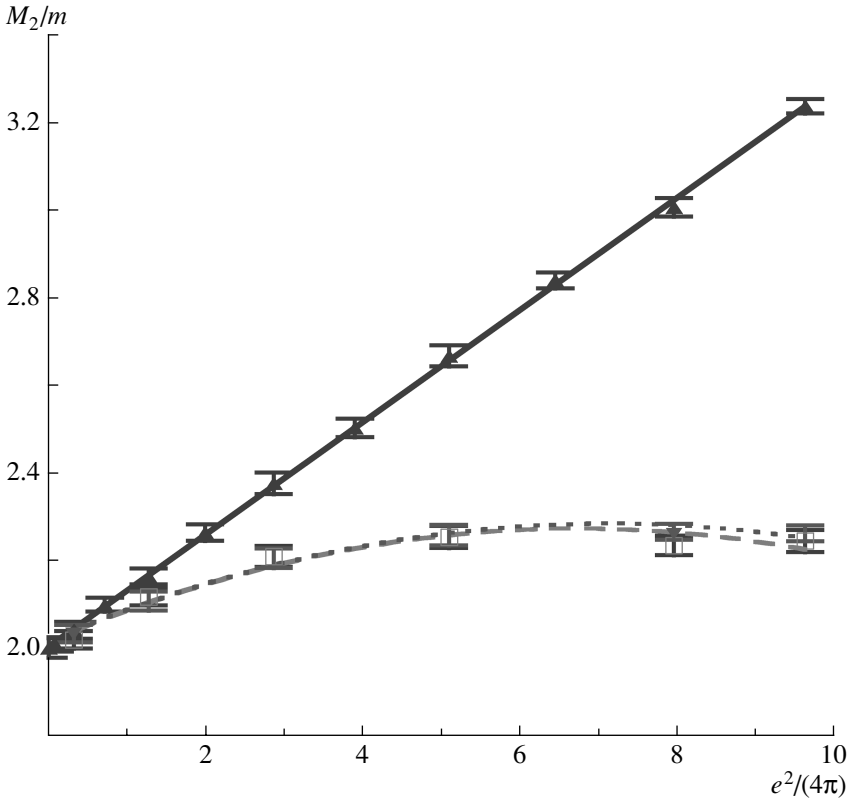


**Fig. 7.** Two-body bound-state mass for SQED in $3+1$ dimensions. Closed triangles are $2M_1$, open squares are $M_2^{\rm exch}$, and closed inverted triangles are $M_2^{\rm tot}$. Here, $\mu/m = 0.15$, $\Lambda_1/\mu = 3$, and $\Lambda_2/\mu = 5$. The smooth curves are fits to the "data." Note that $M_2^{\rm exch} = M_2^{\rm tot}$ to within errors.

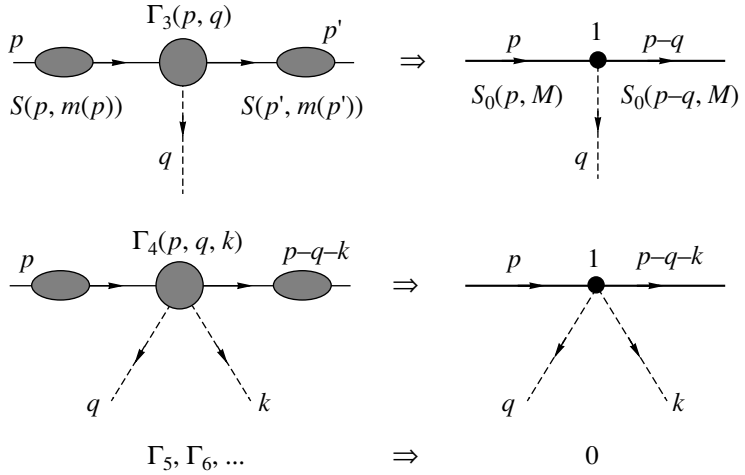


**Fig. 8.** The correct two-body result can be obtained by simply using a dressed constituent mass and a bare vertex and ignoring the contributions of higher order vertices.

cancel is unexpected. If we were to unquench our calculation or to use a theory without a conserved current, it is reasonable to expect that both the bare interaction and the mass would be renormalized.

Finally, we call attention to a remarkable cancellation that occurs in the one-body calculations. The exact self-energy shown in Fig. 7 (and also in Fig. 4 for different parameters) is nearly linear in $e^2$ [29]. This remarkable fact implies that the exact self-energy is well approximated by the lowest order result from

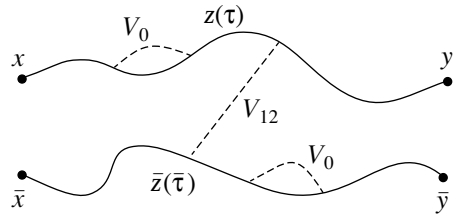


**Fig. 9.** Sample trajectories with self-energy and exchange interactions.

perturbation theory. It is instructive to see how this comes about. If we expand the self-energy to fourth order, expanding each term about the bare mass $m$, we have

$$S_d^{-1}(p^2) = m^2 - p^2 + \Sigma(p^2) \tag{36}$$
$$= m^2 - p^2 + \Sigma_2 + (p^2 - m^2)\Sigma_2' + \Sigma_4,$$

where $\Sigma_\ell = \Sigma_\ell(m^2)$ is the contribution of order $e^\ell$ evaluated at $p^2 = m^2$, $\Sigma' = d\Sigma(p^2)/dp^2$ evaluated at $p^2 = m^2$, and the formula is valid for $p^2 - m^2 \simeq e^2$. Expanding the dressed mass in a power series in $e^2$

$$M_1^2 = m^2 + m_2^2 + m_4^2 + \ldots, \tag{37}$$

where $m_\ell^2$ is the contribution of order $e^\ell$, and substituting it into Eq. (36) gives

$$M_1^2 = m^2 + \Sigma_2 + \Sigma_2'\Sigma_2 + \Sigma_4 + \ldots. \tag{38}$$

The mass is then

$$M_1 = m + \frac{\Sigma_2}{2m} + \frac{4m^2\left[\Sigma_2'\Sigma_2 + \Sigma_4\right] - \Sigma_2^2}{8m^3} + \ldots. \tag{39}$$

The linearity of the exact result implies that the fourth-order term in Eq. (39) must be zero (or very small), and this can be easily confirmed by direct calculations!

The cancellation of the fourth-order mass correction (and all higher orders) is reminiscent of the cancellations between generalized ladders, which explains why quasipotential equations are more effective than the ladder Bethe–Salpeter equation in explaining two-body binding energies. It shows that a simple evaluation of the second-order self-energy at the bare-mass point is more accurate than solution of the Dyson–Schwinger equation in the rainbow approximation.

The general lesson seems to be that attempts to sum a small subclass of diagrams exactly is often less accurate than the approximate summation of a larger class of diagrams.

In the next section, we consider the application of the FSR approach to scalar $\chi^2\phi$ interaction.

## 4. SCALAR $\chi^2\phi$ INTERACTION WITH THE FSR APPROACH

We consider the theory of charged scalar particles $\chi$ of mass $m$ interacting through the exchange of a neutral scalar particle $\phi$ of mass $\mu$. The Euclidean Lagrangian for this theory is given by

$$\mathcal{L} = \chi^*\left[m^2 - \partial^2 + g\phi\right]\chi + \frac{1}{2}\phi(\mu^2 - \partial^2)\phi. \tag{40}$$

The two-body propagator for the transition from the initial state $\Phi_i = \chi^*(x)\chi(\bar{x})$ to final state $\Phi_f = \chi^*(y)\chi(\bar{y})$ is given by

$$G(y,\bar{y}|x,\bar{x}) \tag{41}$$
$$= N\int \mathcal{D}\chi^* \int \mathcal{D}\chi \int \mathcal{D}\phi \Phi_f^* \Phi_i \exp\left[-\int d^4x \mathcal{L}\right].$$

After the usual integration of matter fields is done, the Green's function reduces to

$$G(y,\bar{y}|x,\bar{x}) = N\int \mathcal{D}\phi(\det S) S(x,y) S(\bar{x},\bar{y}) e^{-L_0[\phi]}, \tag{42}$$

with the free Lagrangian $L_0$ and the interacting propagator $S(x,y)$ defined by

$$L_0[\phi] \equiv \frac{1}{2}\int d^4z \phi(z)(\mu^2 - \partial_z^2)\phi(z), \tag{43}$$
$$S(x,y) \equiv \left\langle y \left| \frac{1}{m^2 + H(\hat{z},\hat{p})} \right| x \right\rangle,$$

with the Hamiltonian

$$H(\hat{z},\hat{p}) \equiv \hat{p}^2 - g\phi(\hat{z}). \tag{44}$$

As in the case of scalar QED, we employ the quenched approximation: $\det S \to 1$.

We exponentiate the denominator by introducing an $s$ integration along the imaginary axis with an $\epsilon$ prescription

$$S(x,y) = \int_0^{i\infty} ds\, e^{-s(m^2 + i\epsilon)} \langle y|\exp[-sH]|x\rangle. \tag{45}$$

This representation should be compared with the representation used earlier in SQED Eq. (19). Here, the integration is done along the imaginary axis because $H$ is not positive definite.

Again, a quantum-mechanical path-integral representation can be constructed by recognizing that Lagrangian corresponding to the $H$ of Eq. (44) is given by

$$L(z,\dot{z}) = \frac{\dot{z}^2}{4} + g\phi(z). \tag{46}$$

The path-integral representation for the interacting propagator is therefore

$$S(x,y) = -i\int_0^\infty ds \int \mathcal{D}z \exp\left[is(m^2+i\epsilon)\right. \quad (47)$$
$$\left. -\frac{i}{4}\int_0^s d\tau \dot{z}^2(\tau) + ig\int_0^s d\tau\phi(z(\tau))\right].$$

This representation allows the elimination of the integral over the exchange field $\phi$. The two-body propagator reduces to

$$G = -\int_0^\infty ds \int_0^\infty d\bar{s} \int (\mathcal{D}z)_{xy} \quad (48)$$
$$\times \int (\mathcal{D}\bar{z})_{\bar{x}\bar{y}} e^{iK[z,s]+iK[\bar{z},\bar{s}]} I_\phi,$$

where mass and kinetic term is given by

$$K[z,s] = (m^2+i\epsilon)s - \frac{1}{4s}\int_0^1 d\tau \dot{z}^2(\tau). \quad (49)$$

The field integration $I_\phi$ is a standard Gaussian integration:

$$I_\phi \equiv \int \mathcal{D}\phi \exp\left[+ig\left(\int_0^s d\tau\phi(z(\tau)) \right.\right. \quad (50)$$
$$\left.\left. + \int_0^{\bar{s}} d\bar{\tau}\phi(\bar{z}(\bar{\tau}))\right) - L_0[\phi]\right]$$
$$\equiv \exp(-V_0[z,s] - 2V_{12}[z,\bar{z},s,\bar{s}] - V_0[\bar{z},\bar{s}]),$$

where $V_0$ and $V_{12}$ (self- and exchange-energy contributions in Fig. 9) are defined by

$$V_0[z,s] \quad (51)$$
$$= \frac{g^2}{2}s^2\int_0^1 d\tau \int_0^1 d\tau' \Delta(z(\tau)-z(\tau'),\mu) \equiv s^2 v[z],$$

$$V_{12}[z,\bar{z},s,\bar{s}] \quad (52)$$
$$= \frac{g^2}{2}s\bar{s}\int_0^1 d\tau \int_0^1 d\bar{\tau}\Delta(z(\tau)-\bar{z}(\bar{\tau}),\mu).$$

It should be noted that the interaction terms explicitly depend on the $s$ variable, which was not the case for SQED. The interaction kernel $\Delta$ is given by

$$\Delta(x,\mu) = \int \frac{d^4p}{(2\pi)^4}\frac{e^{ip\cdot x}}{p^2+\mu^2} \quad (53)$$
$$= \frac{\mu}{4\pi^2|x|}K_1(\mu|x|).$$

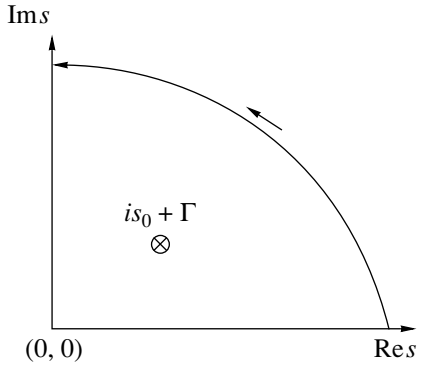


**Fig. 10.** Wick rotation in the $s$ integration.

In order to be able to compute the path integral over trajectories, a discretization of the path integral is needed:

$$(\mathcal{D}z)_{xy} \to (N/4\pi s)^{2N}\prod_{i=1}^{N-1}\int d^4z_i. \quad (54)$$

The $s$ dependence is crucial for correct normalization. After discretization, the one-body propagator takes the following form:

$$G = i\left(\frac{N}{4\pi}\right)^{2N}\int\prod_{i=1}^{N-1}dz_i\int_0^\infty\frac{ds}{s^{2N}} \quad (55)$$
$$\times\exp\left[im^2s - i\frac{k^2}{4s} - s^2v[z]\right],$$

where $v[z]$ was defined in Eq. (51). This is an oscillatory and regular integral and it is not convenient for Monte Carlo integration. The origin of the oscillation is the fact that the $s$ integral was defined along the imaginary axis,

$$\text{Rep. 1: } S(x,y) \quad (56)$$
$$= \left\langle y\left|\int_0^{-i\infty} ds\exp[-s(m^2-\partial^2+g\phi+i\epsilon)]\right|x\right\rangle.$$

In earlier works [4, 27], a nonoscillatory FSR was used,

$$\text{Rep. 2: } S(x,y) \quad (57)$$
$$= \left\langle y\left|\int_0^{\infty} ds\exp[-s(m^2-\partial^2+g\phi)]\right|x\right\rangle.$$

Rep. 2 leads to a nonoscillatory and divergent result

$$G \propto \int_0^\infty \frac{ds}{s^{2N}}\exp\left[-m^2s - \frac{k^2}{4s} + s^2v[z]\right], \quad (58)$$

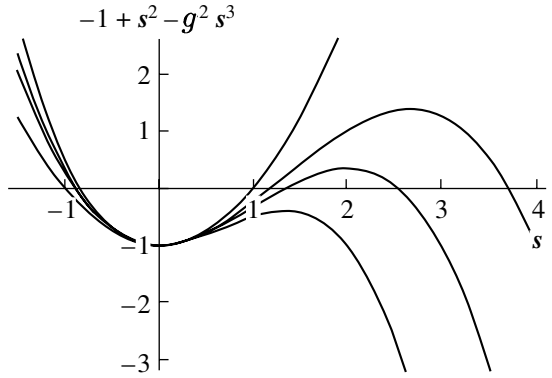


**Fig. 11.** As the effective coupling strength $\boldsymbol{g}^2$ is increased, the stationary point disappears.

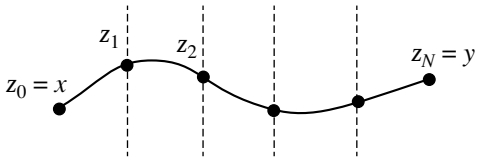


**Fig. 12.** Number of steps a particle takes between initial and final coordinates is discretized. The space–time is *continuous* and there are *no space–time boundaries*.

and the large-$s$ divergence was regulated by a cutoff $\Lambda$. This is not a satisfactory prescription since it relies on an arbitrary cutoff. Later, it was shown [28, 30] that the correct procedure is to start with Rep. 1 and make a Wick rotation such that the final result is nonoscillatory and regular. The implementation of Wick rotation, however, is nontrivial. Consider the $s$-dependent part of the integral for the one-body propagator

$$G \propto \int_0^\infty \frac{ds}{s^{2N}} \exp\left[im^2 s - i\frac{k^2}{4s} - s^2 v[z]\right]. \qquad (59)$$

It is clear that a replacement of $s \to is$ leads to a divergent result. The problem with the Wick rotation (Fig. 10) comes from the fact that the $s$ integral is infinite both along the imaginary axis and along the contour at infinity. These two infinities cancel to yield a finite integral along the real axis. As $g \to 0$, the dominant contribution to the $s$ integral in Eq. (59) comes from the stationary point

$$s = is_0 \simeq i\frac{k}{2m}. \qquad (60)$$

Therefore, one might suppress the integrand away from the stationary point by introducing a damping factor $R$:

$$R(s, s_0) \equiv 1 - (s - is_0)^2/\Gamma^2. \qquad (61)$$

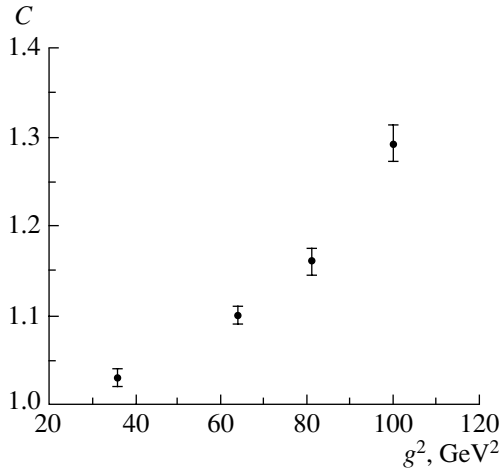


**Fig. 13.** The dependence of the peak of the $s$ distribution on the coupling strength for the two-body bound state is shown. The peak location is given by $s_0 = CT/2m$. There is no real solution for $C$ beyond $g^2 = 100$ GeV$^2$.

With this factor, the integral in Eq. (59) is modified:

$$G \propto \int_0^\infty \frac{ds}{s^{2N}} \exp\left[im^2 s - i\frac{k^2}{4s} - \frac{s^2 v[z]}{R^2(s, s_0)}\right]. \qquad (62)$$

This modification allows us to make a Wick rotation since the contribution of the contour at infinity now vanishes. However, this procedure relies on the fact that there exists a stationary point. It can be seen from the original expression Eq. (59) that this is not always true. According to the original integral, the stationary point is given by the following equation:

$$im^2 + i\frac{k^2}{4s^2} - 2sv[z] = 0. \qquad (63)$$

Introducing $s = i\boldsymbol{s}k/(2m)$ and $\boldsymbol{g}^2 = k\langle v[z]\rangle/m^3$, this equation becomes

$$-1 + \boldsymbol{s}^2 - \boldsymbol{g}^2\boldsymbol{s}^3 = 0. \qquad (64)$$

The stationary point is determined by the first intersection of a cubic plot with the positive $s$ axis as shown in Fig. 11. The plot in Fig. 11 shows that, as the effective coupling strength $\boldsymbol{g}^2$ is increased, the curve no longer crosses the positive $\boldsymbol{s}$ axis. Therefore, beyond a critical coupling strength, the stationary point vanishes and mass results should be unstable. Limiting discussion to cases where the original expression Eq. (59) has a critical point, we now turn to perform a Wick rotation on the modified expression Eq. (62). The Wick rotation in Eq. (62) amounts to a simple replacement $s \to is$, and a regular, nonoscillatory integral is found:

$$G \propto \int_0^\infty \frac{ds}{s^{2N}} \exp\left[-m^2 s - \frac{k^2}{4s} + \frac{s^2 v[z]}{R^2(is, s_0)}\right]. \qquad (65)$$

At first glance, it seems that the new integral always has a stationary point determined by the following equation:

$$-m^2 + \frac{k^2}{4s^2} + \frac{2sv[z]}{R^2(is,s_0)} - s^2 v[z]\frac{(R^2(is,s_0))'}{R^2(is,s_0)} = 0. \quad (66)$$

The key point to remember is that the stationary point we find after the Wick rotation should be the same stationary point we had before the Wick rotation. This is required to make sure that the physics remains the same after the Wick rotation. Self-consistency therefore requires that the stationary point after the Wick rotation be at $s = is_0$. In that case, $R(is_0, s_0) = 1$ and $(R^2(is_0, s_0))' = 0$ and Eq. (66) determining the critical point reduces to the earlier original Eq. (63).

The regularization of the ultraviolet singularities is done using Pauli–Villars regularization, which is particularly convenient for numerical integration since it only involves a change in the interaction kernel

$$\Delta(x,\mu) \longrightarrow \Delta(x,\mu) - \Delta(x,\alpha\mu). \quad (67)$$

Calculations of the $\chi^2\phi$ interaction in $3+1$ dimension require numerical Monte Carlo integration. The first step is to represent the particle trajectories by a discrete number of $N+1$ points with boundary conditions given by

$$z_0 = x = (x_1, x_2, x_3, 0), \quad (68)$$
$$z_N = y = (y_1, y_2, y_3, T).$$

The discretization employed in the FSR is for the number of time steps a particle takes in going from the initial time to the final time along a trajectory in a four-dimension coordinate space, as illustrated in Fig. 12. This is very different from the discretization employed in lattice gauge theory. Contrary to lattice gauge theory, in the FSR approach, space–time is continuous and rotational symmetry is respected. An additional important benefit is the lack of space–time lattice boundaries, which allows calculations of arbitrarily large systems using the FSR approach. This feature provides an opportunity for doing complex applications such as calculation of the form factors of large systems.

In doing Monte Carlo sampling, we sample trajectories (lines) rather than gauge-field configurations (in a volume). This leads to a significant reduction in the numerical cost. The ground-state mass of the Green's function is obtained using

$$m_0 = \frac{\int \mathcal{D}Z S'[Z] e^{-S[Z]}}{\int \mathcal{D}Z e^{-S[Z]}}. \quad (69)$$

Sampling of trajectories is done using the standard Metropolis algorithm, which insures that configurations sampled are distributed according to the weight $e^{-S[Z]}$. In sampling trajectories, the final-state (spacial) coordinates of particles can be integrated over, which puts the system at rest and projects out the $S$-wave ground state. As trajectories of particles are sampled, the wave function of the system can be determined simply by storing the final-state configurations of particles in a histogram.

In sampling trajectories, the first step is thermalization. In order to insure that the initial configuration of trajectories has no effect on results, the first 1000 or so updates are not taken into account. Statistical independence of subsequent samplings is measured by the correlation function $X(n)$, defined as

$$X(n) \equiv \frac{\langle m(i)m(i+n)\rangle - \langle m\rangle^2}{\langle m\rangle^2}, \quad (70)$$

where $m(i)$ is the mass measurement at the $i$th update.

In order to insure that the location of the stationary point is self-consistent, as discussed earlier, its location must be determined carefully. The stationary point can be parametrized by $s_0 = CT/(2m)$, where $T/(2m)$ is the location of the stationary point when the coupling strength $g$ goes to zero. As the coupling strength is increased, the stationary point moves to larger values of $s_0$ (recall Fig. 11) and $C$ increases. Eventually, a critical value of the coupling constant is reached beyond which there is no self-consistent stationary point. In order to be able to do Monte Carlo integrations, an initial guess must be made for the location of the stationary point. Self-consistency is realized by insuring that the peak location of the $s$ distribution in the Monte Carlo integration agrees with the initial guess for the stationary point [30]. In Fig. 13, the dependence of the location of the stationary point on the coupling strength for two-body bound states is shown. The figure shows that, beyond the critical point $g^2 \simeq 100$ GeV$^2$, $C$ goes to infinity, implying that there is no stationary point. A similar critical behavior was also observed in [36, 37] within the context of a variational approach. In Fig. 14, FSR two-body bound-state mass results are shown for $m_\chi = 1$ GeV, $\mu_\phi = 0.15$ GeV. These results are all for a Pauli–Villars mass of $3\mu$. Also shown are the predictions of various integral equation calculations.

The FSR calculation sums all ladder and crossed ladder diagrams and excludes the self-energy contributions. According to Fig. 14, all bound-state equations underbind. Among the manifestly covariant equations, the Gross equation (labeled GR in Fig. 14) gives the closest result to the exact calculation obtained by the FSR method. This is due to the fact

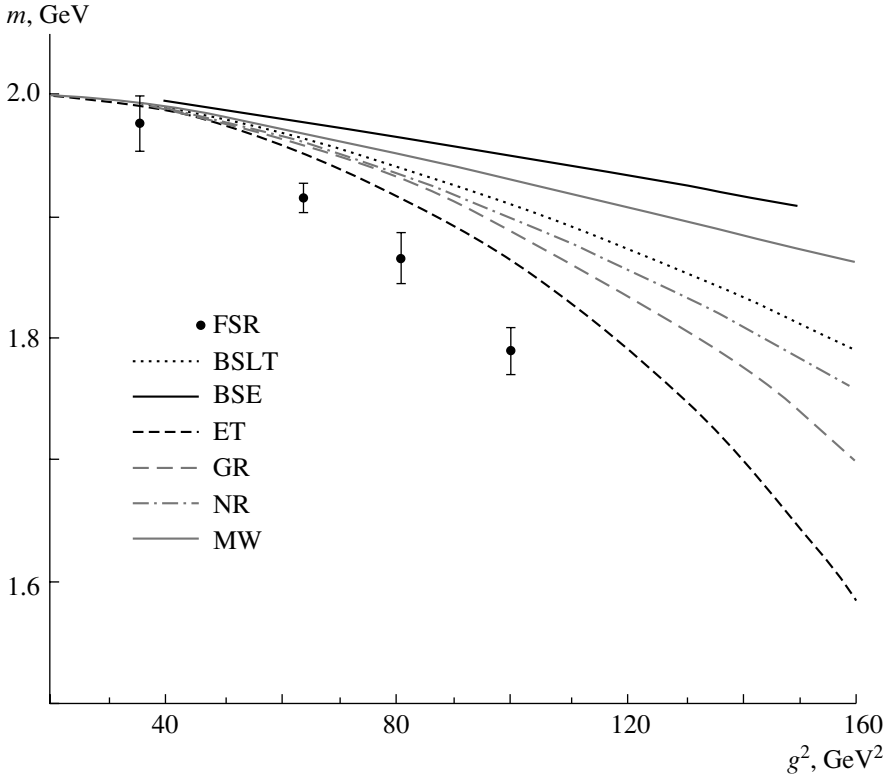


**Fig. 14.** The coupling-constant dependence of the two-body bound-state mass is shown. Beyond the critical coupling strength of $g^2 = 100$ GeV$^2$, the two-body mass becomes unstable. The Bethe–Salpeter equation in the ladder approximation gives the smallest binding. The other models are described in the text.

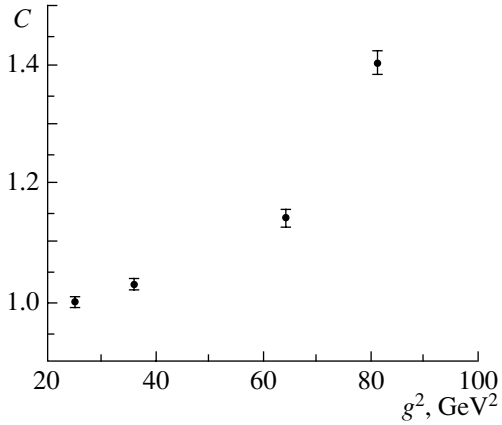


**Fig. 15.** The dependence of the peak of the $s$ distribution on the coupling strength for the three-body bound-state mass is shown. The peak location is given by $s_0 = CT/2m$. There is no real solution for $C$ beyond $g^2 = 81$ GeV$^2$.

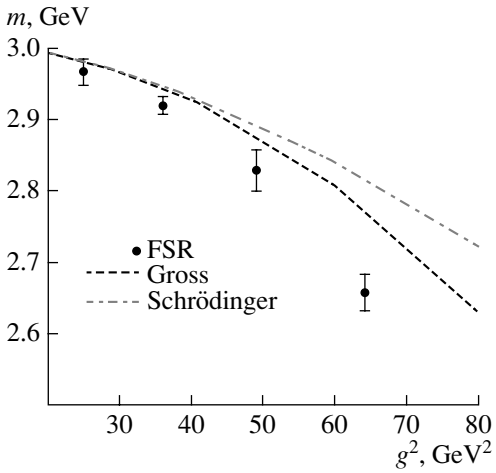


**Fig. 16.** Three-body bound-state results for three equal-mass particles of mass 1 GeV.

that, in the limit of infinitely heavy–light systems, the Gross equation effectively sums all ladder and crossed ladder diagrams. The equal-time equation (labeled ET in Fig. 14) also produces a strong binding, but the inclusion of retardation effects pushes the equal-time results away from the exact results (Mandelzweig–Wallace equation [24], labeled MW in Fig. 14). In particular, the Bethe–Salpeter equation in the ladder approximation (labeled BSE in Fig. 14) gives the lowest binding. Similarly, the Blankenbecler–Sugar–Logunov–Tavkhelidze equation [20, 21] (la-

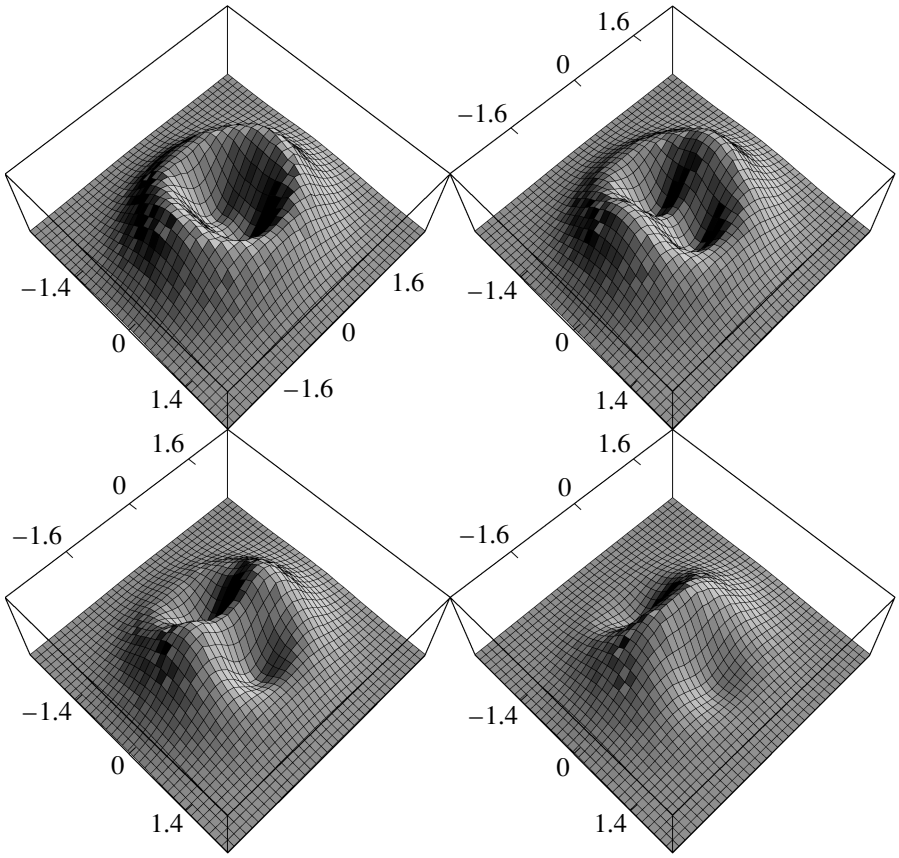


**Fig. 17.** The four panels show the evolution of the probability distribution for the third particle as the distance between the two fixed particles is increased. When the fixed particles are very close to each other, the third particle sees them as a single particle (upper left panel). As the fixed particles are separated, the third particle starts penetrating between them (upper right and lower left panels), and when two fixed particles are far apart (lower right panel), the third particle is most likely to be found between the two fixed particles.

beled BSLT in Fig. 14) gives a very low binding. A comparison of the ladder Bethe–Salpeter, Gross, and the FSR results shows that the exchange of crossed ladder diagrams plays a crucial role.

In Fig. 15, the dependence of the location of the stationary point on the coupling strength for three-body bound states is shown. In Fig. 16, the three-body bound-state results for three equal-mass particles of mass 1 GeV are shown. For the three-body case, the only available results are for the Schrödinger and Gross equations. According to the results presented in Fig. 16, the bound-state equations underbind for the three-body case too. The Gross equation gives the closest result to the exact FSR result.

Determination of the wave function of bound states is done by keeping the final-state configurations of particles in a histogram. For example, for a three-body bound-state system, the probability distribution of the third particle for a given configuration of first and second particles is shown in Fig. 17. In the upper left panel of Fig. 17, the two fixed particles are very close to each other, so that the third particle sees them as a single particle. However, as the fixed particles are separated from each other, the third particle starts having a nonzero probability of being in between the two fixed particles (separation increases as we go from the upper right to lower left panels in Fig. 17). Eventually, when the two fixed particles are far away from each other, the third particle is most likely to be at the origin (the lower right panel in Fig. 17).

Up to this point, the FSR method has been derived and various applications to nonperturbative problems have been presented. In the next section, we discuss the stability of the $\chi^2\phi$ theory.

## 5. STABILITY OF THE SCALAR $\chi^2\phi$ INTERACTION

Scalar field theories with a $\chi^\dagger\chi\phi$ interaction (which we will subsequently denote simply by $\chi^2\phi$) have been used frequently without any sign of instability, despite an argument in 1952 by Dyson [38] suggesting instability and a proof in 1960 by Baym [39]

showing that the theory is unstable. For example, it is easy to show that, for a limited range of coupling values $0 \le g^2 \le g^2_{\rm crit}$, the simple sum of bubble diagrams for the propagation of a single $\chi$ particle leads to a stable ground state, and it is shown in [13] that a similar result also holds for the exact result in the "quenched" approximation. However, if the scalar $\chi^2\phi$ interaction is unstable, then this instability should be observed even when the coupling strength $g$ is vanishingly small $g^2 \to 0^+$, as pointed out recently by Rosenfelder and Schreiber [36] (see also [40]). Both the simple bubble summation and the quenched calculations do not exhibit this behavior. Why do the simple bubble summation and the exact quenched calculations produce stable results for a finite range of coupling values?

A clue to the answer is already provided by the simplest semiclassical estimate of the ground-state energy. In this approximation, the ground-state energy is obtained by minimizing

$$E_0 = m^2\chi^2 + \frac{1}{2}\mu^2\phi^2 - g\phi\chi^2, \qquad (71)$$

where $m$ is the bare mass of the matter particles, and $\mu$ the mass of the "exchanged" quanta, which we will refer to as the mesons. The minimum occurs at

$$E_0 = m^2\chi^2 - g^2\frac{\chi^4}{2\mu^2}. \qquad (72)$$

This is identical to a $\chi^4$ theory with a coupling of the wrong sign, as discussed by Zinn-Justin [41]. The ground state is therefore stable under fluctuations of the $\chi$ field about zero, provided

$$g^2 < g^2_{\rm crit} = \frac{2m^2\mu^2}{\chi^2}. \qquad (73)$$

This simple estimate suggests that the theory is stable over a limited range of couplings if the strength of the $\chi$ field is finite. We now develop this argument more precisely and show under what conditions it holds.

Before presenting new results, we lay the foundation using the variational principle. In the Heisenberg representation, the fields are expanded in terms of creation and annihilation operators that depend on time,

$$\chi(t,\mathbf{r}) = \int d\tilde{k}_m[a(k)e^{-ik\cdot x} + b^\dagger(k)e^{ik\cdot x}], \qquad (74)$$
$$\phi(t,\mathbf{r}) = \int d\tilde{k}_\mu[c(k)e^{-ik\cdot x} + c^\dagger(k)e^{ik\cdot x}],$$

where $x = \{t, \mathbf{r}\}$ and

$$d\tilde{k}_m \equiv \frac{d^3k}{(2\pi)^3\cdot 2E_m(k)} \qquad (75)$$

with $E_m(k) = \sqrt{m^2+k^2}$. The equal-time commutation relations are

$$[a(k), a^\dagger(k')] = (2\pi)^3\cdot 2E_m(k)\delta^3(k-k'). \qquad (76)$$

The Lagrangian for the $\chi^2\phi$ theory is

$$\mathcal{L} = \chi^\dagger[\partial^2 - m^2 + g\phi]\chi + \frac{1}{2}\phi(\partial^2-\mu^2)\phi, \qquad (77)$$

and the Hamiltonian $H$ is a normal ordered product of interacting (or dressed) fields $\phi_d$ and $\chi_d$:

$$H[\phi_d,\chi_d,t] = \int d^3r : \left\{\left(\frac{\partial\chi_d}{\partial t}\right)^2 + (\nabla\chi_d)^2 + m^2\chi_d^2 + \frac{1}{2}\left[\left(\frac{\partial\phi_d}{\partial t}\right)^2 (\vec{\nabla}\phi_d)^2 + \mu^2\phi_d^2\right] - g\chi_d^2\phi_d\right\} : . \qquad (78)$$

This Hamitonian conserves the difference between the number of matter and the number of antimatter particles, which we denote by $n_0$. Eigenstates of the Hamiltonian will therefore be denoted by $|n_0,\lambda\rangle$, where $\lambda$ represents the other quantum numbers that define the state. Hence, allowing for the fact that the eigenvalue may depend on the time,

$$H[\phi_d,\chi_d,t]|n_0,\lambda\rangle = M_{n_0,\lambda}(t)|n_0,\lambda\rangle. \qquad (79)$$

In the absence of an exact solution of (79), we may estimate it from the equation

$$M_{n_0,\lambda}(t) = \langle n_0,\lambda|H[\phi_d,\chi_d,t]|n_0,\lambda\rangle = \langle n_0,\lambda|U^{-1}(t,0)H[\phi,\chi,0]U(t,0)|n_0,\lambda\rangle \equiv \langle n_0,\lambda,t|H[\phi,\chi,0]|n_0,\lambda,t\rangle, \qquad (80)$$

where $U(t,0)$ is the time translation operator which carries the Hamiltonian from time $t=0$ to later time $t$. We have also chosen $t=0$ to be the time at which the interaction is turned on, $\phi_d(t) = U^{-1}(t,0)\times\phi(0)U(t,0)$, and the last step simplifies the discussion by permitting us to work with a Hamiltonian constructed from the free fields $\phi$ and $\chi$. [If the interaction were turned on at some other time $t_0$, we would obtain the same result by absorbing the additional phases $\exp(\pm iEt_0)$ into the creation and annhilation operators.]

At $t=0$, the Hamiltonian in normal order reduces to

$$H[\phi,\chi,0] = \int d\tilde{k}_m E_m(k)\mathcal{N}_0(k,k) + \int d\tilde{p}_\mu E_\mu(p)c^\dagger(p)c(p) - \frac{g}{2}\int\frac{d\tilde{k}_m d\tilde{k'}_m}{\omega(k-k')}\mathcal{N}_1(k,k')[c^\dagger(k'-k) + c(k-k')], \qquad (81)$$

where

$$\mathcal{N}_0(k,k') = \left\{a^\dagger(k)a(k') + b^\dagger(k)b(k')\right\}, \quad (82)$$
$$\mathcal{N}_1(k,k') = \mathcal{N}_0(k,k') + \left\{a^\dagger(k)b^\dagger(-k') + a(-k)b(k')\right\},$$

and $\omega(k) = \sqrt{\mu^2 + \mathbf{k}^2}$. To evaluate the matrix element (80), we express the the eigenstates as a sum of free particle states with $n_0$ matter particles, $n_{\text{pair}}$ pairs of $\chi\bar{\chi}$ particles, and $\ell$ mesons:

$$|n_0,\lambda,t\rangle \equiv |n_0,\alpha(t),\beta(t)\rangle \quad (83)$$
$$= \frac{1}{\gamma(t)}\sum_{n_{\text{pair}}=0}^{\infty}\sum_{\ell=0}^{\infty}\alpha_{n_{\text{pair}}}(t)\beta_\ell(t)|n_0,n_{\text{pair}},\ell\rangle,$$

where $\gamma(t)$ is a normalization constant (defined below), the time dependence of the states is contained in the time dependence of the coefficients $\alpha(t)$ and $\beta(t)$, and

$$|n_0,n_{\text{pair}},\ell\rangle \quad (84)$$
$$\equiv \int \frac{|k_1,\ldots,k_{n_1};q_1,\ldots,q_{n_2};p_1,\ldots,p_\ell\rangle}{\sqrt{(n_0+n_{\text{pair}})!n_{\text{pair}}!\ell!}}.$$

In this equation, $n_1 = n_0 + n_{\text{pair}}$, $n_2 = n_{\text{pair}}$, and

$$\int = \int\prod_{i=1}^{n_1} d\tilde{k}_i f(k_i)\prod_{j=1}^{n_2} d\tilde{q}_j f(q_j)\prod_{l=1}^{\ell} d\tilde{p}_l g(p_l), \quad (85)$$

where $f(k)$ and $g(p)$ are momentum wave functions and the particle masses in $d\tilde{k}$ and $d\tilde{p}$ have been suppressed; their values should be clear from the context. The normalization of the functions $f(p)$ and $g(p)$ is chosen to be

$$\int d\tilde{k} f^2(k) = \int d\tilde{p} g^2(p) \equiv 1, \quad (86)$$

which leads to the normalization

$$\langle n_0',n_{\text{pair}}',\ell'|n_0,n_{\text{pair}},\ell\rangle = \delta_{n_0',n_0}\delta_{n_{\text{pair}}',n_{\text{pair}}}\delta_{\ell',\ell}, \quad (87)$$
$$\langle n_0,\lambda,t|n_0,\lambda,t\rangle = 1,$$

if $\gamma(t) = \alpha(t)\beta(t)$ with

$$\alpha^2(t) = \sum_{n_{\text{pair}}=0}^{\infty}\alpha_{n_{\text{pair}}}^2(t) = \alpha(t)\cdot\alpha(t), \quad (88)$$
$$\beta^2(t) = \sum_{\ell=0}^{\infty}\beta_\ell^2(t) = \beta(t)\cdot\beta(t).$$

The expansion coefficients $\{\alpha_{n_{\text{pair}}}(t)\}$ and $\{\beta_\ell(t)\}$ are vectors in infinite-dimensional spaces.

In principle, the scalar cubic interaction in four dimensions requires ultraviolet regularization. However, the issue of regularization and the question of stabilty are qualitatively unrelated. For example, the cubic interaction is also unstable in dimensions lower than four, where there is no need for regularization. The ultraviolet regularization would have an effect on the behavior of functions $f(p)$ and $g(p)$, which are left unspecified in this discussion except for their normalization.

The matrix element (80) can now be evaluated. Assuming that $f(k) = f(-k)$ and $g(k) = g(-k)$, it becomes

$$M_{n_0,\lambda}(t) = \{n_0 + 2L(t)\}\,\tilde{m} + G(t)\tilde{\mu} \quad (89)$$
$$- gV\{n_0 + 2L(t) + 2L_1(t)\}\sqrt{G_1(t)},$$

where the constants $\tilde{m}$, $\tilde{\mu}$, and $V$ are

$$\tilde{m} \equiv \int d\tilde{k} E_m(k) f^2(k), \quad (90)$$
$$\tilde{\mu} \equiv \int d\tilde{p} E_\mu(p) g^2(p),$$
$$V \equiv \int \frac{d\tilde{k}_m d\tilde{k}_m' f(k) f(k') g(k-k')}{\sqrt{m^2 + (\mathbf{k}-\mathbf{k}')^2}},$$

and the time-dependent quantities are

$$L(t) = \sum_{n_{\text{pair}}=0}^{\infty}\frac{n_{\text{pair}}\alpha_{n_{\text{pair}}}^2(t)}{\alpha^2(t)}, \quad (91)$$
$$G(t) = \sum_{\ell=0}^{\infty}\frac{\ell\beta_\ell^2(t)}{\beta^2(t)},$$
$$L_1(t) = \sum_{n_{\text{pair}}=1}^{\infty}\frac{\sqrt{n_0+n_{\text{pair}}}\sqrt{n_{\text{pair}}}\alpha_{n_{\text{pair}}}(t)\alpha_{n_{\text{pair}}-1}(t)}{\alpha^2(t)},$$
$$\sqrt{G_1(t)} = \sum_{\ell=1}^{\infty}\frac{\sqrt{\ell}\beta_\ell(t)\beta_{\ell-1}(t)}{\beta^2(t)}.$$

Note that $L$ and $G$ are the average number of matter pairs and mesons, respectively, in the intermediate state.

The variational principle tells us that the correct mass must be equal to or larger than (89). This inequality may be simplified by using the Schwarz inequality to place an upper limit on the quantities $L_1$ and $G_1$. Introducing the vectors

$$f_1 = \{\alpha_1,\sqrt{2}\alpha_2,\ldots\} = \{\sqrt{n}\alpha_n\}, \quad (92)$$
$$f_2 = \{\sqrt{n_0+1}\alpha_0,\sqrt{n_0+2}\alpha_1,\ldots\}$$
$$= \{\sqrt{n_0+n}\alpha_{n-1}\},$$
$$h = \{\beta_1,\sqrt{2}\beta_2,\ldots\} = \{\sqrt{\ell}\beta_\ell\},$$

we may write

$$L_1(t) = \frac{f_1(t)\cdot f_2(t)}{\alpha^2(t)} \le \frac{\sqrt{f_1^2(t)f_2^2(t)}}{\alpha^2(t)} \quad (93)$$

$$= \sqrt{L(t)\{n_0 + 1 + L(t)\}},$$

$$\sqrt{G_1(t)} = \frac{h(t)\cdot\beta(t)}{\beta^2(t)} \le \frac{\sqrt{h^2(t)\beta^2(t)}}{\beta^2(t)} = \sqrt{G(t)}.$$

Hence, suppressing explicit reference to the time dependence of $L$ and $G$ and using the inequalities (93), Eq. (89) can be converted into an inequality

$$M_{n_0,\lambda}(t) \ge (n_0 + 2L)\,\tilde{m} + G\tilde{\mu} \qquad (94)$$
$$- gV\left\{\left(\sqrt{n_0 + 1 + L} + \sqrt{L}\right)^2 - 1\right\}\sqrt{G}.$$

Minimization of the ground-state energy with respect to the average number of mesons $G$ occurs at

$$\sqrt{G_0} = \frac{gV}{2\tilde{\mu}}\left\{\left(\sqrt{n_0 + 1 + L} + \sqrt{L}\right)^2 - 1\right\}. \quad (95)$$

At this minimum point, the ground-state energy is bounded by

$$M_{n_0,\lambda}(t) \ge \{n_0 + 2L\}\,\tilde{m} - \mu G_0. \qquad (96)$$

If we continue with the minimization process, we would obtain $M_{n_0,\lambda}(t) \to -\infty$ as $L \to \infty$, providing no lower bound and hence suggesting that the state is unstable. However, if $L$ is finite, this result shows that the ground state is stable for couplings in the interval $0 < g^2 < g^2_{\text{crit}}$ with

$$g^2_{\text{crit}} \equiv \frac{4\tilde{\mu}\tilde{m}(n_0 + 2L)}{V^2\left\{\left(\sqrt{n_0 + 1 + L} + \sqrt{L}\right)^2 - 1\right\}^2}. \quad (97)$$

This interval is nonzero if the number of matter particles, $n_0$, and the average number of $\chi\bar{\chi}$ pairs, $L$, are finite. In particular, if there are no $Z$ diagrams or $\chi\bar{\chi}$ loops in the intermediate states, then the ground state will be stable for a limited range of values of the coupling.

This result also suggests strongly that the system is unstable when $g^2 > g^2_{\text{crit}}$ or when $L \to \infty$ (implying that $g^2_{\text{crit}} \to 0$). However, since Eq. (96) is only a lower bound, our argument does not provide a proof of these latter assertions.

We now discuss the effect of the $Z$ graphs and of the matter loops on the stability. Using FSR, we will show that the ground state is (i) stable when $Z$ diagrams are included in intermediate states, but (ii) unstable when matter loops are included.

The covariant trajectory $z(\tau)$ of the particle is parametrized in the FSR as a function of the proper time $\tau$. In $\chi^2\phi$ theory, the FSR expression for the one-body propagator for a dressed $\chi$ particle in the quenched approximation in Euclidean space was given qualitatively in Eq. (65). The detailed expression, needed in the following discussion, is

$$G(x,y) = \int_0^\infty ds \left[\frac{N}{4\pi s}\right]^{2N} \prod_{i=1}^{N-1}\int d^4 z_i \qquad (98)$$
$$\times \exp\left\{-K[z,s] - V[z,s_r]\right\}.$$

Here, the integrations are over all possible particle trajectories (discretized into $N$ segments with $N-1$ variables $z_i$ and boundary conditions $z_0 = x$ and $z_N = y$), and the kinetic and self-energy terms are

$$K[z,s] = m^2 s + \frac{N}{4s}\sum_{i=1}^{N}(z_i - z_{i-1})^2, \qquad (99)$$

$$V[z,s] = -\frac{g^2 s^2}{2N^2}\sum_{i,j=1}^{N}\Delta\left(\delta z_{ij},\mu\right), \qquad (100)$$

where $\Delta(z,\mu)$ is the Euclidean propagator of the meson (suitably regularized), $\delta z_{ij} = \frac{1}{2}(z_i + z_{i-1} - z_j - z_{j-1})$, and

$$s_r \equiv \frac{s}{R(s,s_0)} = \frac{s}{1 + (s - s_0)^2/\Gamma^2}. \qquad (101)$$

(The need for the substitution $s \to s_r$ was discussed above in Section 4.)

In preparation for a discussion of the effects of $Z$ diagrams and loops, we first discuss the stability of Eq. (98) when neither $Z$ diagrams nor loops are present. To make the discussion explicit, consider the one-body propagator in $0+1$ dimension. Since the integrals converge, we make the crude approximation that each $z_i$ integral is approximated by one point (since we are excluding $Z$ diagrams, the points may lie along the classical trajectory). If the boundary conditions are $z_0 = 0$ and $z_N = T$, the points along the classical trajectory are $z_i = iT/N$, and

$$K[z,s] = m^2 s \qquad (102)$$
$$+ \frac{N}{4s}\sum_{i=1}^{N}(z_i - z_{i-1})^2 = m^2 s + \frac{T^2}{4s}.$$

If the interaction is zero, this has a stationary point at $s = s_0 = T/(2m)$, giving

$$K[z,s] = K_0 = mT, \qquad (103)$$

yielding the expected free particle mass $m$. [Note that half of this result comes from the sum over $(z_i - z_{i-1})^2$.] The potential term (100) may be similarly evaluated; it gives a negative contribution that reduces the mass.

We now turn to a discussion of the effect of $Z$ diagrams. For the simple estimate of the kinetic energy,

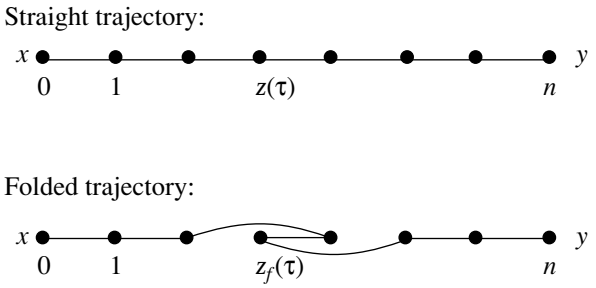


**Fig. 18.** It is possible to create particle–antiparticle pairs using folded trajectories. However, folded trajectories are suppressed by the kinematics, as discussed in the text.

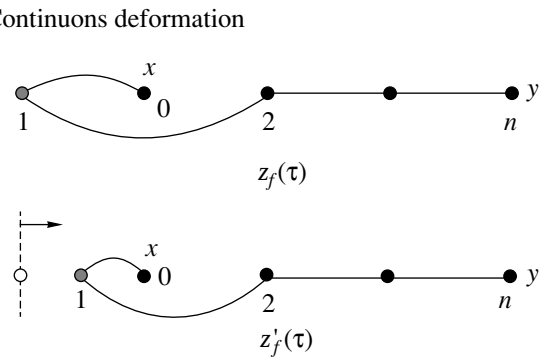


**Fig. 19.** A folded trajectory at the end point of the path and a similar one with $z_1$ closer to $z_0$.

Eq. (102), we choose integration points $z_i = iT/N$ uniformly spaced along a line. The classical trajectory connects these points without doubling back, so that they increase monotonically with proper time $\tau$. However, since the integration over each $z_i$ is independent, there also exist trajectories where $z_i$ does not increase monotonically with $\tau$. In fact, for every choice of integration points $z_i$, there exist trajectories with $z_i$ monotonic in $\tau$ and trajectories with $z_i$ nonmonotonic in $\tau$. The latter double back in time and describe $Z$ diagrams in the path-integral formalism. Two such trajectories that pass through the same points $z_i$ are shown in Fig. 18. These two trajectories contain the same points, $z_i$, but ordered in different ways, and both occur in the path integral.

Now, since the total self-energy is the sum of potential contributions $V[z,s]$ from all $(z_i, z_j)$ pairs, irrespective of how these coordinates are ordered, it must be the same for the straight trajectory $z(\tau)$ and the folded trajectory $z_f(\tau)$:

$$V[z_f, s] = V[z, s]. \quad (104)$$

However, according to Eq. (99), the kinetic energy of the folded trajectory is larger than the kinetic energy of the straight trajectory

$$K[z_f, s] > K[z, s], \quad (105)$$

because it includes some terms with larger values of $(z_i - z_{i-1})^2$. Since the kinetic energy term is always positive, the folded trajectory ($Z$ graph) is always suppressed (has a larger exponent) compared with a corresponding unfolded trajectory (provided, of course, that $g^2 < g^2_{\text{crit}}$).

This argument holds only for cases where the trajectory does not double back to times before $z_0 = 0$ or after $z_N = T$. An example of such a trajectory is shown in Fig. 19 (upper panel). Here, we compare this folded trajectory to another folded trajectory $z'_f$ with point $z_1$ closer to the starting point $z_0$ (lower panel of Fig. 19). This new folded trajectory has points spaced closer together, so that the kinetic energy is smaller and the potential energy is larger, and therefore

$$K[z_f, s] - V[z_f, s] > K[z'_f, s] - V[z'_f, s]. \quad (106)$$

It is clear that the larger the folding in the trajectory, the less energetically favorable the path, and the most favorable path is again an unfolded trajectory with no points outside of the limits $z_0 < z_i < z_N$.

While these arguments have been stated in $0 + 1$ dimension for simplicity, they are not dependent on the number of dimensions and can be extended to the realistic case of $1 + 3$ dimensions.

We conclude that a calculation in the quenched approximation, where the creation of particle–antiparticle pairs can only come from $Z$ graphs, must be more stable (produce a larger mass) than a similar calculation without any $\chi\bar{\chi}$ pairs. The quenched $\chi^2\phi$ theory therefore is bounded by the same limits given in Eq. (97). This conclusion supports, and is supported by, the results of [13, 30, 32] which show, in the quenched approximation, that the $\chi^2\phi$ interaction is stable for a finite range of coupling strengths.

It is now clear that the instability of $\chi^2\phi$ theory must be due to either (i) the possibility of creating an infinite number of closed $\chi\bar{\chi}$ loops or (ii) the presence of an infinite number of matter particles (as in an infinite medium). Indeed, the original proof given by Baym used the possibility of loop creation from the vacuum to prove that the vacuum was unstable.

These results provide justification for the stability of relativistic one-boson-exchange models that usually exclude matter loops but may include $Z$ diagrams of all orders. Our argument cannot be easily extended to symmetric $\phi^3$ theories where it is impossible to make a clear distinction between $Z$ diagrams and loops.

## 6. CONCLUSIONS

In this paper, we have given a summary of results for scalar interactions obtained with the use of the FSR representation. The FSR approach uses a covariant path-integral representation for the trajectories of particles. Reduction of field-theoretical path integrals to path integrals involving particle trajectories reduces the dimensionality of the problem and the associated computational cost.

Applications of the FSR approach to one- and two-body problems, in particular, shows that uncontrolled approximations in field theory may lead to significant deviations from the correct result. Our results indicate that use of the Bethe–Salpeter equation in the ladder approximation to solve the two-body bound-state problem is a poor approximation. For the scalar theories examined here, a better approximation to the two-body problem is obtained using the Gross equation in the ladder approximation. Similarly, use of the rainbow approximation for the one-body problem gives a poorer result than simply calculating the self-energy to second order! In all of these cases, the explanation for these results seems to be that the crossed diagrams (such as crossed ladders) play an essential role, canceling contributions from higher order ladder or rainbow diagrams.

## ACKNOWLEDGMENTS

We thank Prof. Yu. Simonov for his leadership in this field and for many useful discussions. It is a pleasure to contribute to this volume celebrating his 70th birthday.

This work was supported in part by the DOE grant DE-FG02-93ER-40762 and DOE contract DE-AC05-84ER-40150, under which the Southeastern Universities Research Association (SURA) operates the Thomas Jefferson National Accelerator Facility.

## REFERENCES

1. R. P. Feynman, Phys. Rev. **80**, 440 (1950).
2. Julian S. Schwinger, Phys. Rev. **82**, 664 (1951).
3. Yu. A. Simonov, Nucl. Phys. B **307**, 512 (1988).
4. Yu. A. Simonov, Nucl. Phys. B **324**, 67 (1989).
5. Yu. A. Simonov, Yad. Fiz. **54**, 192 (1991) [Sov. J. Nucl. Phys. **54**, 115 (1991)].
6. Yu. A. Simonov and J. A. Tjon, Ann. Phys. (N.Y.) **228**, 1 (1993).
7. Yu. A. Simonov and J. A. Tjon, *Michael Marinov Memorial Volume: Multiple Facets of Quantization and Supersymmetry,* Ed. by M. Olshanetsky and A. Vainshtein (World Sci., Singapore, 2002), p. 369; hep-ph/0201005.
8. Yu. A. Simonov, J. A. Tjon, and J. Weda, Phys. Rev. D **65**, 094013 (2002).
9. Yu. A. Simonov and J. A. Tjon, Ann. Phys. (N.Y.) **300**, 54 (2002); hep-ph/0205165.
10. E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).
11. F. Gross and J. Milana, Phys. Rev. D **43**, 2401 (1991).
12. P. C. Tiemeijer and J. A. Tjon, Phys. Rev. C **49**, 494 (1994).
13. C. Savkli and F. Gross, Phys. Rev. C **63**, 035208 (2001); hep-ph/9911319.
14. M. Levine, J. Wright, and J. A. Tjon, Phys. Rev. **154**, 1433 (1967).
15. N. Nakanishi, Prog. Theor. Phys. Suppl. **43**, 1 (1969).
16. N. Nakanishi, Prog. Theor. Phys. Suppl. **95**, 1 (1988).
17. T. Nieuwenhuis and J. A. Tjon, Few-Body Syst. **21**, 167 (1996).
18. C. Savkli and F. Tabakin, Nucl. Phys. A **628**, 645 (1998); hep-ph/9702251.
19. E. E. Salpeter, Phys. Rev. **87**, 328 (1952).
20. A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **29**, 380 (1963).
21. R. Blankenbecler and R. Sugar, Phys. Rev. **142**, 1051 (1966).
22. F. Gross, Phys. Rev. **186**, 1448 (1969).
23. F. Gross, Phys. Rev. C **26**, 2203 (1982).
24. S. J. Wallace and V. B. Mandelzweig, Nucl. Phys. A **503**, 673 (1989).
25. Taco Nieuwenhuis, Yu. A. Simonov, and J. A. Tjon, Few-Body Syst. Suppl. **7**, 286 (1994).
26. Taco Nieuwenhuis and J. A. Tjon, Phys. Lett. B **355**, 283 (1995).
27. Taco Nieuwenhuis and J. A. Tjon, Phys. Rev. Lett. **77**, 814 (1996); hep-ph/9606403.
28. C. Savkli, F. Gross, and J. Tjon, Phys. Rev. C **60**, 055210 (1999); hep-ph/9906211.
29. C. Savkli, F. Gross, and J. Tjon, Phys. Rev. D **62**, 116006 (2000); hep-ph/9907445.
30. C. Savkli, Comput. Phys. Commun. **135**, 312 (2001); hep-ph/9910502.
31. F. Gross, C. Savkli, and J. Tjon, Phys. Rev. D **64**, 076008 (2001); nucl-th/0102041.
32. C. Savkli, Czech. J. Phys. **51B**, 71 (2001); hep-ph/0011249.
33. C. Savkli, F. Gross, and J. Tjon, Phys. Lett. B **531**, 161 (2002); nucl-th/0202022.
34. B.-F. Ding, Nucl. Phys. B (Proc. Suppl.) **90**, 127 (2000); nucl-th/0008048.
35. D. R. Phillips, S. J. Wallace, and N. K. Devine, Phys. Rev. C **58**, 2261 (1998); nucl-th/9802067.
36. R. Rosenfelder and A. W. Schreiber, Phys. Rev. D **53**, 3337 (1996); nucl-th/9504002.
37. R. Rosenfelder and A. W. Schreiber, Phys. Rev. D **53**, 3354 (1996); nucl-th/9504005.
38. F. J. Dyson, Phys. Rev. **85**, 631 (1952).
39. G. Baym, Phys. Rev. **117**, 886 (1960).
40. B.-F. Ding and J. W. Darewych, J. Phys. G **26**, 907 (2000); nucl-th/9908022.
41. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, 1989).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 861–869. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 894–903.*


# Theoretical Evidence for a Tachyonic Ghost-State Contribution to the Gluon Propagator in High-Energy, Forward Quark–Quark "Scattering"*

**A. I. Karanikas**** **and C. N. Ktorides*****

*Physics Department Nuclear & Particle Physics Section, University of Athens, Panepistimiopolis, Athens, Greece*

Received May 13, 2004; in final form, September 20, 2004

**Abstract**—Implications stemming from the inclusion of nonperturbative confining effects, as contained in the stochastic vacuum model of H. Dosch and Yu.A. Simonov, are considered in the context of a (hypothetical) quark–quark "scattering process" in the Regge kinematical region. In a computation wherein the nonperturbative input enters as a correction to established perturbative results, a careful treatment of infrared divergences is shown to imply the presence of an effective propagator associated with the existence of a linear term in the static potential. An equivalent statement is to say that the modified gluonic propagator receives a contribution from a tachyonic ghost state, an occurrence which is fully consistent with earlier suggestions made in the context of low-energy QCD phenomenology.


## 1. INTRODUCTION

From the theoretical point of view, forward scattering at very high energies (Regge kinematics) presents a situation where one can readily apply eikonal approximation techniques [1]. In this context, such processes provide grounds for exploring long-distance properties of the underlying fundamental theory for the implicated interaction. For the particular case of QCD, long-distance behavior constitutes a fundamental issue whose exploration is not only relevant to high-energy processes but also to low-energy phenomenology.

In the present paper, we revisit the (idealized) problem of quark–quark "scattering" in the Regge limit, which has been extensively studied within the framework of pQCD [1–5], with the aim to extend the aforementioned analyses in a direction which takes into account confining aspects of the theory. Specifically, we shall rely on the premises of the stochastic vacuum model (SVM), pioneered by Dosch and Simonov [6] for the explicit purpose of accomodating the confinement property of QCD. A similar approach has been pursued by Nachtmann [7] by employing a different methodology from the one we shall adopt in this work. Different will also be the scope of the present analysis.

The construction of the SVM is motivated by the intention to incorporate established observations/results regarding the structure of the QCD vacuum [8] into a well-defined theoretical framework. In particular, it summarizes all that is known and/or surmised about its properties through a set of three axioms, which are expressed in terms of field strength, as opposed to field potential, correlators. The underlying stochasticity assumption for the vacuum state facilitates the application of the cumulant expansion [9], which describes the factorization rules for higher order gluon field strength correlators in terms of two-point ones. One of the first results arrived at through the SVM is the deduction of the area law for the static Wilson loop, i.e., confinement. Specific applications of the SVM scheme, including comparisons with lattice results, can be found, e.g., in [10].

A concrete, as well as practical, way to apply the SVM scheme to specific situations has been suggested by Simonov [11]. The idea is to use the background gauge fixing method [12] and assign the background gauge fields with the task of becoming the agents of the nonperturbative dynamics. Specifically, one employs the gauge potential splitting $A^a_\mu = \alpha^a_\mu + B^a_\mu$ with the $\alpha^a_\mu$ being associated with the usual perturbative-field modes. The $B^a_\mu$, on the other hand, enter as *dynamical* fields, assigned with the task of carrying the nonperturbative physics through field strength correlators which adhere to the cumulant expansion rules.

*This article was submitted by the authors in English.

**e-mail: akaranik@cc.uoa.gr

***e-mail: cktorid@cc.uoa.gr

Following our previous work of [5], we find it convenient to employ the FFS-worldline casting of QCD, originally pioneered by Fock [13], Feynman [14], and Schwinger [15] and most recently revived in path-integral versions (see [16–18]). The reason for such a choice is that the eikonal approximation acquires a straightforward realization in this scheme, since, in the perturbative context at least, it can be readily implemented by restricting one's considerations to straight worldline paths. As it will turn out, the inclusion of input from the SVM will produce a kind of deformation of the eikonal paths, with low-energy consequences, which represent subleading, perturbative–nonperturbative interference effects. The conditions which justify the relevant computation will be made explicit in the text. Suffice it to say, at this point, that it was explicitly demonstrated in [19] that the FFS-worldline formulation of QCD, in combination with the background gauge field splitting, is ideally suited for providing a framework within which one can directly and efficiently apply the premises of the SVM.

The main result of this paper is the following: Once (subleading) contributions incorporating nonperturbative, as induced by the SVM, corrections to the perturbative expressions are taken into account, then a consistent analysis of the infrared issues entering the quark–quark high-energy forward-"scattering" process reveals that the gluonic propagator exhibits a behavior which can be interpreted in terms of the presence of a "tachyonic mass pole." Arguments in favor of such an occurrence, with important phenomenological as well as theoretical implications, have been promoted, from different perspectives, in several papers [20–22].

The organization of the paper is as follows. In the next section, we present the basic formulas related to the "amplitude" for the high-energy "quark–quark scattering process" in the forward direction and display their FFS-worldline form. Section 3 focuses its attention on infrared issues associated with the pertubative–nonperturbative interference effects under consideration in this study. The resulting expression for the amplitude is shown to be equivalent to the introduction of an effective propagator, which is associated with the presence of a linear term in the static potential. Section 4 extends the implications of the aforementioned result to issues related to renormalization: modified running coupling constant and summation of leading logarithms via the Callan–Symanzyk equation. Finally, the technical manipulations leading to the main result of Section 3 are displayed in the Appendix.

## 2. PRELIMINARY CONSIDERATIONS

Consider an idealized quark–quark scattering process in the Regge limit, defined by $s/m^2 \to \infty$, $s \gg t(= -q^2 = \mathbf{q}_\perp^2)$. The amplitude is given by

$$T_{ii'jj'}(s/m^2, q_\perp^2/\lambda^2) = \int d^2b e^{-i\mathbf{q}\cdot\mathbf{b}} E_{ii'jj'}(s/m^2, 1/b^2\lambda^2), \tag{1}$$

where $\lambda$ is an "infrared" scale; $b$ is the impact distance in the transverse plane to the direction of the colliding quarks (assumed to be traveling along the $x_3$ axis); and $E_{ii'jj'}$, which incorporates the dynamics of the process, is specified, in Euclidean spacetime, by [5]

$$E_{ii'jj'} = \left\langle P\exp\left[ig\int\limits_{-\infty}^{\infty} d\tau v_1\cdot A(v_1\tau)\right]_{ii'} \times P\exp\left[ig\int\limits_{-\infty}^{\infty} d\tau v_2\cdot A(v_2\tau+b)\right]_{jj'}\right\rangle_A^{\mathrm{conn}}. \tag{2}$$

The $v_1$ and $v_2$ are constant four-velocities characterizing the respective eikonal lines of the quarks participating in the scattering process and "conn" stands for "connected." In Minkowski space (and in light cone coordinates), one has $v_1 \simeq (v_1^+, 0, \mathbf{0}_\perp)$, $v_2 \simeq (0, v_2^-, \mathbf{0}_\perp)$, $b \simeq (0, 0, \mathbf{b}_\perp)$ with $v_1^+ \simeq v_2^- \simeq \frac{1}{\sqrt{2}}\sqrt{s}/m$.

Employing the gauge field splitting $A_\mu^a = \alpha_\mu^a + B_\mu^a$, the expectation values with respect to field configurations acquire the form $\langle\cdots\rangle_A = \langle\cdots\rangle_{\alpha,B}$. Expanding in terms of powers of the perturbative field components, one obtains

$$E_{ii'jj'} = \left\langle P\exp\left[ig\int\limits_{-\infty}^{\infty} d\tau v_1\cdot B(v_1\tau)\right]_{ii'} \times P\exp\left[ig\int\limits_{-\infty}^{\infty} d\tau v_2\cdot B(v_2\tau+b)\right]_{jj'}\right\rangle_B^{\mathrm{conn}} - g^2 \int\limits_{-\infty}^{\infty} ds_2 \int\limits_{-\infty}^{\infty} ds_1 \left\langle P\exp\left[ig\int\limits_{s_1}^{\infty} d\tau v_1\cdot B(v_1\tau)\right]_{ik} \times P\exp\left[ig\int\limits_{-\infty}^{s_1} d\tau v_1\cdot B(v_1\tau)\right]_{li'} \times P\exp\left[ig\int\limits_{s_2}^{\infty} d\tau v_2\cdot B(v_2\tau+b)\right]_{jm} \tag{3}$$

$$\times P\exp\left[ig\int\limits_{-\infty}^{s_2} d\tau v_2\cdot B(v_2\tau+b)\right]_{nj'} t^a_{kl}t^b_{mn}$$

$$\times v_{2\mu}v_{1\nu}iG^{ba}_{\mu\nu}(v_2s_2+b,v_1s_1)\Big\rangle_B + \mathcal{O}(g^4\langle\alpha^4\rangle).$$

Our objective in this paper is to study the behavior of the amplitude as $|b|\to 0$. Accordingly, contributions attributed exclusively to the nonperturbative, background terms will be ignored given that, by definition, they are finite in this limit. This narrows the expression of computational interest to the following one:

$$E_{ii'jj'} = -g^2\frac{N_c}{N_c^2-1}t^a_{ii'}t^a_{jj'}\int\limits_{-\infty}^{\infty} ds_2 \qquad (4)$$

$$\times\int\limits_{-\infty}^{\infty} ds_1 v_{2\mu}v_{1\nu}\left\langle\frac{1}{N_c}\mathrm{Tr}_A iG_{\mu\nu}(v_2s_2+b,v_1s_1)\right\rangle_B$$

$$+\mathcal{O}(g^4\langle\alpha^4\rangle)+\mathcal{O}(b^2).$$

The propagator

$$iG^{ba}_{\mu\nu}(v_2s_2+b,v_1s_1)\equiv\langle\alpha^b_\mu(v_2s_2+b)\alpha^a_\nu(v_1s_1)\rangle$$

acquires the following worldline expression:

$$iG^{ba}_{\mu\nu}(v_2s_2+b,v_1s_1) \qquad (5)$$

$$=\int\limits_0^\infty dT\int\limits_{\substack{x(0)=v_1s_1\\ x(T)=v_2s_2+b}}\mathcal{D}x(t)\exp\left[-\frac{1}{4}\int\limits_0^T dt\dot{x}^2(t)\right]$$

$$\times P\left(\exp\left[g\int\limits_0^T dtJ\cdot F+ig\int\limits_0^T dt\dot{x}\cdot B\right]\right)^{ba}_{\mu\nu}.$$

In the above formula, $(J\cdot F)_{\mu\nu}=J^{\alpha\beta}_{\mu\nu}F_{\alpha\beta}$, with $J^{\alpha\beta}_{\mu\nu}$ being the generators for the spin-1 representation of the Lorentz group.[1)] It should also be noted that we have employed the notation $B^{ba}_\mu=B^c_\mu(t^c_G)^{ab}=-B^c_\mu f^{abc}$.

To close this section, let us briefly comment on gauge-symmetry-related issues. Given the "idealized" process under consideration, our handling of gauge invariance will be to let the two worldlines of the "colliding" quarks extend to infinity in both directions and impose the boudary conditions $A_\mu[x(t)]\to 0$ as $t_{\mathrm{Eucl}}\to\pm\infty$. In this way, the overall worldline configuration introduces a Wilson loop in the path integrals, given that the endpoints of the two trajectories can now be joined at $+\infty$ and at $-\infty$. These, of course, do not correspond to boundary conditions for a scattering process per se; however, our only objective in this work is to extract long-distance implications based on the exchanges taking place in the immediate vicinity of the points of closest approach between the two worldlines. The study of realistic situations involving the scattering of physically observable particle entities in the Regge limit, using the presently adopted methodology, is under current consideration and the relevant analysis will be presented in the near future. Finally, concerning the issue of gauge fixing for the $B$-field sector, our choice is prompted by the intention to rely on field strength correlators for describing nonperturbative dynamics [11]. It, accordingly, becomes convenient to work in the Fock–Schwinger (FS) gauge [23, 24]. The latter is specified by

$$B^a_\mu(x) = -\int\limits_{x_0}^{x} du_\nu(\partial_\mu u_\rho)F^a_{\rho\nu}(u) \qquad (6)$$

$$=-(x-x_0)_\nu\int\limits_0^1 d\alpha\alpha F_{\mu\nu}(x_0+\alpha(x-x_0)).$$

Of course, the arbitrary point $x_0$ should not enter any gauge-invariant expression.

## 3. INFRARED ISSUES ASSOCIATED WITH THE PROPAGATION OF GLUONIC MODES IN A CONFINING ENVIRONMENT

Consider the quantity defined by

$$I(l)\equiv\frac{1}{N_c^2-1}v_{2\mu}v_{1\nu}\langle i\mathrm{Tr}_A G_{\mu\nu}(l)\rangle_B \qquad (7)$$

$$=\frac{N_c}{N_c^2-1}v_{2\mu}v_{1\nu}\int\limits_0^\infty dT\int\limits_{\substack{x(0)=0\\ x(T)=l}}\mathcal{D}x(t)$$

$$\times\exp\left[-\frac{1}{4}\int\limits_0^T dt\dot{x}^2(t)\right]\left\langle\frac{1}{N_c}\mathrm{Tr}_A P\right.$$

$$\left.\times\exp\left(g\int\limits_0^T dtJ\cdot F+ig\int\limits_0^T dt\dot{x}\cdot B\right)_{\mu\nu}\right\rangle_B,$$

where we have introduced $l_\mu\equiv v_{2\mu}s_2-v_{1\mu}s_1-b_\mu$. It describes the propagation of the perturbative gluon modes in the presence of the background gauge field modes $B^a_\mu$ and, in this sense, it is expected to incorporate confinement effects associated with the SVM.

[1)] Its incorporation into the worldline form of the propagator serves to signify the spin of the accomodated modes.

Generally speaking, one would expect that, in a study of a physically relevant process with the full (and proper) inclusion of nonperturbative effects, no need would arise for the introduction of an infrared cutoff to regulate the various expressions entering the computation at long distances. An infrared scale should, in other words, naturally arise suppressing contributions from very large distances ($|l| \to \infty$). Given that the object of the present study is to investigate perturbative/nonperturbative interference, long-distance effects in the (nonphysical) process of quark–quark high-energy "collision" in the forward direction, it becomes necessary to regulate infrared divergences associated with the upper limit of the $T$ integral. Our choice of introducing the infrared cutoff is via the replacement

$$\int_0^\infty (\cdots) dT \to \int_0^\infty dT e^{-T\lambda^2} (\cdots).$$

On a simple dimensional basis and given the length scales entering the problem, one could make the association $\lambda \propto \sigma |l|$.

Upon expanding the exponential term in Eq. (7), one obtains

$$I(l) = \frac{v_1 \cdot v_2}{4\pi^2 |l|^2} - \frac{2N_c^2}{N_c^2 - 1} v_{2\mu} v_{1\nu} \tag{8}$$
$$\times \int_0^\infty dT e^{-T\lambda^2} \int_0^T dt_2 \int_0^T dt_1 \theta(t_2 - t_1)$$
$$\times \int\limits_{\substack{x(0)=0 \\ x(T)=l}} \mathcal{D}x(t) \exp\left[ -\frac{1}{4} \int_0^T dt \dot{x}^2(t) \right]$$
$$\times \left\{ \delta_{\mu\nu} \frac{1}{N_c} \mathrm{Tr}_F \langle g\dot{x}(t_2) \cdot B(x(t_2)) g\dot{x}(t_1) \right.$$
$$\times B(x(t_1)) \rangle_B + \frac{2}{N_c} \mathrm{Tr}_F \langle g F^c_{\mu\rho}(x(t_2))$$
$$\left. \times g F^c_{\rho\nu}(x(t_1)) \rangle_B \right\} + \mathcal{O}(\langle g^4 F^4 \rangle_B).$$

Observe now that, in the FS gauge, the following relation holds,

$$\mathrm{Tr}_F \langle g B_{\mu_2}(x(t_2)) g B_{\mu_1}(x(t_1)) \rangle_B \tag{9}$$
$$= (x_2 - x_0)_{\nu_2} (x_1 - x_0)_{\nu_1} \int_0^1 d\alpha_2 \alpha_2 \int_0^1 d\alpha_1 \alpha_1$$
$$\times \mathrm{Tr}_F \langle g F_{\mu_2\nu_2}(x_0 + \alpha_2(x_2 - x_0))$$
$$\times g F_{\mu_1\nu_1}(x_0 + \alpha_1(x_1 - x_0)) \rangle_B,$$

which brings into play the field strength correlator.

Setting $u_i = x_0 + \alpha_i x(t)$, $i = 1, 2$, one writes

$$\frac{1}{2N_c} \langle g F^c_{\mu_2\nu_2}(u_2) g F^c_{\mu_1\nu_1}(u_1) \rangle_B \tag{10}$$
$$= \frac{1}{N_c} \mathrm{Tr}_F \langle \phi(x_0, u_2) g F_{\mu_2\nu_2}(u_2)$$
$$\times \phi(u_2, x_0) \phi(x_0, u_1) g F_{\mu_1\nu_1}(u_1) \phi(u_1, x_0) \rangle_B$$
$$\equiv \Delta^{(2)}_{\mu_2\nu_2, \mu_1\nu_1}(u_2 - u_1),$$

where

$$\phi(x_0, u_i) = P \exp\left( ig \int_{u_i}^{x_0} dv \cdot B(v) \right)$$

and is unity in the FS gauge. Its insertion serves to underline the gauge invariance of the field strength correlator.

With the above in place and upon making in Eq. (8) the redefinition $t_i \to T t_i$, $i = 1, 2$, one determines

$$I(l) = \frac{v_1 \cdot v_2}{4\pi^2 |l|^2} - \frac{2N_c^2}{N_c^2 - 1} v_1 \cdot v_2 \int_0^1 d\alpha_2 \alpha_2 \tag{11}$$
$$\times \int_0^1 d\alpha_1 \alpha_1 \int_0^1 dt_2 \int_0^1 dt_1 \theta(t_2 - t_1) \int_0^\infty dT T^2 e^{-T\lambda^2}$$
$$\times \int\limits_{\substack{x(0)=0 \\ x(T)=l}} \mathcal{D}x(t) \exp\left[ -\frac{1}{4} \int_0^1 dt \dot{x}^2(t) \right]$$
$$\times \left\{ 16 \frac{v_{2\mu} v_{1\nu}}{v_2 \cdot v_1} \Delta^{(2)}_{\mu\rho, \rho\nu}(x(t_2) - x(t_1)) \right.$$
$$+ \frac{1}{T^2} \dot{x}_{\mu_2}(t_2) x_{\nu_2}(t_2) \dot{x}_{\mu_1}(t_1) x_{\nu_1}(t_1) \Delta^{(2)}_{\mu_2\nu_2, \mu_1\nu_1}$$
$$\left. \times [\alpha_2 x(t_2) - \alpha_1 x(t_1)] \right\} + \mathcal{O}(\langle g^4 F^4 \rangle_B).$$

On a kinematic basis, the correlator can be represented as follows [6, 10]:

$$\Delta^{(2)}_{\mu_2\nu_2, \mu_1\nu_1}(z_2 - z_1) \tag{12}$$
$$= (\delta_{\mu_2\mu_1}\delta_{\nu_2\nu_1} - \delta_{\mu_2\nu_1}\delta_{\nu_2\mu_1}) D(z^2)$$
$$+ \frac{1}{2} \frac{\partial}{\partial z_{\mu_1}} \left[ (z_{\mu_2}\delta_{\nu_2\nu_1} - z_{\nu_2}\delta_{\mu_2\nu_1}) D_1(z^2) \right]$$
$$+ \frac{\partial}{\partial z_{\nu_1}} \left[ (z_{\nu_2}\delta_{\mu_2\mu_1} - z_{\mu_2}\delta_{\nu_2\mu_1}) D_1(z^2) \right].$$

One notices [10, 11] that the first term enters as a distinct feature of the non-Abelian nature of the gauge symmetry (it is not present, e.g., in QED). According

to the premises of the SVM, it is associated with the (QCD) string tension $\sigma$ by [10, 11]

$$\int_0^\infty dz^2 D(z^2) = \frac{1}{\pi}\int d^2 z D(z^2) \equiv \frac{2}{\pi}\sigma. \qquad (13)$$

Now, the central objective the present analysis is to determine first-order contributions to the amplitude coming, via the SVM, from the nonperurbative/confining sector of QCD. The corresponding lowest order correction is expected, on dimensional grounds, to be of the order of the string tension $\sigma$. This implies, as already pointed out by Simonov [25], that we shall set aside the $D_1$ term entering the kinematical analysis of the correlator, according to Eq. (12), which cannot furnish contributing terms of dimension $[m]^2$. In the Appendix, the following expression for the quantity $I(l)$ is established:

$$I(l) = \frac{1}{4\pi^2}\frac{v_1\cdot v_2}{|l|^2} \qquad (14)$$
$$\times\left[1+\alpha\sigma|l|^2\ln\left(C\frac{\sigma}{\lambda^2}\right)+\mathcal{O}(\sigma^2|l|^4)\right],$$

where

$$\alpha \equiv \frac{3}{\pi}\frac{N_c}{N_c^2-1}(1-\kappa)$$

with the constant parameter estimated to be $\kappa \simeq 0.5$. As pointed out by Simonov [21], the first—and most important—term contributing to $\alpha$ comes from the paramagnetic, attractive interaction between the spin of the gluons with the nonperturbative background field [cf. Eq. (5)]. It should also be noted that the constant $C$ entering the argument of the logarithm is connected with the choice of parametrization of $D(z^2)$ (see, e.g., [7]). In the context of a corresponding result having to do with an amplitude for a physically relevant, hence protected from infrared divergences, process, then any such parameter would disappear.

Suppose the following problem is now posed: Given the above results, which pertain to the gluon propagation in a confining environment, look for an equivalent effective particle-like mode propagation, summarizing their full content. In this spirit, we shall proceed to assess the possibility that the gist of all we have done up to here can be reproduced via the introduction of an effective propagator. Following [20], we make the substitution

$$\frac{1}{k^2} \to \frac{1}{k^2}+\frac{\mu^2}{k^4}, \qquad (15)$$

whose additional term signifies the presence of a linear term in the static potential. Then, one would obtain

$$I(l) = v_1\cdot v_2\int\frac{d^4k}{(2\pi)^4}e^{ik\cdot l}\left(\frac{1}{k^2}+\frac{\mu^2}{k^4}\right). \qquad (16)$$

The integral is infrared divergent and should require the introduction of a corresponding cutoff. Alternatively, one could restrict the validity of the replacement, according to Eq. (15), to the region $k^2 > \mu^2$. Then, one would determine

$$I(l) = v_1\cdot v_2\int\limits_{k^2>\mu^2}\frac{d^4k}{(2\pi)^2}e^{ik\cdot l}\left(\frac{1}{k^2}+\frac{\mu^2}{k^4}\right) \qquad (17)$$
$$\simeq \frac{v_1\cdot v_2}{|l|^2}\frac{1}{(2\pi)^4}\left[1+\frac{\mu^2|l|^2}{4}\ln\left(\frac{4}{e\mu^2|l|^2}\right)\right.$$
$$\left.+\mathcal{O}(\mu^4|l|^4)\right] \quad \text{for} \quad \mu^2|l|^2<1.$$

Comparing the above result with that of Eq. (14), one deduces

$$\mu^2 = 4\alpha\sigma = \frac{3}{\pi}\frac{N_c}{N_c^2-1}(1-\kappa)\sigma \qquad (18)$$
$$\simeq 2.15\sigma \simeq 0.4\ \mathrm{GeV}^2,$$

in full accord with the phenomenologically determined estimate for the tachyonic "pole" [20–22]. One also observes that $\lambda \sim |l|\sigma$, as per our original supposition.

Returning to the original full expression, which provides the full dynamical input for the amplitude, we write

$$E_{ii'jj'} \simeq g^2 t^a_{ii'}t^a_{jj'}v_1\cdot v_2\int\limits_{-\infty}^{+\infty}ds_2 \qquad (19)$$
$$\times\int\limits_{-\infty}^{+\infty}ds_1\frac{1}{4\pi^2|l|^2}\left[1+\sigma|l|^2\alpha\ln\left(C\frac{\sigma}{\lambda^2}\right)\right]$$
$$\simeq -g^2 t^a_{ii'}t^a_{jj'}v_1\cdot v_2\int\limits_{-\infty}^{+\infty}ds_2\int\limits_{-\infty}^{+\infty}ds_1\frac{1}{4\pi^2|l|^2}$$
$$\times\left[1+\frac{\mu^2|l|^2}{4}\ln\left(\frac{4}{e\mu^2|l|^2}\right)\right] \simeq -g^2 t^a_{ii'}t^a_{jj'}v_1\cdot v_2$$
$$\times\int\limits_{-\infty}^{+\infty}ds_2\int\limits_{-\infty}^{+\infty}ds_1\int\limits_{k^2>\mu^2}\frac{d^4k}{(2\pi)^2}e^{ik\cdot l}\left(\frac{1}{k^2}+\frac{\mu^2}{k^4}\right).$$

Concerning the logarithmic factors entering the above result, it is useful to remark that the various constants appearing in the arguments are not of any particular importance—at least in the approximation we have been working—given that they would disappear with the appropriate choice for the infrared

cutoff. More importantly, thinking in terms of the significance of the above results if they became part of an amplitude corresponding to a physically consistent, hence protected from infrared divergences, process, then any dependence from these constants should be absent.

Going over to Minkowski space, the previous relation assumes the form

$$E_{ii'jj'} \simeq -g^2 t^a_{ii'} t^a_{jj'} i v_1 \cdot v_2 \tag{20}$$
$$\times \int\limits_{-\infty}^{+\infty} ds_2 \int\limits_{-\infty}^{+\infty} ds_1 \int\limits_{k^2>\mu^2} \frac{d^4k}{(2\pi)^2} e^{-ik\cdot l}\left(-\frac{1}{k^2}+\frac{\mu^2}{k^4}\right).$$

The integration leads to the result

$$v_1 \cdot v_2 \int\limits_{-\infty}^{+\infty} ds_2 \int\limits_{-\infty}^{+\infty} ds_1 e^{-ik\cdot v_1 s_1 - ik\cdot v_2 s_2} \tag{21}$$
$$= (2\pi)^2 v_1 \cdot v_2 \delta(k\cdot v_1)\delta(k\cdot v_2)$$
$$= (2\pi)^2 \coth\gamma \delta(k_+)\delta(k_-),$$

where $\gamma$ is determined by

$$\cosh\gamma \equiv \frac{v_1\cdot v_2}{|v_1||v_2|} = \frac{s}{2m^2} \overset{s/m^2\gg 1}{\Rightarrow} \gamma \tag{22}$$
$$\simeq \ln(s/m^2) \Rightarrow \coth\gamma \simeq 1.$$

It follows

$$E_{ii'jj'} \simeq -\frac{g^2}{4\pi} t^a_{ii'} t^a_{jj'} i \coth\gamma f(b^2\mu^2), \tag{23}$$

where

$$f(b^2\mu^2) = \frac{1}{\pi} \int\limits_{k_\perp^2>\mu^2} d^2k_\perp e^{ik_\perp\cdot b}\left(\frac{1}{k_\perp^2}+\frac{\mu^2}{k_\perp^4}\right). \tag{24}$$

One immediately notices that, if $\mu^2 b^2 \ll 1$, then

$$f(\mu^2 b^2) \simeq \ln\left(\frac{4e}{\mu^2 b^2}\right), \tag{25}$$

which recovers the known perturbative result—with an infrared cutoff given by $\lambda^2 \equiv \mu^2/(4e)$.

As $b$ grows, while remaining in the region $\mu^2 b^2 < 1$, one finds

$$f(\mu^2 b^2) \simeq \ln\left(\frac{4e}{\mu^2 b^2}\right)\left(1-\frac{\mu^2 b^2}{4}\right)+\frac{1}{2}\mu^2 b^2. \tag{26}$$

In turn, this gives

$$E_{ii'jj'} \simeq -\frac{g^2}{4\pi} t^a_{ii'} t^a_{jj'} i \coth\gamma \tag{27}$$
$$\times \left\{\ln\left(\frac{4e}{\mu^2 b^2}\right)\left(1-\frac{\mu^2 b^2}{4}\right)+\frac{1}{2}\mu^2 b^2\right\}.$$

## 4. SUMMATION OF LARGE LOGARITHMS

The presence of terms $\sim g^2 \ln\left(1/b^2\mu^2\right)$, entering through the function $f(b^2\mu^2)$, imposes the need of their summation in the perturbative series. In the absence of the background field, i.e., in the framework of pQCD, it is well known that such a summation can be accomplished by employing the renormalization group strategies, which, for the quark–quark scattering process under consideration, can be justified on the basis that $1/b$ plays the role of an ultraviolet cutoff. As Simonov has shown [11], the presence of the background field does not alter the Callan–Symanzyk (CS) equation, relevant for the summation. The physical basis on which this is so can be articulated by the following two arguments:

(i) Contributions from the nonperturbative sector do not introduce additional divergences, given that they are finite at short distances.

(ii) Dimension-carrying quantities arising from the nonperturbative sector (correlators) are structured in terms of combinations of renormalization group invariant quantities $gB$; i.e., they behave as external momenta, as opposed to masses which are subject to renormalization.

Consequently, the called-for renormalization group evolution follows the footsteps of the procedure employed in the purely perturbative analysis of the same process [5]. In this connection, it is recalled [2, 26] that the Wilson contour configuration associated with $E_{ii'jj'}$ mixes with one corresponding to a pair of closed loops resulting from an alternative way of identifying the points at infinity [5]. It is associated with

$$\bar{E}_{ij'ji'} = -\frac{\alpha_S}{\pi} c_F(\gamma\coth\gamma - 1)\delta_{ij'}\delta_{ji'} \tag{28}$$
$$\times f(b^2\mu^2) + \frac{\alpha_S}{\pi} c_F(\gamma\coth\gamma - 1 - i\pi\coth\gamma) t^a_{ij'}$$
$$\times t^a_{ji'} i\coth\gamma f(b^2\mu^2) + \mathcal{O}(\alpha_S^2).$$

Accordingly, and upon introducing, in shorthand notation, $W_1 \equiv \delta_{ii'}\delta_{jj'} + E_{ii'jj'}$ and $W_2 \equiv \delta_{ij'}\delta_{ji'} + \bar{E}_{ij'ji'}$, the CS equation assumes the form

$$\left[M\frac{\partial}{\partial M} + \beta(g)\frac{\partial}{\partial g}\right] W_a = -\Gamma_{ab} W_b, \tag{29}$$
$$a, b = 1, 2,$$

with $M$ playing the role of the ultraviolet cutoff whose running takes place between some lower scale (at which corresponding initial conditons are set) and an upper scale set by $1/b$. The anomalous dimension matrix $\Gamma_{ab}$, computed in the context of perturbation theory (see [2, 3, 5]), reads

$$(\Gamma_{ab}) = \frac{\alpha_S}{\pi}\begin{pmatrix} -\dfrac{i\pi}{N_c}\coth\gamma & i\pi\coth\gamma \\ -\gamma\coth\gamma + 1 + i\pi\coth\gamma + 1 & N_c(\coth\gamma - 1) - \dfrac{i\pi}{N_c}\coth\gamma \end{pmatrix} + \mathcal{O}(\alpha_S^2). \quad (30)$$

What does change, with respect to the perturbative analysis, on account of the presence of nonperturbative background contributions is the dependence of the $W_a$ on the $B$-field correlators; i.e., one has $W_a = W_a[\{\Delta^{(n)}\}, M, g]$. Given that the computation has taken into account only the two-point correlator, the extra dependence of the $W_a$ will involve the string tension [cf. Eq. (13)]. The inclusion of this additional dimensional parameter will have its effects on the running coupling constant.

With this in mind, let us recast Eq. (29) in integral form:

$$W_a[\sigma, M_2, g_B(M_2)] \quad (31)$$
$$= \left\{ P\exp\left[ -\int_{M_1}^{M_2} \frac{dM}{M}\Gamma(g_B(M)) \right] \right\}_{ab} \times W_b[\sigma, M_1, g_B(M_1)]$$

with the path ordering becoming necessary because the anomalous dimension matrices do not commute with each other. Concerning the integration limits, a consistent choice, given the premises of the present calculation, is to take $M_2 = 1/b$ and set $M_1 = 1/b_0$ with $b_0^2\sigma < 1$. It is observed that the nonperturbative input enters $W_a$ not only through their explicit dependence on the string constant, but also—which is the most important—through a running coupling constant $g_B$, which obeys the equation

$$M\frac{\partial}{\partial M} g_B(M) = \beta(g_B(M)). \quad (32)$$

The solution of the latter calls for initial conditions which are influenced by the presence of the nonperturbative background and specifically by $\sigma$. Such matters have been studied by Simonov in [25].

Turning our attention to the "deformation" (to the one-loop order) of the running coupling constant, on account of its additional dependence on the background field, we proceed as follows. Knowing the perturbative result to order $\alpha_S$, we go to Eq. (31) and present its solution in the form

$$W_1[\sigma, 1/b, \alpha_B(1/b)] = \delta_{ii'}\delta_{jj'} - i\coth\gamma \quad (33)$$
$$\times \left[ f(b_0^2\mu^2) \int_{1/b_0^2}^{1/b^2} \frac{d\tau}{\tau}\alpha_B(\tau) \right] t^a_{ii'}t^a_{jj'} + \mathcal{O}(\alpha_B^2).$$

It follows that

$$\int_{1/b_0^2}^{1/b^2} \frac{d\tau}{\tau}\alpha_B(\tau) = \alpha_S(1/b^2)f(b^2\mu^2) \quad (34)$$
$$- \alpha_S(1/b_0^2)f(b_0^2\mu^2) + \mathcal{O}(\alpha_S^2).$$

Upon comparing Eqs. (27) and (33), one obtains

$$\alpha_B(\tau) = \alpha_S(\tau)\left\{ 1 + \frac{\mu^2}{4\tau}\left[ \ln\left(\frac{4\tau}{\mu^2}\right) - 2 \right] \right\} + \mathcal{O}(\alpha_S^2). \quad (35)$$

It should be noted that the validity of the above results holds for $\tau/\mu^2 > 1$ and

$$\alpha_S(\tau) = \frac{4\pi}{\beta_0}\frac{1}{\ln(\tau/\Lambda^2)} < 1.$$

An indicative estimate, on the basis of Eq. (35), is that, if $\alpha_S \simeq 0.5$, then $\alpha_B \simeq 0.5(1 + 0.05)$. Following [3, 5], one surmises that the amplitude $A$ for the process under consideration behaves as

$$A \sim \exp\left[ -\frac{N_c}{2\pi}\ln\left(\frac{s}{m^2}\right) \int_{1/b_0^2}^{1/b^2} \frac{d\tau}{\tau}\alpha_B(\tau) + \mathcal{O}(\alpha_S^2) \right] \quad (36)$$
$$\propto \exp\left[ -\frac{\alpha_S}{2\pi}N_c\ln\left(\frac{s}{m^2}\right) f(b^2\mu^2) \right],$$

from which one reads a Reggeized behavior for the gluon. The difference from the usual purely perturbative result is that the function $f(b^2\mu^2)$ is now connected with the modified propagator, as per Eq. (24).

In conclusion, we have demonstrated that the nonperturbative input, through the SVM, to the analysis of a hypothetical quark–quark scattering process in the Regge kinematical region produces a result which, in a phenomenological context, has been argued to be extremely attractive in reproducing low-energy hadron phenomenology. In a sense, this investigation could be considered as a

special example, which justifies Simonov's more general argumentation [21] according to which the perturbative–nonperturbative interference in static QCD interactions at small distances implies the presence of a linear term in the potential.

## ACKNOWLEDGMENTS

We are grateful to our colleague Prof. F.K. Diakonos for programming the computations for the constants $c_i$.

We would also like to acknowledge financial support through the research program "Pythagoras" (grant no. 016) and by the General Secretariat of Research and Technology of the University of Athens.

## *APPENDIX*

Given the set of defining worldline formulas given by Eqs. (11)–(15) in Section 3, we proceed to derive Eq. (16). In the course of the derivation, the various quantities and parameters appearing in the last part of the section are specified.

Writing

$$D(z^2) = \int_0^\infty dp \tilde{D}(p) e^{-pz^2}, \tag{A.1}$$

one determines

$$I(l) = \frac{1}{4\pi^2}\frac{v_1\cdot v_2}{|l|^2}\frac{2N_c^2}{N_c^2-1} v_1\cdot v_2 \tag{A.2}$$
$$\times \int_0^1 d\alpha_2\alpha_2 \int_0^1 d\alpha_1\alpha_1 \int_0^1 dt_2 \int_0^1 dt_1 \theta(t_2-t_1)$$
$$\times \int_0^\infty dp\tilde{D}(p)[48Q(p;t_2,t_1)$$
$$- R(p;t_2,t_1,\alpha_2,\alpha_1)] + \mathcal{O}(\langle g^4F^4\rangle_B),$$

where we have made the change $t_i \to Tt_i$, $i = 1, 2$, and have introduced the quantities

$$Q(p;t_2,t_1) \equiv \left(\frac{\pi}{p}\right)^2 \int_0^\infty dT e^{-T\lambda^2} \tag{A.3}$$
$$\times \int \frac{d^4q}{(2\pi)^4} e^{-q^2/(4p)} \int_{\substack{x(0)=0\\ x(1)=l}} \mathcal{D}x(t)$$
$$\times \exp\left[-\frac{1}{4T}\int_0^1 dt\dot{x}^2(t)\right] \exp[iq\cdot(x(t_2)-x(t_1))]$$

and

$$R(p;t_2,t_1,\alpha_2,\alpha_1) \equiv \left(\frac{\pi}{p}\right)^2 \int_0^\infty dT e^{-T\lambda^2} \tag{A.4}$$
$$\times \int \frac{d^4q}{(2\pi)^4} e^{-q^2/(4p)} (\delta_{\mu_2\nu_2}\delta_{\mu_1\nu_1} - \delta_{\mu_2\nu_1}\delta_{\mu_1\nu_2})$$
$$\times \int_{\substack{x(0)=0\\ x(T)=l}} \mathcal{D}x(t) \exp\left[-\frac{1}{4}\int_0^T dt\dot{x}^2(t)\right] \dot{x}_{\mu_2}(t_2)$$
$$\times x_{\nu_2}(t_2)\dot{x}_{\mu_1}(t_1)x_{\nu_1}(t_1)$$
$$\times \exp[iq\cdot[\alpha_2 x(t_2) - \alpha_1 x(t_1)]]$$

with $|l|$, as defined in the text.

The above path integrals can be executed by employing standard techniques, given that "particle" action functionals are quadradic (plus a linear term) [16–18]. Ignoring terms giving contributions $\mathcal{O}(b^2)$ and using condensed notation from hereon, one determines

$$Q = \frac{1}{16}\frac{1}{p^2}\int_0^\infty dT e^{-T\lambda^2} \int \frac{d^4q}{(2\pi)^4} \tag{A.5}$$
$$\times \exp\left[-\frac{q^2}{4p} - Tq^2G_{12}\right][1 + \mathcal{O}(l^2q^2)]$$

and

$$R = \frac{1}{16}\frac{1}{p^2}\int_0^\infty dT e^{-T\lambda^2} \int \frac{d^4q}{(2\pi)^4} \tag{A.6}$$
$$\times \exp\left[-\frac{q^2}{4p} - Tq^2K_{12}\right]$$
$$\times [c_0 + Tq^2c_1 + T^2q^4c_2 + \mathcal{O}(l^2q^2)],$$

where the following one-dimensional particle-propagator-type quantities have been introduced:

$$\Delta_{12} = \Delta(t_1,t_2) \equiv t_1(1-t_2)\theta(t_2-t_1) \tag{A.7}$$
$$+ t_2(1-t_1)\theta(t_1-t_2),$$

$$G_{12} = G(t_2,t_1) = \Delta(t_2,t_2) + \Delta(t_1,t_1) \tag{A.8}$$
$$- 2\Delta(t_1,t_2) = |t_2-t_1|(1-|t_2-t_1|)$$

and

$$K_{12} = K(t_2,t_1) = \alpha_2^2\Delta(t_2,t_2) \tag{A.9}$$
$$+ \alpha_1^2\Delta(t_1,t_1) - 2\alpha_1\alpha_2\Delta(t_1,t_2).$$

The coefficients entering Eq. (A.6) are given by the expressions

$$c_0 = -72\Delta_{12}\frac{\partial}{\partial t_1}\Delta_{12}\frac{\partial}{\partial t_2}\Delta_{12}, \tag{A.10}$$

$$c_1 = 48\Delta_{12}\partial_2\Delta_{12}\partial_1 K_{12} \qquad (A.11)$$
$$+ 24\alpha_1\partial_1\Delta_{12}\partial_2\Delta_{12}(\alpha_1\Delta_{11} - \alpha_2\Delta_{12})$$
$$+ 24\alpha_2\partial_1\Delta_{12}\partial_2\Delta_{12}(\alpha_2\Delta_{22} - \alpha_1\Delta_{12})$$
$$- 12(\alpha_1^2\Delta_{12}\partial_1\Delta_{11}\partial_2\Delta_{12} + \alpha_2^2\Delta_{12}\partial_2\Delta_{22}\partial_1\Delta_{12}$$
$$- \alpha_1\alpha_2\Delta_{12}\partial_1\Delta_{11}\partial_2\Delta_{22} - \alpha_1\alpha_2\Delta_{12}\partial_1\Delta_{12}\partial_2\Delta_{12})$$
$$- 12(\alpha_2\Delta_{22} - \alpha_1\Delta_{12})(\alpha_1\partial_1\Delta_{11}\partial_2\Delta_{12}$$
$$- \alpha_2\partial_1\Delta_{12}\partial_2\Delta_{12}) - 12(\alpha_1\Delta_{11} - \alpha_2\Delta_{12})$$
$$\times (\alpha_2\partial_2\Delta_{22}\partial_1\Delta_{12} - \alpha_1\partial_1\Delta_{12}\partial_2\Delta_{12})$$

and

$$c_2 = 24(\alpha_2\Delta_{22} - \alpha_1\Delta_{12}) \qquad (A.12)$$
$$\times (\alpha_1\Delta_{11} - \alpha_2\Delta_{12})\partial_1 K_{12}\partial_2\Delta_{12}.$$

Given the above, the "paramagnetic" contribution to Eq. (A.2) becomes

$$I_p = 12\frac{2N_c}{N_c^2-1}v_1\cdot v_2\int\limits_0^\infty dp\tilde{D}(p)\int\limits_0^1 dt_2 \qquad (A.13)$$
$$\times\int\limits_0^1 dt_1\theta(t_2-t_1)Q(p;t_2,t_1)$$
$$= \frac{12}{16}\frac{2N_c^2}{N_c-1}\frac{v_1\cdot v_2}{4\pi^2}\int\limits_0^\infty\frac{dp}{p}\tilde{D}(p)$$
$$\times\left[\ln\left(4e^{-\gamma_E}\frac{p}{\lambda^2}\right)+\mathcal{O}(\lambda^2/p)\right],$$

and since

$$\int\limits_0^\infty\frac{dp}{p}\tilde{D}(p) = \int\limits_0^\infty dz^2\tilde{D}(z^2) \qquad (A.14)$$
$$= \frac{1}{\pi}\int\limits_0^\infty d^2z\tilde{D}(z^2) \equiv \frac{2}{\pi}\sigma,$$

Eq. (A.13) gives

$$I_p = \frac{3N_c}{N_c^2-1}\frac{v_1\cdot v_2}{4\pi^2}\sigma\ln\left(C\frac{\sigma}{\lambda^2}\right). \qquad (A.15)$$

This furnishes the correction term from the background gauge field contributions entering Eq. (14) in the text. The constant $C$ entering the above result depends on the parametrization of $D(z^2)$. Following the one of Nachtmann [7], one determines $C = 39.65$. The numerical computation of the factor $\kappa$, based on the expressions for the $c_i$, as given by Eqs. (A.10)–(A.12), produces the value $\kappa \simeq 0.5$.

## REFERENCES


1. H. Cheng and T. T. Wu, *Expanding Protons: Scattering at High Energies* (M. I. T. Press, Cambridge, 1987).
2. L. N. Lipatov, Nucl. Phys. B **309**, 379 (1988).
3. G. P. Korchemsky, Phys. Lett. B **325**, 459 (1994); I. A. Korchemskaya and G. P. Korchemsky, Nucl. Phys. B **437**, 127 (1995).
4. D. Kabat, Comments Nucl. Part. Phys. **20**, 325 (1992).
5. A. I. Karanikas and C. N. Ktorides, Phys. Rev. D **59**, 016003 (1999).
6. H. G. Dosch and Yu. A. Simonov, Phys. Lett. B **205**, 339 (1988).
7. O. Nachtmann, *Lectures Given at the 35th International University School of Nuclear and Particle Physics "Perturbative and Nonperturbative Aspects of Quantum Field Theory", Schladming, Austria, 1996*, HD-THEO-96-38.
8. G. K. Savvidi, Phys. Lett. B **71B**, 133 (1977).
9. N. G. Van Kampen, Physica **74**, 215, 239 (1974); Phys. Rep. **24**, 172 (1976).
10. A. Di Giacomo, H. G. Dosch, V. I. Shevchenko, and Yu. A. Simonov, Phys. Rep. **372** (2002).
11. Yu. A. Simonov, Lect. Notes Phys. **479**, 144 (1995); Yad. Fiz. **65**, 140 (2002) [Phys. At. Nucl. **65**, 135 (2002)].
12. G. 't Hooft, *The Background Field Method in Gauge Field Theories, in "Karpacz 1975, Proceedings"* (Acta Universitatis Wratislaviensis, Wroclaw, 1976), No. 368, Vol. 1, p. 345.
13. V. A. Fock, Izv. Akad. Nauk USSR, OMEN, 557 (1937).
14. R. P. Feynman, Phys. Rev. **80**, 440 (1950).
15. J. Schwinger, Phys. Rev. **82**, 664 (1951).
16. M. J. Strassler, Nucl. Phys. B **385**, 145 (1992).
17. M. Reuter, M. G. Schmidt, and C. Schubert, Ann. Phys. (N.Y.) **259**, 313 (1997).
18. A. I. Karanikas and C. N. Ktorides, J. High Energy Phys. **9911**, 033 (1999).
19. Yu. A Simonov and J. A. Tjon, Ann. Phys. (N.Y.) **300**, 54 (2002).
20. K. G. Chetyrkin, S. Narison, and V. I. Zakharov, Nucl. Phys. B **550**, 353 (1999).
21. Yu. A. Simonov, Phys. Rep. **320**, 265 (1999).
22. S. L. Huber, M. Reuter, and M. G. Schmidt, Phys. Lett. B **462**, 158 (1999).
23. V. A. Fock, Phys. Z. Sowjetunion **12**, 404 (1937).
24. J. Schwinger, Phys. Rev. **82**, 664 (1951).
25. Yu. A. Simonov, Usp. Fiz. Nauk **166**, 337 (1996) [Phys. Usp. **39**, 313 (1996)].
26. R. A. Brandt, F. Neri, and M.-A. Sato, Phys. Rev. D **24**, 879 (1981); R. A. Brandt, A. Gocksch, M.-A. Sato, and F. Neri, Phys. Rev. D **26**, 3611 (1982).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 870–889. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 904–924.*


# Instantons or Monopoles: Dyons at Finite Temperature*

**E.-M. Ilgenfritz[1], B. V. Martemyanov[2]**, M. Müller-Preussker[1], and A. I. Veselov[2]**

Received May 21, 2004; in final form, September 22, 2004

**Abstract**—This article in honor of Yurii Antonovich Simonov's 70th birthday reviews some recent work related to the semiclassical approach to QCD at finite temperature based on classical solutions with nontrivial holonomy. By cooling Monte Carlo generated lattice $SU(2)$ gauge fields, we investigate approximate solutions of the classical field equations as a possible starting point for rethinking the semiclassical approximation of the path integral. We show that old findings of cooling have to be reinterpreted in terms of Kraan–van Baal solutions with generically nontrivial holonomy instead of Harrington–Shepard caloron solutions with trivial holonomy. The latter represent only a subclass of possible topological configurations and for $T < T_c$ seem to become suppressed due to quantum fluctuations. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

It is a great honor for us to be invited to give a tribute to Yurii Antonovich Simonov on the occasion of his 70th birthday. All of us, in one or another context, had the opportunity to work together with him and gratefully remember this. More than that, his ideas had an important impact on our thinking on the vacuum structure of gluodynamics and QCD.

For long time, the vacuum was imagined as a random medium (dilute gas or less dilute liquid) of instantons [1–5] described by a density of about 1 fm$^{-4}$ and a radius $\rho = 0.3$–$0.5$ fm, and this picture was easily extended to finite temperature with the instantons being replaced by Harrington–Shepard (HS) calorons [6, 7] and where the temperature acted to regulate the infrared explosion with growing $\rho$ (for recent reviews, see [8, 9]). The starting point for the $T = 0$ model was the assumption of a repulsion between instantons at distance $d$ ($d^2 > \text{const} \cdot \rho_1\rho_2$) [4] which seemed to solve many puzzles at once. This idea then has opened the way to the instanton liquid picture of the vacuum [5]. This picture was rather successful for doing hadron phenomenology [8], i.e., for understanding the behavior of light quarks.

The importance of the existence of instantons for the understanding of gluodynamics, however, is still under debate, and this makes us uncomfortable about the status of instantons. The confining aspect of the vacuum was mainly ignored by the physicists working with the instanton model, while instantons were mainly ignored by the confinement community, which concentrated instead on monopoles and vortices [10, 11] (in the context of projecting Yang–Mills theory to $U(1)^{N-1}$ and $Z(N)$ Abelian gauge theories, correspondingly). There have been claims for some time [12] that random instantons could create a string tension of the right magnitude. Finally, among other not realistic features of the model pointed out, it became clear that one of the main counterarguments raised by Yurii Antonovich against the model [13], the violation of approximate Casimir scaling in the instanton model, cannot be circumvented [14].

Yurii Antonovich, who has designed a model-independent scheme for dealing with confinement, spectroscopy, and other hadronic features in terms of $n$-point field strength cumulants [15], with only the two-point correlator getting involved in the leading (so-called Gaussian) approximation, from this point of view has always been skeptical that instantons alone could give a description of all these aspects. Instead, he has pointed out how instantons would be modified on top of other, more infrared ("confining background"), fields without losing the role that they have to play for chiral symmetry breaking.

This different starting point did not preclude that we together once estimated the strength and correlation length of the field strength correlator in the dilute instanton gas model [16], surprisingly with parameters in the above-mentioned ballpark. In a follow-up paper to this work, we have found, however, that the field strength correlator would not be defined correctly without taking the interplay of instantons (instanton correlations) into account [17].

During the 1990s, there were numerous attempts to identify instanton-like structures, the density and

*This article was submitted by the authors in English.

[1)]Institut für Physik, Humboldt-Universität zu Berlin, Germany.

[2)]Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia.

**E-mail: martemja@itep.ru

size (or, better, size distribution) of instantons, from ab initio lattice gauge field samples. This task turned out to be much more difficult than expected. Apart from the very last time, all these studies relied on very subjective tools of smoothing the UV gauge field fluctuations (cooling [18–20], blocking and inverse blocking [21–23], APE smearing [24]). Whereas the existence of hot spots (very localized regions of large field strength, where it turns out self-dual or anti-self-dual) immediately caught the eye, the number of these objects and the sizes were strongly dependent on details and prejudices. Only recently, thanks to the mastering of chirally improved [25] or chirally perfect (so-called overlap) Dirac operators [26], the hope emerged to reach the goal of an objective identification of lumps of topological charge density. But still, the correct interpretation of the lumps is a task which poses new problems.

Yurii Antonovich, in the mid-1990s, proposed his own model of the vacuum structure which was thought to replace the instanton model. One could say that it started from a certain limit of the HS caloron which itself can be understood as a chain of instantons, enforcing the Euclidean (thermal) periodicity $b = 1/T$, looked at for time $t$ in the interval $0 < t < b$. In the formal limit $\rho \to \infty$, one gets a self-dual or anti-self-dual configuration (independent of $b$) which has a magnetic and an electric charge. Some of us participated in studies together with Yurii Antonovich, which were devoted to the properties of such a dyon gas. Being essentially extended particles, dyons are characterized by parameters like mass, the direction of their world line in Euclidean space, and a color orientation inherited from the HS caloron [27].

All these degrees of freedom should be averaged over in the full dyon gas. It was realized that dyons can be superposed in the singular gauge such that the topological charge is additive, the residual interaction between dyons was calculated [28], and a mean-field approximation was developed [29]. The connection to finite-temperature Yang–Mills theory was understood but not used. This model had an unwanted feature [30]: more-than-linear confinement. The idea of screening was put forward [31] in order to cure this "superconfinement." To make dyons the essential constituents of topological density, some degree of freedom must have been missing in the model.

Coming back to the attempts to identify the topological structure of a lattice field ensemble, even the most conservative method called "cycling" [22], a series of blocking and inverse blocking steps, which in principle has the potential to unambiguously return a collection of instantons and anti-instantons [32] thought to encode the infrared structure of each of the ab initio gauge fields, seems to ignore some important information. Neither the corresponding monopole structure of the true configurations [33] nor the mesonic two-point functions [34] could be reproduced on the reconstructed configurations based on the (anti-)instanton sizes and loci alone.

There were two ideas about what could be missing. One of them focused on the eventual role of field strength correlations between the hot spots which were not recordable by the cycling method. This could have been measured by Yurii Antonovich's gauge-invariant field strength correlator, restricted to the loci of instantons. However, smoothing and other methods ("restricted overimproved cooling") have failed [35, 36] to detect these correlations. Thus, it is fair to say that the uncorrelated instanton model correctly represents the result of cooling, provided the cooling is stopped at the right (average) multiplicity of hot spots. The other idea was that a nontrivial holonomy in one of the four Euclidean directions might have escaped attention and therefore a corresponding constant background $A_\mu$ should be recorded and used to model the lattice field [34]. This construction was helpful for reconstruction of the mesonic correlators just empirically. At this stage, realizing a possible correlation between holonomy and the topological lump structure was out of scope, while the message was clear: holonomy and topological structure are working together. Whereas in an "infinite" volume a well-defined holonomy in one of the spacelike directions might be an artifact of the actually finite spatial box, the temporal holonomy (i.e., the set of eigenvalues of the Polyakov loop) is an integral part of the physics at finite temperature. This remark includes also the confinement–deconfinement transition.

An important hint how to proceed further came with the discovery of (anti-)self-dual caloron solutions with nontrivial holonomy by Lee and Lu [37] and Kraan and van Baal [38, 39]. The works by Kraan and van Baal (KvB) [38, 39] clarified the special role of the HS caloron as a finite-$T$ instanton with trivial holonomy and the special character of the dyon which had been considered in the dyon gas model. The generic case is now that the caloron can consist of $N_c$ dyons sharing the action (and topological charge) of the whole caloron according to the eigenvalues of the asymptotic holonomy. For trivial holonomy, only one of the dyons "eats" the whole action (and topological charge). If the now massless constituent dyons are moved to infinity, the remaining one turns into the dyon considered in the dyon gas model mentioned above. The long-range properties of the dyonic caloron constituents are similar to those dyons, except for the fact that electric and magnetic charge (and topological charge) fluctuate in response to the background holonomy. For suitable constituents approaching each other in configurations of

higher topological charge (even more so in generically non(anti-)self-dual fields), various nonlinear superpositions will be realized. Constituents of one caloron approaching each other lead to configurations which are also not static and resemble more instantons despite the nontrivial holonomy. This is typically what happens if the inverse temperature becomes big compared to the distance between calorons.

Very recently, Diakonov [9] discussed the possible consequences of the KvB solutions as background fields in the semiclassical approximation of the Yang–Mills path integral. Together with Petrov and young coworkers [40], he went through the difficult task of computing the one-caloron functional determinant as a prerequisite for establishing a KvB caloron gas or liquid approximation. The hope is that the path integral could finally be represented in terms of the dyon constituents instead of the full calorons.

In the present review, we will describe the results of investigations of calorons with nontrivial holonomy and of constituent dyons obtained by cooling of Monte Carlo lattice gauge field configurations [41–47]. While the new type of classical solutions with an action $S = 8\pi^2/g_0^2$ does remember the physical (equilibrium) situation only through the nontrivial holonomy it inherited from the confined phase, the dependence on inverse temperature $b$ and volume has important consequences for the rebuilding of the caloron model, in particular, for its working near the confinement–deconfinement phase transition. Higher action plateaus reflect more about the structure of the equilibrium configurations, but this is not yet completely understood. It seems more important to study the Polyakov loop structure of lumps of action which become visible after APE smearing [24]. Another way to reveal the dissociation into dyon constituents is provided by the specific localization behavior of zero modes of the Dirac operator depending on the boundary conditions imposed on the fermion fields [48, 49]. Gattringer and his coworkers [50] have recently made a first attempt to identify such dyon constituents even in equilibrium lattice fields with the help of the zero modes of a chirally improved lattice Dirac operator.

This paper, in which we shall restrict ourselves to the case of $SU(2)$ lattice gauge fields, is organized as follows. In Section 2, we will give a theoretical introduction to the objects under consideration: instantons, calorons, calorons with nontrivial holonomy, etc. In Section 3, we shall briefly discuss the recent work by Diakonov and coworkers [40]. Section 4 provides all necessary lattice definitions, in particular, the observables considered in order to identify calorons with nontrivial holonomy. Then, in Section 5, at lowest action plateaus for temperatures near but mostly below the deconfining temperature $T_c$, we describe the search for (approximate) solutions of the lattice equation of motion. We shall show that the nontrivial holonomy plays an important role and shall see solutions carrying all attributes of KvB solutions. But we shall also discuss some findings related to by-products ("artifacts") of the cooling process (for instance, dyon–antidyon pairs, left over from partial caloron–anticaloron annihilations), as well as of the finiteness of the volume with torus topology (Dirac sheets). In Section 6, we demonstrate under what circumstances KvB solutions become dissociated into dyon pairs or recombine into calorons, the latter still having the internal nontrivial holonomy substructure. We shall see that, even on a symmetric torus, instantons with nontrivial holonomy show up, for which the analytic form is not yet known. Section 7 presents the first results from ensembles obtained at higher action plateaus. Section 8 presents the conclusion and possible directions of future investigations.

## 2. CALORONS WITH NONTRIVIAL HOLONOMY IN CONTINUUM $SU(2)$ YANG–MILLS THEORY

In the following, we consider the case of $SU(2)$ pure gauge theory in continuum 4D Euclidean space. We assume periodicity of the gauge fields in the fourth, i.e., imaginary time, direction, as it is required for the path-integral representation of the partition function in the nonzero temperature case. Classical instanton solutions periodic in $x_4$ are therefore called calorons [6] and have been constructed generalizing the one-instanton solution [1] in the singular gauge with the help of the so-called 't Hooft ansatz [51]

$$A_\mu(x) = \frac{1}{g_0}\frac{\tau_a}{2}\bar\eta^a_{\mu\nu}\partial_\nu \log\phi(x), \qquad (1)$$
$$\mu,\nu = 1,\ldots,4,$$

with an infinite chain of equidistant singularities

$$\phi(x) = 1 + \sum_{n=-\infty}^{+\infty} \frac{\rho^2}{(\mathbf{x}-\mathbf{x}_0)^2 + (t-nb)^2} \qquad (2)$$
$$= 1 + \frac{\pi\rho^2}{br}\frac{\sinh(2\pi r/b)}{\cosh(2\pi r/b) - \cos(2\pi t/b)}.$$

Here, $\bar\eta^a_{\mu\nu}$ denotes the anti-self-dual 't Hooft tensor; $\tau_a/2$, $a = 1,2,3$, are the generators of the $SU(2)$ group; $r = |\mathbf{x}-\mathbf{x}_0|$ and $t = x_4 - x_4^0$ determine the spatial and temporal distance to the caloron center, whereas $b$ gives the time period to be identified with the inverse temperature $T^{-1}$; for simplicity, mostly $b = 1$ is assumed in what follows; $\rho$ denotes the scale size of the solution.

The action and the topological charge of the caloron solution—integrated over one periodicity

strip of width $b$—correspond to the one-instanton values $S = S_{\rm inst} \equiv 8\pi^2/g_0^2$ and $Q_t = 1$, respectively. The anti-instanton solution with the opposite charge $Q_t = -1$ is easily obtained by replacing the 't Hooft tensor $\bar\eta^a_{\mu\nu}$ by the self-dual one $\eta^a_{\mu\nu}$ [2].

The gauge potentials of the (anti)caloron solutions fall off at spatial infinity, providing a trivial asymptotic holonomy, i.e.,

$$\mathcal{P}(\mathbf{x}) = P\exp\left(i\int\limits_0^b A_4(\mathbf{x},t)dt\right) \to \mathcal{P}_\infty = \pm\mathbf{1} \in \mathbb{Z}_2 \quad \text{for} \quad |\mathbf{x}| \to \infty, \tag{3}$$

whereas

$$\mathcal{P}(\mathbf{x}) = P\exp\left(i\int\limits_0^b A_4(\mathbf{x},t)dt\right) \to \mp\mathbf{1} \quad \text{for} \quad \mathbf{x} \to \mathbf{x}_0 \tag{4}$$

holds complementary at the center of the (anti)caloron.

The HS chain of instantons is aligned along the temporal direction with an identical orientation in color space. In the large-scale limit, $\rho \to \infty$, and after applying an appropriate gauge transformation, it turns into a static solution which can be identified with a BPS monopole [52] in Euclidean space, where the fourth component $A_4$ plays the role of the Higgs field in the adjoint representation. The solution has both electric and magnetic charge. This motivates its name "dyon" ($D$).

In order to generalize the solution to the case of nontrivial holonomy, it is useful to rewrite the HS caloron (1), (2) in terms of separate "isospin" components

$$A_\mu(x) = \frac{1}{2}\bar\eta^3_{\mu\nu}\tau_3\partial_\nu \log\phi(x) + \frac{1}{2}\phi(x)\mathrm{Re}\left((\bar\eta^1_{\mu\nu} - i\bar\eta^2_{\mu\nu})(\tau_1 + i\tau_2)\partial_\nu\chi(x)\right) \tag{5}$$

with ($b = 1$)

$$\phi(x) = \psi(x)/\hat\psi(x), \tag{6}$$
$$\psi(x) = \cosh(2\pi r) - \cos(2\pi t) + \frac{\pi\rho^2}{r}\sinh(2\pi r),$$
$$\hat\psi(x) = \cosh(2\pi r) - \cos(2\pi t),$$
$$\chi(x) = 1 - \frac{1}{\phi} = \frac{\pi\rho^2\sinh(2\pi r)}{\psi r}.$$

A caloron with nontrivial holonomy [38, 39] is a solution with two spatial centers. It can be viewed—in the limit ($\rho/b \ll 1$)—as an instanton chain, where each of the instantons is rotated compared to the previous one by an angle $4\pi\omega$ in color space, $\omega$ being the parameter of holonomy, as we will see below. The rotation axis is arbitrary; for definiteness, the third axis is taken. The solution is made periodic by a nonperiodic (in time) gauge transformation $g(x) = e^{-2\pi i t\omega\tau_3}$. The explicit expression of the KvB caloron solution looks as follows:

$$A^{\rm per}_\mu = \frac{1}{2}\bar\eta^3_{\mu\nu}\tau_3\partial_\nu\log\phi + \frac{1}{2}\phi\mathrm{Re}((\bar\eta^1_{\mu\nu} - i\bar\eta^2_{\mu\nu})(\tau_1 + i\tau_2)(\partial_\nu + 4\pi i\omega\delta_{\nu,4})\tilde\chi) + \delta_{\mu,4}\cdot 2\pi\omega\tau_3, \tag{7}$$

where

$$\phi(x) = \psi(x)/\hat\psi(x), \tag{8}$$
$$\psi(x) = -\cos(2\pi t) + \cosh(4\pi r\bar\omega)\cosh(4\pi s\omega) + \frac{r^2+s^2+\pi^2\rho^4}{2rs}\sinh(4\pi r\bar\omega)\sinh(4\pi s\omega) + \pi\rho^2(s^{-1}\sinh(4\pi s\omega)\cosh(4\pi r\bar\omega) + r^{-1}\sinh(4\pi r\bar\omega)\cosh(4\pi s\omega)),$$
$$\hat\psi(x) = -\cos(2\pi t) + \cosh(4\pi r\bar\omega)\cosh(4\pi s\omega) + \frac{r^2+s^2-\pi^2\rho^4}{2rs}\sinh(4\pi r\bar\omega)\sinh(4\pi s\omega),$$
$$\tilde\chi = \frac{\pi\rho^2}{\psi}\left\{e^{-2\pi i t}s^{-1}\sinh(4\pi s\omega) + r^{-1}\sinh(4\pi r\bar\omega)\right\}$$

instead of Eq. (6). The holonomy parameters $\omega$ and $\bar\omega$ are related to each other: $\bar\omega = 1/2 - \omega, 0 \le \omega \le 1/2$; $r = |\mathbf{x} - \mathbf{x}_1|$ and $s = |\mathbf{x} - \mathbf{x}_2|$ are the 3D distances to the locations of the two centers of the new caloron solution. The distance between the centers $d \equiv |\mathbf{x}_1 - \mathbf{x}_2|$ is connected with the scale size and the width of the time periodicity strip through

$$\pi\rho^2/b = d. \tag{9}$$

The time component of the caloron potential becomes nonzero at spatial infinity, providing the nontrivial asymptotic holonomy

$$\mathcal{P}(\mathbf{x}) = P\exp\left(i\int\limits_0^b A_4(\mathbf{x},t)dt\right) \to \mathcal{P}_\infty = e^{2\pi i\omega\tau_3} \quad \text{for} \quad |\mathbf{x}| \to \infty, \tag{10}$$

whereas

$$\mathcal{P}(\mathbf{x}) = P\exp\left(i\int\limits_0^b A_4(\mathbf{x},t)dt\right) \to \pm\mathbf{1} \quad \text{for} \quad \mathbf{x} \to \mathbf{x}_{1,2} \tag{11}$$

holds simultaneously at the two spatial centers $\mathbf{x}_1, \mathbf{x}_2$ of the KvB caloron [38, 53].

In terms of $\omega$, the normalized trace of the holonomy, the Polyakov loop, which we shall take as a direct measure of the holonomy, at spatial infinity becomes

$$L(\mathbf{x}) \equiv \frac{1}{2}\mathrm{tr}\mathcal{P}(\mathbf{x}) \to L_\infty \equiv \frac{1}{2}\mathrm{tr}\mathcal{P}_\infty = \cos(2\pi\omega), \tag{12}$$

whereas it shows a dipole structure with opposite peak and dip values $\pm 1$ at the centers of the solution.

From Eqs. (8), it can be easily seen that, for trivial asymptotic holonomy, i.e., for $\omega \to 0$ or $\bar{\omega} \to 0$, and by sending one of the positions of the centers to infinity, the "old" HS caloron emerges as described by Eqs. (5), (6).

On the other hand, for nontrivial asymptotic holonomy, we can distinguish two limiting cases. First, when the separation $d$ between the centers becomes sufficiently large, two well-separated constituents emerge which are static in time. The "mass" ratio of these dissociated constituents is equal to $\bar{\omega}/\omega$. Since the full solution is self-dual, the ratio is the same for the action as for the equal-sign topological charge carried by the constituents, the latter summing up to one unit of topological charge, $Q_t = 1$. The separated constituents form a pair of BPS monopoles (or dyons) with opposite magnetic charges. In the following, we will call this limiting case a $DD$ pair. Second, in the small-distance limit, we shall observe just one lump of action and integer topological charge which looks very similar to the HS caloron. Thus, we denote this nondissociated KvB solution by caloron. Still, this one-lump solution contains the typical internal dipole structure for the local Polyakov loop variable as mentioned above.

The action density in all three cases described above can be expressed by a simple formula [38]

$$s(x) = -\frac{1}{2}\partial_\mu^2\partial_\nu^2 \log\psi(x). \tag{13}$$

For well-separated dyons, when the functions $\phi(x)$ and $\psi(x)$ are almost time-independent, the strongest time dependence comes from the first part of the function $\tilde{\chi}(x)$. This dependence is represented by the phase $e^{-2\pi i t}$ and is nothing else but the homogeneous rotation of the first dyon, which has $L(\mathbf{x}_1) = -1$, in color space around the third axis with angle $2\pi$ over the period. The second dyon with $L(\mathbf{x}_2) = +1$ is static. Such a relative rotation of two dyons (which form a monopole–antimonopole pair) gives the so-called Taubes winding necessary to produce a unit topological charge from a monopole–antimonopole pair [54]. One can detect this rotation in a gauge-invariant fashion by looking at the gauge-invariant field strength correlator defined on each constant-time slice and watching its evolution over the periodicity interval $b$ [44].

Finally, let us comment on the zero-mode eigenfunctions of the fermionic massless Dirac operator in the background of the KvB solutions. They have been studied analytically in [48] and [49]. One finds closed solutions depending on the type of (anti)periodic boundary conditions (b.c.) imposed on the fermion fields in the imaginary time direction. In the case of well-separated dyon pairs, i.e., for $d = \pi\rho^2/b \gg 1$, the zero-eigenmode densities become very simple expressions,

$$|\psi^-(x)|^2 = -\frac{1}{4\pi}\partial_\mu^2\left[\tanh(2\pi r\bar{\omega})/r\right] \tag{14}$$

for antiperiodic b.c.,

$$|\psi^+(x)|^2 = -\frac{1}{4\pi}\partial_\mu^2\left[\tanh(2\pi s\omega)/s\right]$$

for periodic b.c.

This means that the zero-mode eigenfunctions are always localized around one of the constituents of the KvB solution, for antiperiodic b.c. at that constituent which has $L(\mathbf{x}_1) = -1$ at its center $\mathbf{x}_1$. Under the switching to periodic b.c. for the fermion fields, the zero-mode localization jumps to the other constituent monopole of the gauge field. Therefore, the fermionic zero modes provide a convenient way to identify a monopole-pair structure in the gauge fields.

## 3. A SEMICLASSICAL APPROACH TO CONFINEMENT BASED ON THE KvB SOLUTIONS

For two decades, the only known generalization of the instanton solution to finite temperature was the HS caloron [6]. In [7], alternatives were discussed but soon abandoned. The argument against solutions with nontrivial asymptotic holonomy was the following. The one-loop effective action obtained by integrating out short-wavelength fluctuations on top of constant $A_4$ (in the absence of a nonvanishing electric or magnetic field) was known to be [55]

$$S_{\mathrm{eff}}(A_4) = \int_0^{1/T} dt \int d^3x P(\phi) \tag{15}$$

with $\phi = \sqrt{A_4^a A_4^a} + 1$ and

$$P(\phi) = \frac{1}{3T(2\pi)^2}\phi^2\bar{\phi}^2. \tag{16}$$

Here, $\phi$ is related to the holonomy parameter $\omega$ introduced in the previous section by $L = \cos(2\pi\omega) = \cos(\phi/(2T))$. The notation $\bar{\phi}$ is shorthand to denote $2\pi T - \phi$. For a single dyon or a caloron with asymptotically nontrivial holonomy $\omega$, the consequence

would be a positive free-energy density $f_{\rm pert} = P(\phi)$ of the background, compared to the cases of trivial holonomy, $\omega = 0$ or $\omega = 1/2$. For the topological object under discussion, embedded in an infinite volume, this would mean that its contribution to the partition function is infinitely suppressed.

This argument was questioned first by Diakonov in [9]. He pointed out that, independent of details of the decomposition of the caloron amplitude into dyon amplitudes, the total free-energy density of a dyon gas would eventually reverse the argument, i.e., overcompensate the positive free-energy density in the region of nontrivial holonomy. A sketchy way to construct such a mechanism, in the case of $SU(2)$ gluodynamics, goes as follows [9]. Realize that the asymptotic holonomy finally determines how the action (and topological charge) of the caloron is distributed among the two dyonic constituents. Let us label them $M$ and $L$, carrying magnetic and electric charges $M = (+,+)$ and $L = (-,-)$. They are sources of self-dual fields. Correspondingly, an anticaloron is composed of $\bar{M} = (+,-)$ and $\bar{L} = (-,+)$, which are sources of anti-self-dual fields. The respective share of the total action $S = S_{\rm inst}$ is $2\omega S_{\rm inst}$ and $(1-2\omega)S_{\rm inst}$. Equal sharing is possible only in the background of maximally nontrivial holonomy, $\omega = 1/4$. Assuming the factorization of the caloron amplitude into that of a dyon pair, the partition function of the dyon and antidyon gases would be

$$Z_{\rm dyon} = \sum_{N_M,N_{\bar{M}},N_L,N_{\bar{L}}} \frac{1}{N_M!N_{\bar{M}}!N_L!N_{\bar{L}}!} \qquad (17)$$
$$\times \left[\int d^3z|\phi|^3 \exp\left(-\frac{8\pi^2}{g^2}\frac{|\phi|}{2\pi T}\right)\right]^{N_M+N_{\bar{M}}}$$
$$\times \left[\int d^3z|2\pi T-\phi|^3 \right.$$
$$\left. \times \exp\left(-\frac{8\pi^2}{g^2}\frac{|2\pi T-\phi|}{2\pi T}\right)\right]^{N_L+N_{\bar{L}}}.$$

Then the dyon pressure

$$T\frac{\log Z_{\rm dyon}}{V} = 2|\phi|^3 \exp\left(-\frac{8\pi^2}{g^2}\frac{|\phi|}{2\pi T}\right) \qquad (18)$$
$$+ 2|2\pi T-\phi|^3 \exp\left(-\frac{8\pi^2}{g^2}\frac{|2\pi T-\phi|}{2\pi T}\right)$$

would give a negative contribution to the free-energy density. Invoking a running coupling depending on temperature $T$ in the usual way would lead to a dyon free-energy density

$$f_{\rm dyon} = -cT\left[|\phi|^3\left(\frac{\Lambda}{\pi T}\right)^{\frac{22}{3}\frac{|\phi|}{2\pi T}}\right. \qquad (19)$$
$$\left. + |2\pi T-\phi|^3\left(\frac{\Lambda}{\pi T}\right)^{\frac{22}{3}\frac{|2\pi T-\phi|}{2\pi T}}\right],$$

which becomes minimal at maximally nontrivial holonomy. For sufficiently low temperature, it could be strong enough to overcompensate the positive perturbative free-energy density $f_{\rm pert}$.

Very recently, the $SU(2)$-caloron amplitude was evaluated by Diakonov and coworkers [40]. Based on this knowledge, the conditions have been clarified under which the dissociation into dyonic constituents actually occurs. This is a necessary prerequisite for the free-energy density eventually to acquire a form resembling (19). In [40], the temperature $T_c$ was estimated at which trivial holonomy becomes unstable.

The fluctuation determinants have been evaluated by integrating the derivatives with respect to $\phi$, beginning from $\phi = 0$ (trivial caloron). The prefactor has been fixed also by connecting the result to the limiting case of the trivial caloron dealt with in [7]. A closed expression has been obtained for the single-caloron amplitude written in terms of the dyon coordinates $\mathbf{z}_1$ and $\mathbf{z}_2$ in the limit of dyon separation $r_{12} = |\mathbf{z}_1 - \mathbf{z}_2| \gg 1/T$:

$$\mathcal{Z}_{\rm KvB} = \int d^3z_1 d^3z_2 T^6 C\left(\frac{8\pi^2}{g^2}\right)^4 \qquad (20)$$
$$\times \left(\frac{\Lambda e^{\gamma_{\rm E}}}{4\pi T}\right)^{22/3}\left(\frac{1}{Tr_{12}}\right)^{5/3}\left(2\pi + \frac{\phi\bar{\phi}}{T}r_{12}\right)$$
$$\times (\phi r_{12}+1)^{\frac{4\phi}{3\pi T}-1}\left(\bar{\phi}r_{12}+1\right)^{\frac{4\bar{\phi}}{3\pi T}-1}$$
$$\times \exp\left[-V^{(3)}P(\phi) - 2\pi r_{12}P''(\phi)\right],$$

with a constant $C \approx 1$ and $P(\phi)$ as in Eq. (16). The second derivative of the latter,

$$P''(\phi) = \frac{1}{\pi^2 T}\left[\pi T\left(1-\frac{1}{\sqrt{3}}\right) - \phi\right] \qquad (21)$$
$$\times \left[\bar{\phi} - \pi T\left(1-\frac{1}{\sqrt{3}}\right)\right],$$

determines the leading dyon–dyon interaction, which is a pure quantum effect. $\Lambda$ is the scale parameter of the Pauli–Villars regularization scheme. When the dyon separation is much larger than their core sizes, $r_{12} \gg 1/\phi > 1/T$, $r_{12} \gg 1/\bar{\phi} > 1/T$, the expression simplifies further to

$$\mathcal{Z}_{\rm KvB} = \int d^3z_1 d^3z_2 T^6 (2\pi)^{8/3} C\left(\frac{8\pi^2}{g^2}\right)^4 \qquad (22)$$
$$\times \left(\frac{\Lambda e^{\gamma_{\rm E}}}{4\pi T}\right)^{\frac{22}{3}}\left(\frac{\phi}{2\pi T}\right)^{\frac{4\phi}{3\pi T}}\left(\frac{\bar{\phi}}{2\pi T}\right)^{\frac{4\bar{\phi}}{3\pi T}}$$

$$\times \exp\left[-2\pi r_{12} P''(\phi)\right] \exp\left[-V^{(3)} P(\phi)\right],$$

where the terms $(\phi/(2\pi T))^{4\phi/3\pi T}$ and $\left(\bar{\phi}/(2\pi T)\right)^{4\bar{\phi}/3\pi T}$ can be understood as fugacities of $M$ and $L$ dyons, respectively. The derivation allows one to estimate a linear-plus-Coulomb-like interaction between the constituents and relates the (leading) linear one to the sign of $P''(\phi)$. For $|L| > 0.787597$, it is attractive and the integral over $r_{12}$ converges and permits one to define the fugacity of the compound caloron. Based on this knowledge, in [40], the critical temperature $T_c = 1.125\Lambda$ has been estimated as the temperature below which trivial holonomy becomes unstable. At this temperature, the gluodynamic system starts to relax to larger $\phi$. For $|L| < 0.787597$, the dyon interaction becomes repulsive, and the caloron dissociates into separate dyons.

This fully dissociated case, however, is not covered by the derivation given in [40]. The statistical mechanics of the fully dissociated dyon gas still needs to be developed more in depth in order to justify the sketchy scenario leading to a free-energy density (19) with holonomy becoming stabilized at $\omega = 1/4$ (confinement).

## 4. LATTICE OBSERVABLES USED TO DETECT DYONS AND CALORONS

Now we are going to describe what lattice gauge theory can tell about the holonomy of semiclassical background fields. We concentrate on $SU(2)$ gauge fields discretized on an asymmetric lattice with periodic boundary conditions (p.b.c.) in all four directions. In a very few cases, also fixed holonomy boundary conditions (f.h.b.c.) have been implemented. For the latter, all timelike links at the spatial boundary $\Omega$ are fixed to equal Abelian values such that $L(\mathbf{x}) = 0$ for $\mathbf{x} \in \Omega$ in accordance with the vanishing order parameter in the confinement phase.

The respective ensembles of configurations have been created by heat-bath Monte Carlo using the standard Wilson plaquette action ($\beta = 4/g_0^2$),

$$S = \sum_{\mathbf{x},t} s(\mathbf{x},t) = \sum_{\mathbf{x},t} \sum_{\mu<\nu} s(\mathbf{x},t;\mu,\nu), \qquad (23)$$

$$s(\mathbf{x},t;\mu,\nu) = \beta\left(1 - \frac{1}{2}\mathrm{tr} U_{x,\mu\nu}\right),$$

$$U_{x,\mu\nu} = U_{x,\mu} U_{x+\hat{\mu},\nu} U^{\dagger}_{x+\hat{\nu},\mu} U^{\dagger}_{x,\nu}.$$

The lattice size will be $N_s^3 \times N_t$ with varying spatial extension $N_s = 8$ to 24 and with $T^{-1} \equiv N_t = 4, 5$, or 6. For $N_t = 4$, for instance, the model is well known to undergo the deconfinement phase transition at a critical coupling $\beta_c \simeq 2.299$ [56].

We generate the quantum gauge field ensemble $\{U_{x,\mu}\}$ by simulating the canonical partition function using the standard heat-bath Monte Carlo method. The equilibrium field configurations will be cooled by iteratively minimizing the action $S$. Usually, cooling in one form or another is used in order to smooth out short-range fluctuations, while (initially) leaving some long-range physics intact. The cooling method applied here is the standard relaxation method described a long time ago in [18].

This method, if applied without any further limitation, easily finds approximate solutions of the lattice field equations as shoulders (plateaus) of action in the relaxation history. Under certain circumstances, this defines and preserves the total topological charge of a configuration. However, the short-range structure of the vacuum fields is changed. Still, the type of classical solutions which are being selected depends on the phase which the quantum ensemble $\{U_{x,\mu}\}$ is meant to describe. We want to investigate smoothed fields at different stages of cooling by using a stopping criterion which selects the plateaus in a given interval of action.

The smoothed fields have been analyzed according to the spatial distributions of the following observables:

**Action density** computed from the local plaquette values:

$$s(\mathbf{x},t) = \sum_{\mu<\nu} s(\mathbf{x},t;\mu,\nu). \qquad (24)$$

**Topological density** computed with the standard discretization and averaged over the time variable:

$$q_t(\mathbf{x}) = -\frac{1}{N_t}\frac{1}{2^4 \cdot 32\pi^2} \qquad (25)$$

$$\times \sum_t \left( \sum_{\mu,\nu,\rho,\sigma=\pm1}^{\pm4} \epsilon_{\mu\nu\rho\sigma} \mathrm{tr}\left[U_{x,\mu\nu} U_{x,\rho\sigma}\right] \right).$$

**Polyakov loop** defined as

$$L(\mathbf{x}) = \frac{1}{2}\mathrm{tr} \prod_{t=1}^{N_t} U_{\mathbf{x},t,4}. \qquad (26)$$

**Abelian magnetic fluxes and monopole charges** defined within the maximally Abelian gauge (MAG). The latter is found by maximizing the gauge functional

$$A[g] = \frac{1}{2}\sum_{x,\mu} \mathrm{tr}(U^g_{x,\mu}\tau_3 U^{g\dagger}_{x,\mu}\tau_3) \qquad (27)$$

with respect to gauge transformations $U_{x,\mu} \to U^g_{x,\mu} = g(x)U_{x,\mu}g^\dagger(x+\hat{\mu})$. Abelian link angles $\theta_{x,\mu}$ are then defined by Abelian projection onto the diagonal $U(1)$ part of the link variables $U_{x,\mu} \in SU(2)$. According to the DeGrand–Toussaint prescription [57], a gauge-invariant magnetic flux $\bar{\Theta}_p$ through an oriented plaquette $p \equiv (x, \mu\nu)$ is defined by splitting the plaquette $\Theta_p = \theta_{x,\mu} + \theta_{x+\hat{\mu},\nu} - \theta_{x+\hat{\nu},\mu} - \theta_{x,\nu}$ into $\Theta_p = \bar{\Theta}_p + 2\pi n_p$, $n_p = 0, \pm 1, \pm 2$, such that $\bar{\Theta}_p \in (-\pi, +\pi]$. The (quantized) magnetic charge of an elementary 3-cube $c$ is then

$$m_c = \frac{1}{2\pi} \sum_{p \in \partial c} \bar{\Theta}_p,$$

where $\partial c$ is the boundary of this cube.

**Eigenvalues and eigenmode densities** of the non-Hermitian standard Wilson–Dirac operator

$$\sum_{y,s,\beta} D[U]_{xr\alpha,ys\beta}\psi_{s\beta}(y) = \lambda\psi_{r\alpha}(x), \qquad (28)$$

$$D[U]_{xr\alpha,ys\beta} = \delta_{xy}\delta_{rs}\delta_{\alpha\beta} - \kappa \sum_\mu \{\delta_{x+\hat{\mu},y}(\mathbf{1}_D - \gamma^\mu)_{rs}(U_{x,\mu})_{\alpha\beta} + \delta_{y+\hat{\mu},x}(\mathbf{1}_D + \gamma^\mu)_{rs}(U^\dagger_{y,\mu})_{\alpha\beta}\},$$

are studied with both time-antiperiodic and time-periodic b.c. For our purposes, i.e., for demonstrating the qualitative properties of a caloron solution, on smooth lattice fields it will be sufficient to consider this operator which explicitly breaks chiral invariance. We find the $\lambda$ spectrum and the eigenfunctions with the help of the implicitly restarted Arnoldi method and use the standard ARPACK code package for this aim [58, 59].

**Nonstaticity** $\delta_t$, is defined as

$$\delta_t = \frac{N_t}{4} \sum_{\mathbf{x},t} |s(\mathbf{x}, t+1) - s(\mathbf{x}, t)|/S. \qquad (29)$$

The normalization factor $N_t/4$ has been chosen such that $\delta_t$ will not change for a given KvB solution discretized on lattices with a varying time extent $N_t$. It allows one to discriminate between caloron and $DD$ solutions taking into account that $\delta_t$ depends monotonically on the spatial distance $d$ between the two centers of the solution. The "bifurcation" value for the recombination of dyon pairs into single calorons is $\delta_t^* \simeq 0.27$ for analytic, lattice-discretized KvB solutions of maximal nontrivial asymptotic holonomy $L_\infty = 0$ for which the constituent dyons get equal masses. If $\delta < \delta_t^*$, two dyons can be distinguished by the two maxima of the action density. For $\delta > \delta_t^*$, the dyons will appear recombined into a caloron with only one action density maximum.

**Mean violation** of the equations of motion

$$\Delta = \frac{1}{4N_s^3N_t} \sum_{x,\mu} \frac{1}{2}\mathrm{tr}\left[(U_{x,\mu} - \tilde{U}_{x,\mu})(U_{x,\mu} - \tilde{U}_{x,\mu})^\dagger\right], \qquad (30)$$

$$\tilde{U}_{x,\mu} = c\sum_{\nu\neq\mu}\Big[U_{x,\nu}U_{x+\hat{\nu},\mu}U^\dagger_{x+\hat{\mu},\nu} + U^\dagger_{x-\hat{\nu},\nu}U_{x-\hat{\nu},\mu}U_{x+\hat{\mu}-\hat{\nu},\nu}\Big]$$

is the local link variable representing the solution of the lattice equation of motion, with all degrees of freedom coupled to it held fixed. The factor $c$ is just a normalization of the staple sum such that $\tilde{U}_{x,\mu} \in SU(2)$. The replacement $U_{x,\mu} \to \tilde{U}_{x,\mu}$ is exactly the local cooling step as applied throughout this paper.

## 5. SIMPLEST SOLUTIONS OF THE $SU(2)$ LATTICE FIELD EQUATIONS

In the first part of our investigation, we have searched for topologically nontrivial objects with the lowest possible action, i.e., at $S \simeq S_{\mathrm{inst}}$ and below, late in the cooling history, in order to find systematic dependences of the selected solutions on the spatial boundary conditions and on the temperature of the original Monte Carlo ensemble. First, we started from equilibrium configurations generated in the neighborhood of the deconfinement phase transition. Later on, we lowered the temperature by increasing the lattice extent in the time direction. Cooling was always stopped on plateau values of the action $S < 1.5S_{\mathrm{inst}}$ as a function of the number of iteration steps. We have selected those configurations for which the mean violation of the equations of motion $\Delta$ according to Eq. (30) passed a low-lying minimum.

For each $\beta$, we have scanned the topological content of several hundred cooled configurations. At $N_t = 4$ and $\beta = 2.2$, i.e., in the confinement phase ($T \simeq 0.8T_c$), for both kinds of boundary conditions (p.b.c. and f.h.b.c), we have most frequently found objects which can be identified as KvB solutions in a nondissociated form or dissolved into a dyon pair. Below, we shall demonstrate that they indeed show all known features of KvB solutions despite the fact that, on a 4-torus, solutions of topological charge $Q_t = \pm 1$ do not exist in a mathematically strict sense [60]. The frequency of appearance of $DD$ vs. that of calorons in the ensemble of the detected KvB solutions will be discussed later. Compared with the confinement case, we have seen self-dual KvB solutions to become strongly suppressed at $N_t = 4$, $\beta = 2.4$, i.e., at $T \simeq 1.3T_c$.

Moreover, for $T \leq T_c$ at the level of one-instanton action values, we have detected configurations which

have vanishing total topological charge, $Q_t = 0$, and consist of two distinct topological lumps of opposite charge. They are static in time and can be easily interpreted as $D\bar{D}$ pairs. By tracing back the cooling history, we have convinced ourselves that such pairs can occur through a partial annihilation of $D$ with $\bar{D}$ in superpositions of $DD$ pairs with $\bar{D}\bar{D}$ pairs. These non-self-dual $D\bar{D}$ configurations will be discarded from our further discussion in this section.

Definitely, below the one-instanton action level, we have also detected very stable but purely Abelian objects which in an early work by Laursen and Schierholz [61] were interpreted as 't Hooft–Polyakov monopoles thought to characterize dominantly the deconfinement phase. We shall show below that such configurations are more likely extended Dirac sheets (DS), which, in the thermodynamic limit, should not play a role in the Euclidean path integral.

(i) $DD$ pairs: For a special $DD$ solution found with p.b.c., we show in Figs. 1*a* and 1*b* 2D cuts of the topological charge density $q_t(\mathbf{x})$ and of the Polyakov loop distribution $L(\mathbf{x})$, respectively. We clearly see the opposite-sign peaks of the Polyakov loop variable correlated with the equal-sign maxima of the topological charge density. The boundary values of the Polyakov loop vary slightly because they are not fixed here to a well-defined value. This is the only difference observed between the two types of b.c. In principle, for p.b.c., the holonomy can be arbitrary. As a consequence, the ratio of the action carried by the well-separated dyon constituents can take any value. For the same $DD$ solution, Figures 1*c* and 1*d* show the scatter plot of the 70 lowest complex eigenvalues of the Wilson–Dirac fermion operator according to Eq. (28) for $\kappa = 0.14$, for both time-periodic Fig. 1*c* and time-antiperiodic Fig. 1*d* b.c. for the fermion fields. In both cases, we find one isolated low-lying real eigenvalue which can be related to a zero-mode of the continuum Dirac operator. The corresponding (projected) eigenmode densities $(\psi^\dagger\psi)(x)$ are drawn below, in Figs. 1*e* and 1*f*. They show a localization behavior as analytically proposed in Eq. (14). For the time-antiperiodic b.c., the eigenmode is localized at the dyon exhibiting the negative peak of the Polyakov loop related to the Taubes winding [49]. For the given special solution, we have carried out a fit with the analytic formula for the action density Eq. (13). With the corresponding parameter values $\mathbf{x}_1$, $\mathbf{x}_2$ and $\omega$, we have computed the topological charge density and the Polyakov loop as well as the fermionic mode densities for both kinds of fermionic b.c. from the analytic KvB solution. For all those observables, we found an impressive agreement between the "measured" and "predicted" spatial dependences [44]. Within the MAG, we have searched for the Abelian monopole content of the field configurations under inspection. For static $DD$ solutions, we always find a pair of static (anti)monopoles with world lines coinciding with the centers of the dyons.

(ii) Caloron configurations: In Fig. 2, we show a typical caloron solution, with an approximately 4D rotationally invariant action distribution, also obtained at $\beta = 2.2$ from cooling with p.b.c. Again, we plot 2D cuts for the topological charge density and for the Polyakov loop and present the plot for fermionic eigenvalues together with the 2D cuts of eigenmode density for the distinct real eigenvalue. The full topological charge $Q_t$ is unity. But most important for us is the observation of narrow opposite-sign peaks of the Polyakov loop (the "dipole structure"), which shows that this caloron possesses a nontrivial asymptotic holonomy different from "old" HS calorons, i.e., intrinsic dyon–dyon constituents. The fermionic zero modes for time-periodic and time-antiperiodic b.c. for this configuration are only slightly shifted relative to each other. A reasonable fit with the analytic solution can be obtained, showing that this caloron is nothing but a limiting case of the generic KvB solutions. The typical caloron configurations show, after putting them into MAG, a closed Abelian monopole loop circulating around the maximum of the action density in the 4D space.

At that point, we may conclude that cooling, even with nonfixed holonomy at the spatial boundary, yields almost-classical solutions which show all characteristics of the KvB calorons.

(iii) Dirac sheets: Finally, by cooling with both kinds of spatial b.c., we have found objects becoming very stable against cooling at even lower action, i.e., $S < S_{\mathrm{inst}}$ [43, 45]. Their (color) electric contribution to the action is very small compared with the magnetic contribution. Moreover, they are almost perfectly static. Employing MAG, we have convinced ourselves that they are purely Abelian. In the deconfinement phase, they are the most frequent events [43, 61], whereas in the confinement phase, they occur quite rarely directly in the cooling process. There, they are more likely to appear after $D\bar{D}$ pairs annihilate in the final stage of the relaxation. We have investigated these DS events on lattices $N_s^3 \times N_t$ with $N_t = 4$ and $N_s = 8, 10, 12, 16, 20$ for both confinement ($\beta = 2.2$) and deconfinement ($\beta = 2.4$). The mean actions of these configurations selected by the cooling process are presented in Fig. 3, where its dependence on $N_s/N_t$ is shown to have the tendency $S_{\mathrm{DS}}/S_{\mathrm{inst}} \to N_t/N_s$. The latter dependence is easily understood in terms of a constant Abelian magnetic field $B_x^3 = 4\pi/N_s^2$ the action of which is equal to [62]

$$S_{\mathrm{DS}} = \frac{4}{g_0^2}\left(1 - \cos\left(\frac{B_x^3}{2}\right)\right) N_s^3 N_t \qquad (31)$$

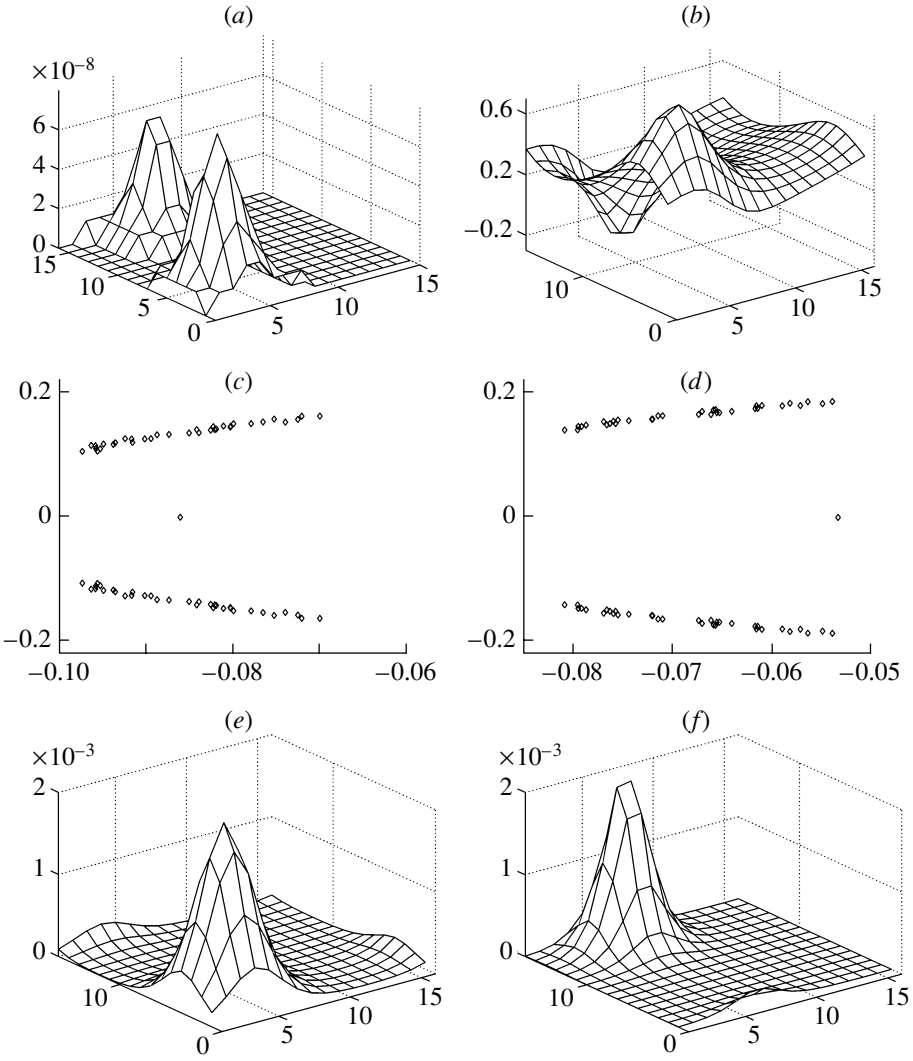


**Fig. 1.** Various portraits of a self-dual *DD* pair obtained by cooling under periodic gluonic boundary conditions. The subpanels show appropriate 2D cuts of the topological charge density in lattice units (*a*) and of the Polyakov loop (*b*), the plot of lowest fermionic eigenvalues (Re, Im) (*c*, *d*), and the 2D cut of the real-mode fermion densities in lattice units (*e*, *f*) for the cases of time-periodic (*c*, *e*) and time-antiperiodic (*d*, *f*) fermionic boundary conditions, respectively ($\beta = 2.2$ and lattice size $16^3 \times 4$).

$$\approx \frac{1}{2g_0^2}(B_x^3)^2 N_s^3 N_t = \frac{8\pi^2}{g_0^2}\frac{N_t}{N_s} = S_{\rm inst}\frac{N_t}{N_s}.$$

All DS events found in the confined phase for $N_s = 20$ (when put into MAG) show exactly such an Abelian magnetic field. For smaller $N_s$ in the confined phase, the fluxes are not always completely homogeneous and absolutely stable, whereas the fluxes are definitely unstable for all $N_s$ in the deconfined phase. As we have shown in [47], this stability pattern [63, 64] is directly related to the nontrivial holonomy observed within the confinement phase (stability is guaranteed for $|L| < \cos(\sqrt{\pi}N_t/N_s)$) and to the trivial one ($L_\infty = \pm 1$), which naturally occurs in the deconfinement phase.

Anyway, the DS configurations should not contribute in the thermodynamic limit. Therefore, in what follows, we shall concentrate on (anti-)self-dual configurations of the KvB type for action values $S \simeq S_{\rm inst}$.

## 6. DYON PAIRS AND THEIR RECOMBINATION INTO CALORONS

The possibility of observing the dyonic constituents of a KvB caloron as separate lumps of action depends on the value of $(\pi\rho/b)^2$. In $SU(2)$, lattice gauge theory (LGT) at $\beta = 2.2$ on a lattice $16^3 \times 4$ fits KvB solutions provided $\rho \approx 2.5a$ ($a$ is the lattice spacing) [42]. With $b = 4a$, we have

$$(\pi\rho/b)^2 \approx 4 \gg 1.$$

This means that dyons become well-separated. On a lattice $16^3 \times 6$ (with $b = 6a$) and at the same $\beta = 2.2$ (i.e., at a temperature 1.5 times lower), the parameter

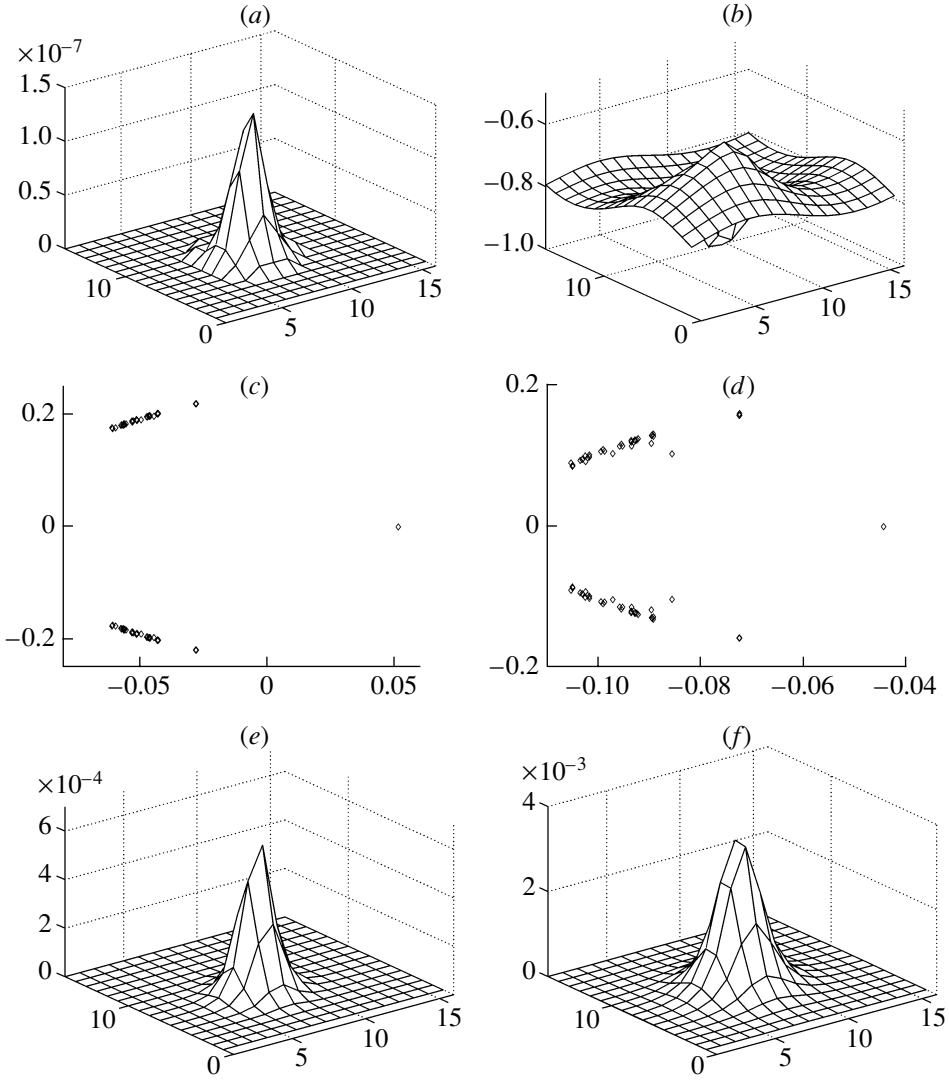


**Fig. 2.** As Fig. 1, but for self-dual caloron configuration.

$(\pi\rho/b)^2$ would be of the order $O(1)$. Then, from this simple arithmetic, one would expect that calorons are not dissociated anymore into separate dyonic lumps.

It is possible to measure the distance between dyons by detecting the peaks of the action density on the lattice only in the case of well-separated objects. Indirectly, we are able to measure this distance by means of the nonstaticity parameter $\delta_t$ according to Eq. (29), since for analytic KvB solutions there is a one-to-one correspondence between the distance and the nonstaticity, if the asymptotic holonomy is given.

We have generated Monte Carlo ensembles each of $O(10^4)$ independent $SU(2)$ gauge field configurations at $\beta = 2.2$ on lattices $16^3 \times N_t$ with $N_t = 4$, 5, and 6.

We have selected (anti-)self-dual solutions after cooling by imposing the following criteria:

The decrease in action has slowed down to $|\Delta S|/S_{\rm inst} < 0.05$.

The action fits into the window $0.5 < S/S_{\rm inst} < 1.25$.

The violation of the equations of motion passes through a minimum.

The efficiency of these conditions was such that 80% ($N_t = 4$), 60% ($N_t = 5$), and 55% ($N_t = 6$) of the equilibrium configurations ended up in a cooled configuration at the one-instanton plateau. These are the classical caloron configurations which are described in the following with respect to their dyonic properties.

We remind the reader that the recombination threshold $\delta_t^* = 0.27$, strictly speaking, reflects the recombination for maximally nontrivial holonomy only, i.e., with an asymptotic value of the Polyakov

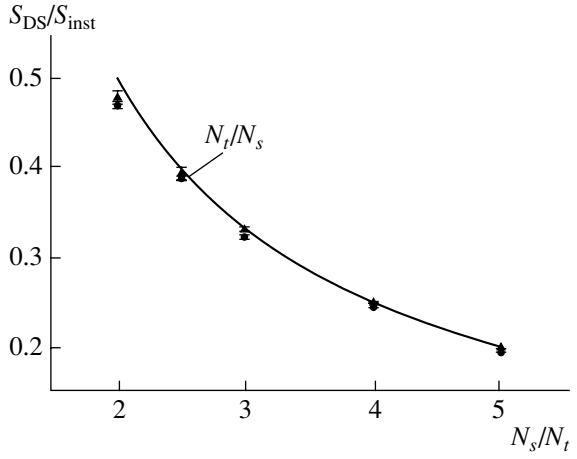


**Fig. 3.** Action of DS events in the confined phase (triangles, $\beta = 2.2$) and in the deconfined one (circles, $\beta = 2.4$) for $N_t = 4$ as a function of the ratio $N_s/N_t$. The curve presents $N_t/N_s$.

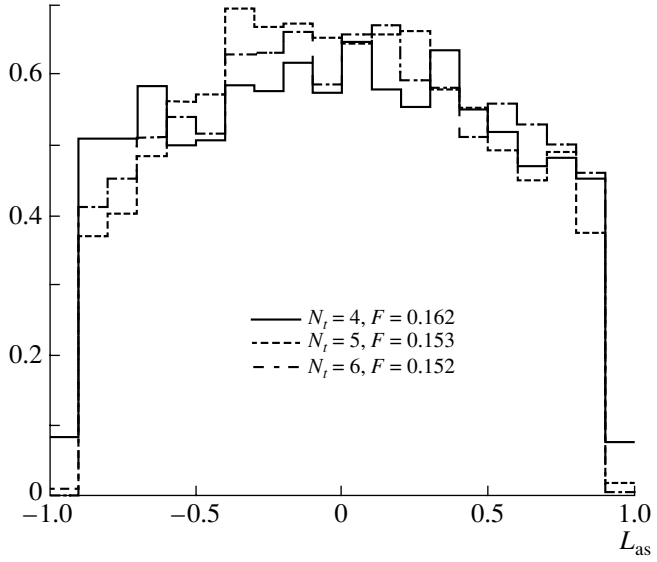


**Fig. 4.** The (normalized) distribution of the asymptotic holonomy $L_\infty \equiv L_{as}$ for cooled configurations at one-instanton action plateaus corresponding to three temperatures below deconfinement at $N_t = 4$, 5, and 6 ($\beta = 2.2$, $N_s = 16$).

line $L_\infty = 0$. If one performs cooling without special restrictions concerning the holonomy, there is no guarantee that the asymptotic holonomy of the caloron configurations still coincides with the average Polyakov line of equilibrium configurations in confinement. In order to define an approximate asymptotic holonomy $L_\infty$ for each cooled configuration, we have determined the average of $L(\mathbf{x})$ over a 3D subvolume where the local 3D action density $s(\mathbf{x})$ is low; for definiteness, $s(\mathbf{x}) < 10^{-4}$. In Fig. 4, we present the distributions of the cooled configurations over $L_\infty$ for the three cases $N_t = 4$, 5, and 6. In the legend, we show the respective volume fraction ($F \approx 0.15$) of the 3D volume (i.e., far from the lumps of action and topological charge) over which the "asymptotic" value $L_\infty$ is defined as an average.

As explained above, the nonstaticity $\delta_t$ is a measure which describes the distance from a perfectly (Euclidean) time-independent configuration. In other words, the distributions of nonstaticity of caloron events obtained by cooling of equilibrium lattice configurations can be considered as a substitute for the distribution in dyon distances $d$. This quantity can be directly measured for cooled lattice gauge field configurations. We show the $\delta_t$ distributions for all our cooling products obtained at $\beta = 2.2$ on $16^3 \times N_t$ lattices in Fig. 5. In an attempt to make a fair comparison with calorons with nontrivial holonomy and to correct for the possible evolution of the asymptotic holonomy away from $L_\infty = 0$ during the cooling process, we defined a subsample by the requirement $|L_\infty| < 1/6$. One can see that the cut with respect to the asymptotic holonomy selects cooled configurations from the flat central part of the histogram shown in Fig. 4. On the other hand, we notice that a considerable fraction of cooled configurations have developed an asymptotic holonomy $|L_\infty| > 1/6$. In Fig. 5*a*, we show the probability distribution over $\delta_t$ for cooled configurations (with an action at the one-instanton plateau) without the cut with respect to the asymptotic holonomy $|L_\infty|$. One can see that a relatively high fraction of configurations, obtained from the Monte Carlo equilibrium with $N_t = 4$, have $\delta_t < \delta_t^* = 0.27$. This means that they would be identifiable as consisting of two constituents by looking for the 3D action density on the lattice. For $N_t = 5$, there are only a minority of cooled configurations which fall below the threshold $\delta_t = 0.27$. No static configurations (according to the nonstaticity criterion) have been found among cooling products at $N_t = 6$. We have repeated the same analysis after applying the cut with respect to the asymptotic holonomy, $|L_\infty| < 1/6$. Then we get modified histograms in $\delta_t$ for the three temperatures. This is shown in Fig. 5*b*. The histograms got more pronounced peaks in $\delta_t$, which are positioned around 0.125, exactly around $\delta_t^* = 0.27$, and around 0.5 for $N_t = 4$, 5, and 6, respectively.

There are other criteria which could be used to characterize a more or less static configuration, for example, the presence of static Abelian monopoles emerging in the maximal Abelian projection. By selecting the subsample of cooled configurations containing a static Abelian monopole–antimonopole pair, we can determine the distance of the caloron constituents $d$ through the distance $R$ between these

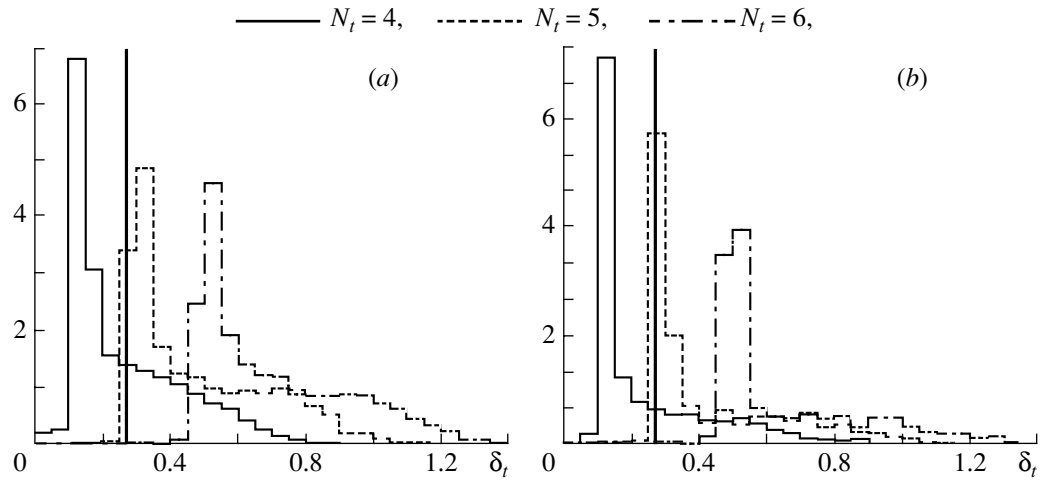


**Fig. 5.** Histograms (of nonnormalized distribution) for the nonstaticity $\delta_t$ after cooling, for three temperatures below the deconfinement transition with $N_t = 4$, 5, and 6 ($\beta = 2.2$, $N_s = 16$): (*a*) without any cut; (*b*) when a cut $|L_{\rm as}| < 1/6$ is applied ($L_{\rm as} \equiv L_\infty$). The thick vertical line marks the value $\delta_t^*$ where the caloron recombines.

(anti)monopoles. We make sure that there is a clear anticorrelation between $R$ and $\delta_t$ as long as static Abelian monopoles are found. Practically all cooled configurations with $\delta_t < \delta_t^*$ possess static Abelian monopole–antimonopole pairs, the locations of which fall close to the opposite peaks of the Polyakov loop $L = \pm 1$. Above $\delta_t^*$, however, the fraction, as well as the distance $R$, rapidly goes to zero. For $N_t = 6$, i.e., at lower temperature, at the peak value around $\delta_t = 0.5$, only 20% of the solutions are still characterized by a static pair of Abelian monopoles, whereas at higher nonstaticity this is never the case.

In order to represent how the Polyakov line behaves inside a lump of action, we have explored the neighborhood of the absolute maximum of the 3D action density (denoted as central point $\mathbf{x}_0$). This position could either be one of the two dyonic lumps (as long as they are separable) or the center of a recombined caloron. For this purpose, we have defined a locally summed-up Polyakov line $L_{\rm tot}$ (including the central point $\mathbf{x}_i$ ($i = 0$) and its six nearest neighbors $\mathbf{x}_i$ ($i = 1, \ldots, 6$)),

$$L_{\rm tot} = \sum_{i=0}^{6} L(\mathbf{x}_i), \qquad (32)$$

and a Polyakov line "dipole moment" over the same set of 3D lattice points with respect to the central point,

$$\mathbf{M}_{\rm tot} = \sum_{i=1}^{6} L(\mathbf{x}_i)(\mathbf{x}_i - \mathbf{x}_0). \qquad (33)$$

The absolute value $|L_{\rm tot}|$ of the first quantity tests the amount of local coherence of the Polyakov line. The absolute value $|\mathbf{M}_{\rm tot}|$ of the second quantity tests the amount of presence of opposite-sign Polyakov lines representing eventually two different constituents inside the same lump of action. Figure 6*a* shows how $|L_{\rm tot}|$ changes with $\delta_t$. For the temperature nearest to the transition, at $N_t = 4$, we see that $|L_{\rm tot}|$ falls from $\approx$4.0 to $\approx$1.0 at $\delta_t \geq 0.5$. We interpret this such that, in the region where constituents can be well separated according to the action density (at small $\delta_t$), they are characterized by a relatively smooth change of the Polyakov line inside. In the region of large $\delta_t$ where they are not separable according to the action density, the Polyakov line changes rapidly in the neighborhood of the absolute maximum of action density. For lower temperatures, $N_t = 5$ and 6, the relationship between these properties of an action cluster and $\delta_t$ is the same; however, separable lumps of action (at low $\delta_t$) become very rare. Figure 6*b* shows how the "dipole moment" $|\mathbf{M}_{\rm tot}|$ of the Polyakov line around a maximum of action density rises with increasing nonstaticity $\delta_t$. In the region where one can separate the constituents according to the action density, the "dipole moment" is small, emphasizing again the homogenity of the Polyakov line around the central point. In the region beyond $\delta_t^*$, the dipole moment gradually stabilizes around a value of 1.5.

In the same way as described so far, we have analyzed configurations obtained by cooling at higher action plateaus. Although the separation between static and nonstatic does not have the clear meaning as for the one-caloron case, the trend is the same: at lower temperature, the lumps of action tend to be more localized also in Euclidean time ("instanton-like").

We have asked whether our findings at lowest action cooling plateaus do really scale with the physical

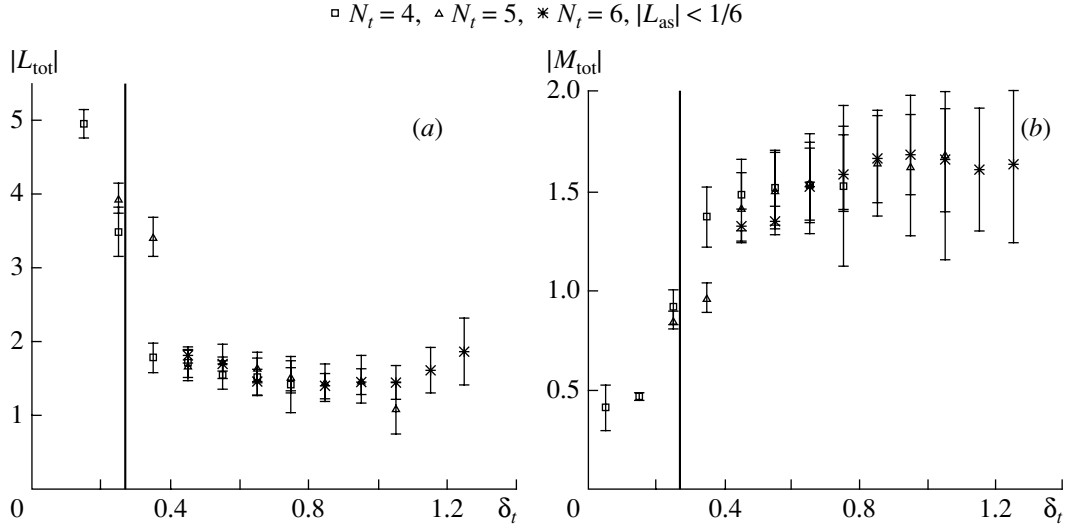


**Fig. 6.** The Polyakov line in the neighborhood of maxima of the action density: ($a$) modulus of the average summed-up Polyakov line $|L_{\rm tot}|$ (Eq. (32)) and ($b$) modulus of the corresponding "dipole moment" $|\mathbf{M}_{\rm tot}|$ (Eq. (33)) vs. nonstaticity $\delta_t$ for the subsample with asymptotic holonomy near zero. The thick vertical line marks the nonstaticity $\delta_t^*$ where the caloron recombines ($\beta = 2.2$, $N_s = 16$).

temperature. Therefore, at fixed lattice size, we have varied $\beta$ within the confinement case (e.g., for $N_t = 6$, we compared $\beta = 2.2$ with 2.36). We did not find a significant change in the $\delta_t$ distributions as long as we stayed within the same phase. This shows that our cooling results for the one-instanton action plateaus do not really depend on the physical temperature of the original Monte Carlo equilibrium gauge fields but rather on the nonsymmetric geometry of the lattice. This will certainly change when we investigate higher action plateaus with improved methods to remove ultraviolet fluctuations (improved cooling, smoothing, or smearing methods).

We have also cooled equilibrium configurations generated with $\beta = 2.2$ on symmetric lattices ($16^4$ representing "zero" temperature). In this case, we have found for the classical configurations at the plateau $S \approx S_{\rm inst}$ a broad distribution of nonstaticity with a maximum around $\delta_t \approx 2$ and with a tail extending beyond 3. These are obviously configurations with an action (topological charge) density well localized in all four Euclidean directions. There is a nontrivial behavior of the Polyakov line inside these nondissociated, instanton-like objects resembling that which has been seen—for the Polyakov line associated with the time direction—in the finite-temperature case for nondissociated calorons. We have mapped the cooled lattice configuration with the help of all four possible definitions of the Polyakov line, which are physically equivalent to each other on a symmetric lattice. Figure 7 shows the profiles of action density, topological charge, and the Polyakov lines (for four possible definitions) as they are seen in appropriate planes intersecting the lump through the center. The latter is defined as the maximum of the 4D action density. For all types of Polyakov lines, the characteristic double structure is seen exactly when the "asymptotic" value of the respective Polyakov line is not close to $\pm 1$. Still, it is not clear how to construct analytically an instanton solution which behaves nontrivially with respect to holonomy in all four directions. A KvB solution discretized on a symmetric lattice, for comparison, shows nontrivial holonomy only for the Polyakov loop in one, i.e., the "time," direction.

## 7. ENSEMBLES OF KvB SOLUTIONS AT HIGHER ACTION PLATEAUS

Within the confinement phase, for $0 < T \leq T_c$ and for both kinds of spatial b.c., we have also studied in some detail semiclassical configurations at higher action plateaus. They represent snapshots of earlier stages of the cooling histories because the stopping criteria were focused on multiples of the instanton action. This study should allow us to observe superpositions of classical solutions studied in Section 5 promising to be relevant for a semiclassical approximation of the nonzero $T$ partition function. So far in the literature, the semiclassical approach to QCD at nonzero temperature is entirely based on HS caloron solutions with trivial holonomy [6, 7]. Our main concern here is whether superpositions of solutions with nontrivial holonomy naturally occur under cooling.

Therefore, we expose equilibrium Monte Carlo lattice gauge field configurations to cooling, this time stopping under criteria which apply to different

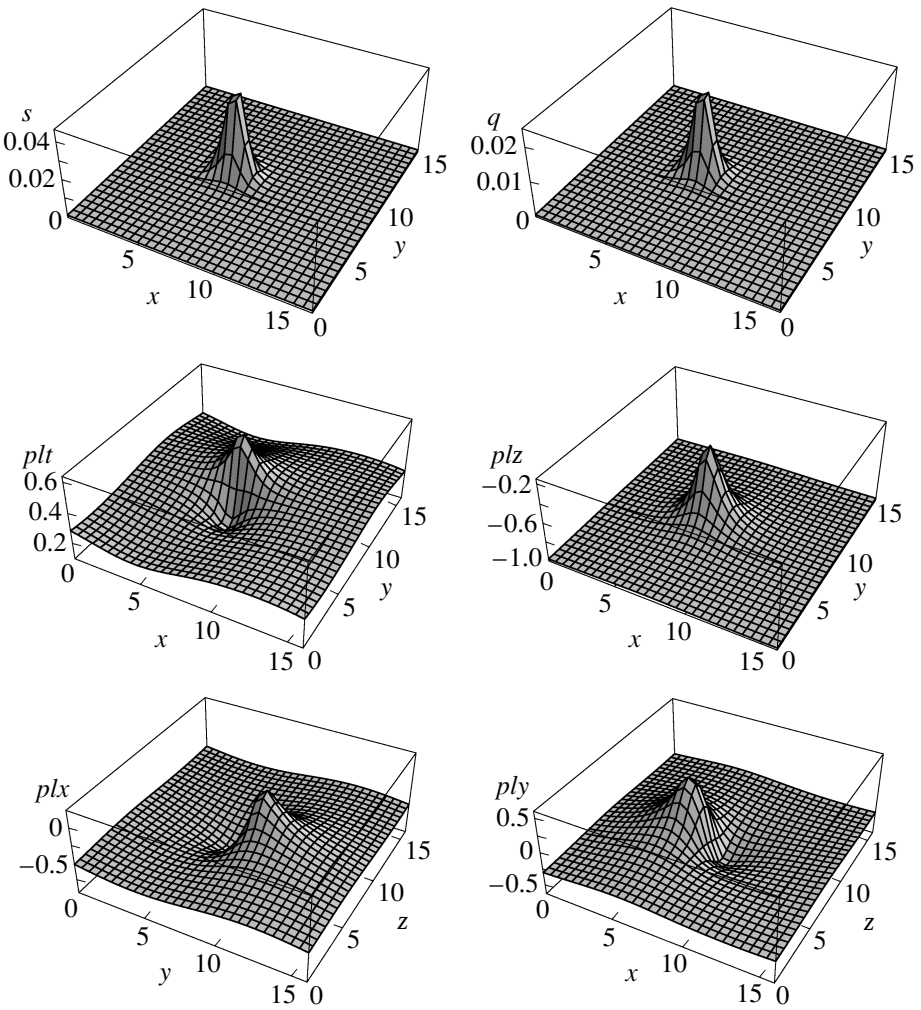


**Fig. 7.** Profiles of the action density ($s$), the topological charge density ($q$) (all in lattice units), and the Polyakov lines ($plt, plz, plx, ply$) calculated along all straight line paths parallel to the four axes for a $16^4$ lattice caloron found by cooling a Monte Carlo generated equilibrium gauge field ($\beta = 2.2$) down to the one-instanton action plateau. The center of the caloron (at the maximum of its action density) was found at the site $(x, y, z, t) = (7, 8, 8, 14)$. The planes shown in the figures cross just at this point.

subsequent action windows. We have been monitoring the landscape of topological density and of the Polyakov line operator as well as the localization of the fermionic zero modes in the semiclassical candidate configurations. Correspondingly, we have been triggering now by the condition

$$(m - 1/2)S_{\text{inst}} < S_n < (m + 1/2)S_{\text{inst}},$$
$$m = 2, 3, \ldots .$$

In particular, we inspected the first (highest) visible plateaus which occurred at various $m$ values, typically in the range $m \simeq 8{-}20$ for $\beta = 2.2$ on a lattice of the size $16^3 \times 4$. Then, we looked at the series of subsequent action plateaus.

To be more definite concerning the moment of taking snapshots under cooling along a plateau, we have additionally used the measure $\Delta$ for the mean violation of the classical lattice field equations per link [see Eq. (30)].

On the first visible plateau, we find a gas of localized lumps carrying topological charge, where an identification in terms of dyons $D$ and/or antidyons $\bar{D}$ is still difficult.

Independent of the kind of b.c. employed, at somewhat lower action plateaus with $m < 10$, we are able to clearly recognize dyons $D$ and antidyons $\bar{D}$ carrying noninteger topological charges. During cooling, more and more of these objects disappear. However, at all plateaus, we observe an even number of peaks of the spatial Polyakov loop landscape $|L(\mathbf{x})|$. For illustration, see Fig. 8, which shows one and the same gauge field configuration at different stages of the cooling process. In this case, f.h.b.c. with $L_\infty = 0$ have been used. At three subsequent lower action

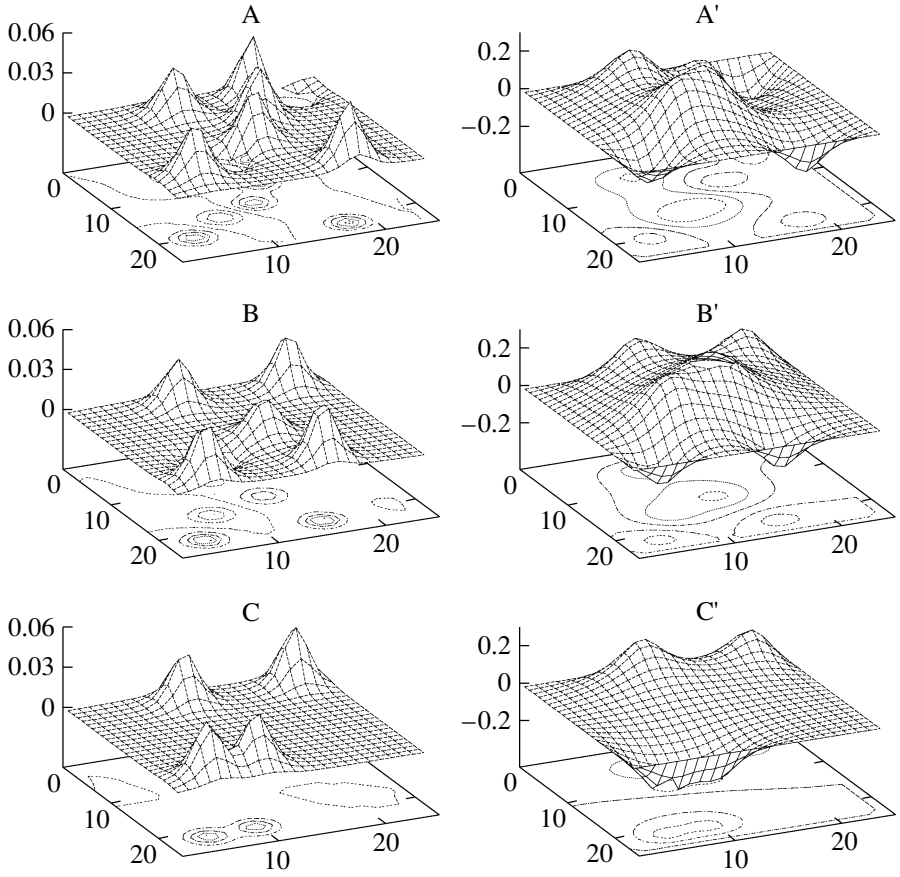


**Fig. 8.** Configurations after 800 (A, A′), 1650 (B, B′), and 7000 (C, C′) cooling steps. A, B, C show 2D projections of the topological charge density $q_t(x)$ (in lattice units) and A′, B′, C′ of the Polyakov loop $L(\mathbf{x})$, respectively. Cooling has been employed with f.h.b.c., $L(\mathbf{x}) = 0$ for $\mathbf{x} \in \Omega$ ($\Omega$ is the 3D volume boundary) ($\beta = 2.2$, lattice size $24^3 \times 4$).

plateaus, we have stopped at iteration steps A (for $m = 4$), B (for $m = 3$), and C (for $m = 2$), respectively, where the mean violation $\Delta$ of the lattice field equations passed through local minima. The corresponding semiclassical field configurations are then displayed in Fig. 8 by means of the 2D-projected (summed over the third coordinate) spatial topological charge density and the 2D-projected Polyakov loop distribution. One can recognize in these figures that, at stage A, we have a superposition of six dyons and two antidyons. The topological charge sector has been independently determined to be $Q_t = 2$. A $D\bar{D}$ pair decays or annihilates from A to B such that we have five dyons and one antidyon at the next stage. The topological charge did not change. Finally, stage C exhibits a superposition of four dyons, again with $Q_t = 2$. The latter configuration is very stable. While it stays at the same action over thousands of cooling sweeps, the nonstationarity $\delta_t$ gradually rises. A closer look then shows that the scale size of one of the dyon pairs shrinks, transforming this pair into a small undissociated and nonstatic caloron, which finally disappears after having turned into a tiny dislocation, strongly violating the equations of motion (as can be monitored by following $\Delta$). The average violation of the equations of motion peaks immediately before the configuration drops to the next plateau. This example shows that we have superpositions of noninteger $Q_t$ lumps, which can be interpreted as described in the previous section. To make sure that this is really the case, we also provide the eigenvalue scatter plots and the density $(\bar{\psi}\psi)(x)$ for the lowest modes of the Wilson–Dirac operator for stage B (see Fig. 9).

We have also studied the Abelian (anti)monopole structure after MAG fixing and Abelian projection and have seen a strong correlation of the peaks of the Polyakov loop with the positions of the (anti)monopoles. The pair structure in terms of Abelian monopoles, occurring on all action plateaus, as well as the annihilation of single (monopole–antimonopole) pairs during further relaxation, provides an additional signal for the topological content as superpositions of nontrivial holonomy caloron, $DD$ ($\bar{D}\bar{D}$), or $D\bar{D}$ pairs.

In order to understand this from the point of view of single-caloron solutions with nontrivial holonomy,

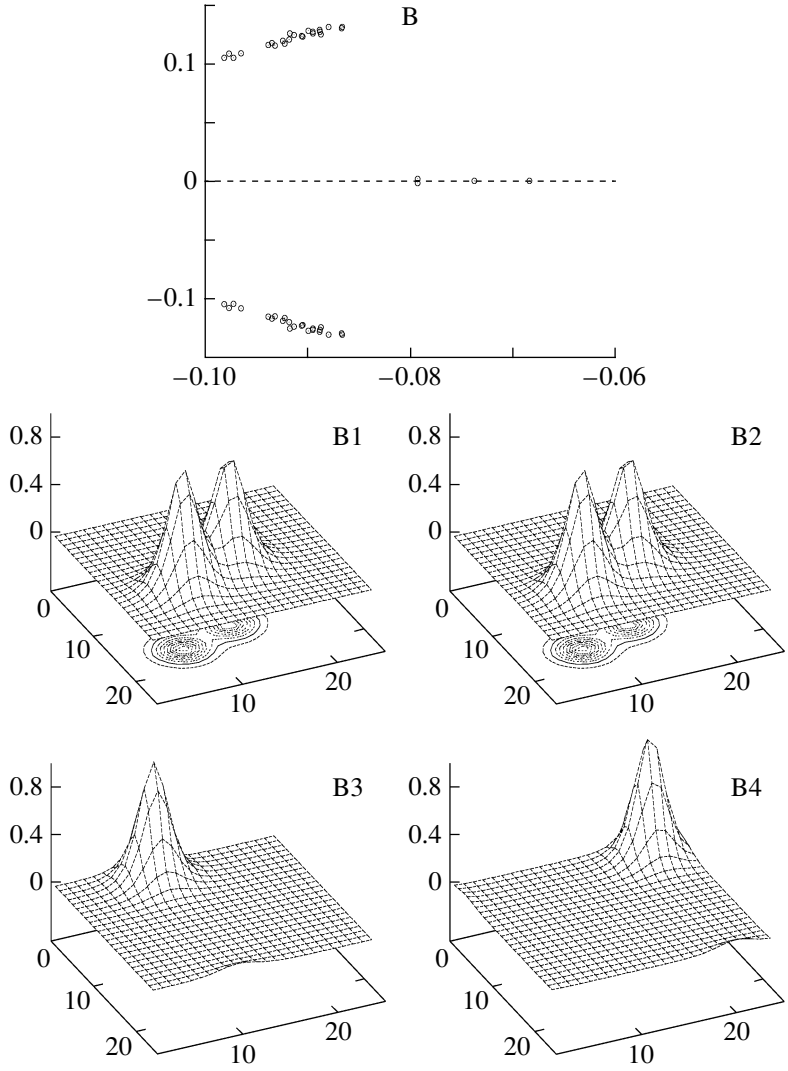


**Fig. 9.** The lattice field configuration depicted in Fig. 8 (B, B′) for 1650 cooling steps (f.h.b.c.). Here, B plots the eigenvalues of the Wilson–Dirac operator in the complex plane for $\kappa = 0.140$ and the case of time-periodic fermionic b.c.; B1, B2 show 2D projections of the fermionic mode densities (in lattice units) related to the two distinct *almost* real eigenvalues, whereas B3, B4 present the densities (in lattice units) related to the two real eigenvalues.

we have to find out whether (approximately) those asymptotic holonomy boundary conditions as typical for the dyonic (antidyonic) semiclassical background excitations are actually present during the cooling process when periodic boundary conditions (and not particular holonomy boundary conditions!) are applied to the full volume. To answer this question, the asymptotic holonomy $L_\infty$ as defined in Section 6 needs to be followed over the process of cooling. In Fig. 10, we show histograms of $L_\infty$ obtained at different plateaus during the cooling histories of an ensemble of $O(200)$ configurations produced at $\beta = 2.2$ on a $16^3 \times 4$ lattice with standard p.b.c. We see a clear peak at $L_\infty = 0$ for higher action plateaus. The distribution is narrower than the pure Haar measure would tell us. However, approaching lower lying plateaus, the real distribution becomes flat. Therefore, at the lowest plateaus, a cut with respect to the asymptotic holonomy $L_\infty$ is necessary to focus on the most static (most nontrivial) calorons.

Concluding, we can state that multicaloron solutions with nontrivial holonomy naturally occur in semiclassical configurations of higher action. From the analytic point of view, at present, only special classes of multicaloron solutions with nontrivial holonomy are known [65]. The study of these solutions shows that the calorons consist of dyon constituents which cannot be easily associated with single calorons. They constitute more complicated objects which hopefully can be described as multidyon events. This should then allow one to parametrize

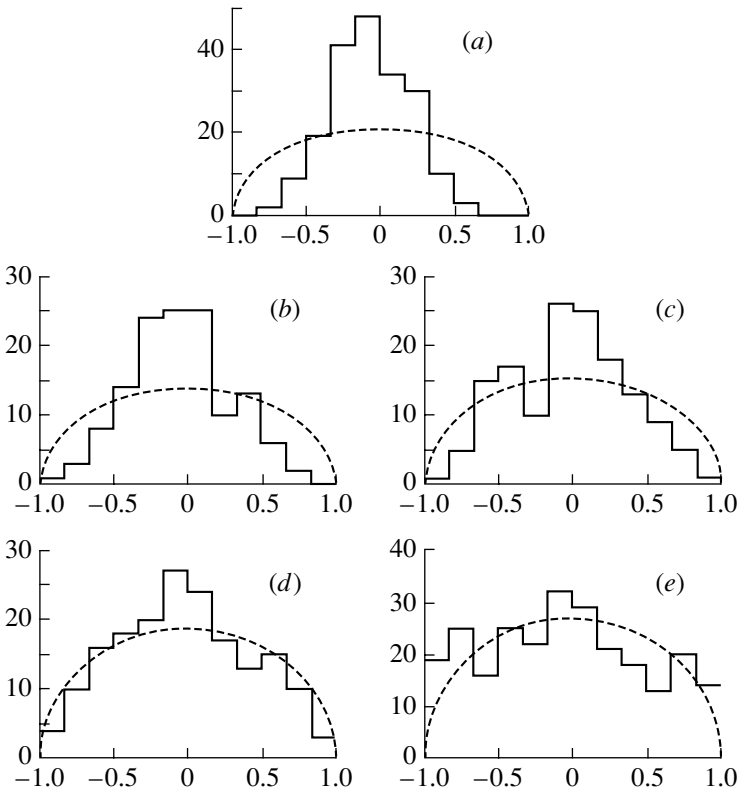


**Fig. 10.** Histogram $P(L_\infty)$ of the values of the Polyakov loop at "infinity" (as explained in the text) seen on the first plateau ($a$) and at plateaus with $m \simeq 4$ ($b$), 3 ($c$), 2 ($d$), and 1 ($e$). For comparison, the distribution expected from the pure Haar measure $P_{\mathrm{Haar}}(L) \sim \sqrt{1-L^2}$ is shown with the same normalization (dashed curves). The equilibrium ensemble of $O(200)$ configurations was generated at $\beta = 2.2$ (for lattice size $16^3 \times 4$, p.b.c.).

the path integral in terms of monopole degrees of freedom, as has been discussed in Section 3.

## 8. CONCLUSIONS AND OUTLOOK

In this review, we tried to demonstrate that the semiclassical treatment of QCD at finite temperature [7] has to be reconsidered taking into account (anti-)self-dual solutions of the field equations with nontrivial holonomy. Such solutions were analytically described by van Baal and his coworkers [38, 39, 65]. A first attempt to treat the path integral with these solutions was recently put forward by Diakonov and collaborators [40].

In order to show that KvB solutions are really relevant, we have used the lattice representation of the Yang–Mills path integral and generated statistical ensembles of finite-temperature fields with the help of the Monte Carlo method. Then, by iteratively minimizing the action, we have searched for semiclassical background configurations. The method is standard and has been known for almost twenty years [18–20], and has been refined and improved in different ways. Anyway, over the years, the outcome, i.e., the (approximate) solutions of the equations of motion, has always been interpreted in terms of standard instantons [1] or calorons [6], since the observables were mostly the action or topological charge densities, as well as the fermionic densities obtained from the low-lying spectrum of the (chirally improved) Dirac operator. Over many years, one has overlooked that these solutions in general have an intrinsic holonomy structure and a nontrivial asymptotic behavior related to nonvanishing potentials far away from the lumps of topological charge. In this paper, we have reviewed attempts to reveal this structure by carefully investigating the spatial Polyakov loop distributions and the localization of fermionic zero modes depending on whether time-periodic or time-antiperiodic boundary conditions for the fermion fields are imposed. The result is surprising. We clearly see a dyonic substructure even though the calorons or instantons produced at lower temperature or on a symmetric torus look like integer charged (topological charge is assumed) pseudoparticles in the 4D Euclidean world.

One is tempted to ask whether these observations are only an artifact of the cooling method. Having in mind that the low-lying fermionic spectrum is relatively robust against smoothing or smearing of the

gauge fields, Gattringer and his coworkers [50] studied the "mode hopping" of the fermionic zero modes of the equlibrium lattice gauge fields when changing the fermionic boundary conditions. They observed that the modes really change their localization and interpreted this as an indication of a dyonic structure a la Kraan–van Baal. This is clearly an interesting observation, although it remains open whether the given interpretation is the only possible one.

Our investigations presented here were restricted to the $SU(2)$ case. In the meantime, preliminary analogous results were also obtained for $SU(3)$ [66]. The $SU(3)$ KvB solutions generically contain three BPS monopoles as caloron constituents. This is a nontrivial generalization of the "old" caloron which by construction is an $SU(2)$ object embedded into $SU(3)$.

Still, it is not clear whether the new and nontrivial structures discussed in this paper will contribute to a better understanding of the mechanism driving the deconfinement transition. How far a certain working picture of a dilute gas of these configurations can be developed and whether it will improve our understanding of quark confinement itself remains an open question to which we hope to come back in the future. Nevertheless, we feel that some drawbacks found in interpreting lattice instantons [32–34] can be cured by classical solutions which properly take the nontrivial holonomy into account.

## ACKNOWLEDGMENTS

We gratefully acknowledge many stimulating discussions with Yu.A. Simonov at ITEP, Moscow. Moreover, we thank Pierre van Baal, Falk Bruckmann, Christof Gattringer, and Dirk Peschka for cooperation and useful discussions.

This work was partly supported by the Russian Foundation for Basic Research (project nos. 02-02-17308, 03-02-19491, and 04-02-16079), by DFG grant no. 436 RUS 113/739/0 and RFBR-DFG grant no. 03-02-04016, by the Federal Program of the Russian Ministry of Industry, Science, and Technology no. 40.052.1.1.1112, and by INTAS-00-00111. Two of us (B.V.M. and A.I.V.) gratefully appreciate the support of Humboldt University, Berlin, where this work was carried out to a large extent. E.-M.I. is supported by DFG (FOR 465).

## REFERENCES

1. A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59B**, 85 (1975).
2. G. 't Hooft, Phys. Rev. D **14**, 3432 (1976); **18**, 2199(E) (1978).
3. C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978).
4. E.-M. Ilgenfritz and M. Müller-Preussker, Nucl. Phys. B **184**, 443 (1981).
5. E. V. Shuryak, Nucl. Phys. B **203**, 93, 116, 140 (1982); D. I. Diakonov and V. Yu. Petrov, Nucl. Phys. B **245**, 259 (1984).
6. B. J. Harrington and H. K. Shepard, Phys. Rev. D **17**, 2122 (1978); **18**, 2990 (1978).
7. D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. **53**, 43 (1981).
8. E. V. Shuryak and T. Schäfer, Rev. Mod. Phys. **70**, 323 (1998).
9. D. I. Diakonov, Prog. Part. Nucl. Phys. **51**, 173 (2003).
10. Y. Nambu, Phys. Rev. D **10**, 4262 (1974); G. 't Hooft, in *Proceedings of the EPS International Conference "High-Energy Physics", Palermo, 1975,* Ed. by A. Zichichi (Editrice Compositori, Bologna, 1976); S. Mandelstam, Phys. Rep. **23**, 245 (1976); A. S. Kronfeld, M. L. Laursen, G. Schierholz, and U. J. Wiese, Phys. Lett. B **198**, 516 (1987); V. Bornyakov *et al.*, Phys. Lett. B **261**, 116 (1991); G. Bali, V. Bornyakov, M. Müller-Preussker, and K. Schilling, Phys. Rev. D **54**, 2863 (1996).
11. G. 't Hooft, Nucl. Phys. B **138**, 1 (1978); L. Del Debbio, M. Faber, J. Greensite, and S. Olejnik, Phys. Rev. D **55**, 2298 (1997); L. Del Debbio *et al.*, Phys. Rev. D **58**, 094501 (1998); K. Langfeld, H. Reinhardt, and O. Tennert, Phys. Lett. B **419**, 317 (1998); M. Engelhardt, K. Langfeld, H. Reinhardt, and O. Tennert, Phys. Lett. B **452**, 301 (1999); Phys. Rev. D **61**, 054504 (2000).
12. M. Fukushima *et al.*, Phys. Lett. B **399**, 141 (1997); Nucl. Phys. B (Proc. Suppl.) **63**, 513 (1998); M. Fukushima, H. Suganuma, and H. Toki, Phys. Rev. D **60**, 094504 (1999).
13. V. I. Shevchenko and Yu. A. Simonov, Phys. Rev. Lett. **85**, 1811 (2000).
14. M. Fukushima, E.-M. Ilgenfritz, and H. Toki, Phys. Rev. D **64**, 034503 (2001).
15. H. G. Dosch, Phys. Lett. B **190**, 177 (1987); H. G. Dosch and Yu. A. Simonov, Phys. Lett. B **205**, 339 (1988); A. Di Giacomo, H. G. Dosch, V. I. Shevchenko, and Yu. A. Simonov, Phys. Rep. **372**, 319 (2002).
16. E.-M. Ilgenfritz, B. V. Martemyanov, S. V. Molodtsov, *et al.*, Phys. Rev. D **58**, 114508 (1998).
17. E.-M. Ilgenfritz, B. V. Martemyanov, and M. Müller-Preussker, Phys. Rev. D **62**, 096004 (2000).
18. E.-M. Ilgenfritz *et al.*, Nucl. Phys. B **268**, 693 (1986).
19. J. Hoek, M. Teper, and J. Waterhouse, Nucl. Phys. B **288**, 589 (1987).
20. M. K. Polikarpov and A. I. Veselov, Nucl. Phys. B **297**, 34 (1988).
21. T. DeGrand, A. Hasenfratz, and De-cai Zhu, Nucl. Phys. B **475**, 321 (1996); **478**, 349 (1996).
22. T. DeGrand, A. Hasenfratz, and T. G. Kovacs, Nucl. Phys. B **505**, 417 (1997).
23. M. Feurstein, E.-M. Ilgenfritz, M. Müller-Preussker, and S. Thurner, Nucl. Phys. B **511**, 421 (1998).
24. T. DeGrand, A. Hasenfratz, and T. G. Kovacs, Nucl. Phys. B **520**, 301 (1998).

25. Ch. Gattringer, Phys. Rev. D **63**, 114501 (2001); Nucl. Phys. B (Proc. Suppl.) **119**, 122 (2003); Ch. Gattringer, I. Hip, and C. B. Lang, Nucl. Phys. B **597**, 451 (2001).
26. D. B. Kaplan, Phys. Lett. B **288**, 342 (1992); H. Neuberger, Phys. Lett. B **417**, 141 (1998); R. Narayanan and H. Neuberger, Nucl. Phys. B **412**, 574 (1994).
27. B. V. Martemyanov, S. V. Molodtsov, Yu. A. Simonov, and A. I. Veselov, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 679 (1995) [JETP Lett. **62**, 695 (1995)].
28. B. V. Martemyanov, S. V. Molodtsov, Yu. A. Simonov, and A. I. Veselov, Yad. Fiz. **60**, 565 (1997) [Phys. At. Nucl. **60**, 490 (1997)].
29. B. V. Martemyanov, Phys. Lett. B **429**, 55 (1998).
30. B. V. Martemyanov and S. V. Molodtsov, Yad. Fiz. **59**, 766 (1996) [Phys. At. Nucl. **59**, 733 (1996)].
31. B. V. Martemyanov and S. V. Molodtsov, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 133 (1997) [JETP Lett. **65**, 142 (1997)].
32. T. De Grand and A. Hasenfratz, Prog. Theor. Phys. Suppl. **131**, 573 (1998).
33. T. G. Kovacs and Z. Schram, Nucl. Phys. B (Proc. Suppl.) **73**, 530 (1999).
34. T. G. Kovacs, Phys. Rev. D **62**, 034502 (2000).
35. E.-M. Ilgenfritz and S. Thurner, hep-lat/9810010.
36. E.-M. Ilgenfritz and Ph. de Forcrand (unpublished).
37. K. Lee and C. Lu, Phys. Rev. D **58**, 025011 (1998).
38. T. C. Kraan and P. van Baal, Phys. Lett. B **435**, 389 (1998).
39. T. C. Kraan and P. van Baal, Nucl. Phys. B **533**, 627 (1998).
40. D. I. Diakonov, N. Gromov, V. Yu. Petrov, and S. Slizovskiy, hep-th/0404042.
41. E.-M. Ilgenfritz, M. Müller-Preussker, and A. I. Veselov, in *Proceedings of the NATO Advanced Research Workshop on Lattice Fermions and Structure of the Vacuum, Dubna, Russia, 1999*; hep-lat/0003025.
42. E.-M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, and A. I. Veselov, Nucl. Phys. B (Proc. Suppl.) **94**, 407 (2001).
43. E.-M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, and A. I. Veselov, Nucl. Phys. B (Proc. Suppl.) **106-107**, 589 (2002).
44. E.-M. Ilgenfritz *et al.*, Phys. Rev. D **66**, 074503 (2002).
45. E.-M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, and A. I. Veselov, hep-lat/0310030; Eur. Phys. J. C (in press).
46. E.-M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, and A. I. Veselov, hep-lat/0402010; Phys. Rev. D (in press).
47. E.-M. Ilgenfritz, B. V. Martemyanov, M. Müller-Preussker, and P. van Baal, hep-lat/0402020; Phys. Rev. D (to appear).
48. M. García Pérez, A. González-Arroyo, C. Pena, and P. van Baal, Phys. Rev. D **60**, 031901 (1999).
49. M. N. Chernodub, T. C. Kraan, and P. van Baal, Nucl. Phys. B (Proc. Suppl.) **83**, 556 (2000).
50. C. Gattringer, Phys. Rev. D **67**, 034507 (2003); C. Gattringer and S. Schaefer, Nucl. Phys. B **654**, 30 (2003); C. Gattringer and R. Pullirsch, hep-lat/0402008.
51. G. 't Hooft (unpublished); R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
52. M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975); E. B. Bogomol'nyi, Yad. Fiz. **24**, 861 (1976) [Sov. J. Nucl. Phys. **24**, 449 (1976)].
53. M. García Pérez, A. González-Arroyo, A. Montero, and P. van Baal, J. High Energy Phys. **9906**, 001 (1999).
54. C. Taubes, Commun. Math. Phys. **86**, 257, 299 (1982).
55. N. Weiss, Phys. Rev. D **24**, 475 (1981); **25**, 2667 (1982).
56. J. Engels, J. Fingberg, and M. Weber, Nucl. Phys. B **332**, 737 (1990).
57. T. A. DeGrand and D. Toussaint, Phys. Rev. D **22**, 2478 (1980).
58. H. Neff *et al.*, Phys. Rev. D **64**, 114509 (2001).
59. See: http://www.caam.rice.edu/software/ARPACK/
60. P. J. Braam and P. van Baal, Commun. Math. Phys. **122**, 267 (1989).
61. M. L. Laursen and G. Schierholz, Z. Phys. C **38**, 501 (1988).
62. V. K. Mitrjushkin, Phys. Lett. B **389**, 713 (1996).
63. M. García Pérez and P. van Baal, Nucl. Phys. B **429**, 451 (1994).
64. P. van Baal, Nucl. Phys. B (Proc. Suppl.) **47**, 326 (1996); Commun. Math. Phys. **94**, 397 (1984).
65. F. Bruckmann and P. van Baal, Nucl. Phys. B **645**, 105 (2002); F. Bruckmann, D. Nogradi, and P. van Baal, hep-th/0404210.
66. C. Gattringer *et al.*, Nucl. Phys. B (Proc. Suppl.) **129-130**, 653 (2004).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 890–891. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 925–926.*


# Klein–Gordon Equation for Quark Pairs in Color Superconductor*

**B. O. Kerbikov****

*Institute of Theoretical and Experimental Physics,*
*Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received July 23, 2004

**Abstract**—The wave equation is derived for quark pairs in color superconductor in the regime of low density/strong coupling. © *2005 Pleiades Publishing, Inc.*

During the last five years, color superconductivity became a compelling topic in QCD—see the review papers [1]. Broadly speaking, we have a fair understanding of color superconductivity physics in the high-density/weak-coupling regime. In the low-density/strong-coupling region, the situation is different. Here, the theory faces the well-known difficulties of nonperturbative QCD and use is made of models like NJL or instanton gas. By low density, we mean quark densities 3–4 times larger than that in normal nuclear matter. Model calculations show (see [2] and references in [1]) that, at such densities, the gap equation acquires a nontrivial solution. This was interpreted [1] as the onset of the Bardeen–Cooper–Schrieffer (BCS) regime, i.e., the formation of the condensate of Cooper pairs made of $u$ and $d$ quarks. It is known, however, that a nonzero value of the gap is only a signal of the presence of fermion pairs. Depending on the dynamics of the system, on the fermion density, and on the temperature, such pairs may be either stable or fluctuating in time, may form a BCS condensate or dilute Bose gas, or undergo a Bose condensation. The continuous evolution from the BCS regime to the regime of Bose–Einstein condensation (BEC) is called the BCS–BEC crossover. Such a transition takes place either by increasing the strength of the interaction or by decreasing the carrier density. The fact that the BCS wave function may undergo a smooth evolution and describe the tightly bound fermion pairs was first noticed long ago [3–6]. According to [5], the remark that "there exists a transformation that carries the BCS into BE state" was originally made by F.J. Dyson in 1957 (i.e., the same year that the BCS paper [7] was published). The BCS–BEC crossover for quarks was first discussed in [8]. The general description of the crossover for quarks will be given elsewhere.

In the present paper, we shall derive an equation for the quark pairs in the low-density/stong-coupling limit. This equation is obtained directly from the mean-field gap equation. Our derivation follows the work by Nozieres and Schmitt-Rink [4], who obtained the Schrödinger equation starting from the BCS solution. Quarks in color superconductor are relativistic particles, and therefore we shall arrive at the Klein–Gordon equation.

Our starting point is the expression for the thermodynamic potential for the two-flavor superconductor (2SC). In the 2SC phase [1], pairing takes place between the $u$ and $d$ quarks, while the $s$ quark is out of the game until the density increases, so that the chemical potential becomes substantially larger than the mass of the $s$ quark. Here, we consider only the $T = 0$ case. The possible coexistence of the chiral and diquark condensates is neglected on the basis of the Anderson theorem [9]. The expression for the thermodynamic potential reads [2, 9]

$$\Omega(T=0,\mu;\,\Delta) = \frac{\Delta^2}{4g} - \frac{2}{\pi^2}\int dq q^2 \left\{\sqrt{(E_q-\mu)^2+\Delta^2} + \sqrt{(E_q+\mu)^2+\Delta^2} + |\mu - E_q| + E_q\right\}, \tag{1}$$

where $E_q = \sqrt{q^2+m^2}$.

The four-fermion interaction constant $g$ has dimension $m^{-2}$. From (1), one obtains the following gap equation:

$$\Delta = \frac{4g}{\pi^2}\int dq q^2 \left(\frac{\Delta}{\mathcal{E}_q} + \frac{\Delta}{\bar{\mathcal{E}}_q}\right), \tag{2}$$

where

$$\mathcal{E}_q = \sqrt{(E_q-\mu)^2+\Delta^2},$$
$$\bar{\mathcal{E}}_q = \sqrt{(E_q+\mu)^2+\Delta^2}.$$

*This article was submitted by the author in English.

** e-mail: borisk@heron.itep.ru

An important remark concerning the structure of Eqs. (1), (2) is due here. We have tacitly assumed that the four-fermion interaction between quarks is pointlike. In the general case, instead of (2), one should write

$$\Delta_p = \frac{4g}{\pi^2}\int dqq^2 V_{pq}\left(\frac{\Delta_q}{\mathcal{E}_q}+\frac{\Delta_q}{\bar{\mathcal{E}}_q}\right). \qquad (3)$$

However, the crude approximation (1), (2) is sufficient to obtain the general structure of the Klein–Gordon equation. Next, making use of the standard Bogolyubov functional, we introduce the wave functions of the quark–quark and antiquark–antiquark pairs

$$\varphi_p = \frac{\Delta}{\mathcal{E}_p}, \quad \chi_p = \frac{\Delta}{\bar{\mathcal{E}}_p}.$$

Then, with a little juggling of (2), (3), we obtain the following set of coupled equations for $\varphi_p$ and $\chi_p$:

$$(\sqrt{p^2+m^2}-\mu)\varphi_p \qquad (4)$$
$$= \frac{4g}{\pi^2}(1-2n_p)\int dqq^2(\varphi_q+\chi_q),$$

$$(\sqrt{p^2+m^2}+\mu)\chi_p \qquad (5)$$
$$= \frac{4g}{\pi^2}(1-2\bar{n}_p)\int dqq^2(\chi_q+\varphi_q),$$

where

$$1-2n_p = \frac{E_p-\mu}{\mathcal{E}_p}, \quad 1-2\bar{n}_p = \frac{E_p+\mu}{\bar{\mathcal{E}}_p}. \qquad (6)$$

These two equations may be recasted into a single Klein–Gordon equation following the standard procedure [10]. Let us define $\psi_p = \varphi_p + \chi_p$ and consider the dilute limit $n_p \ll 1$, $\bar{n}_p \ll 1$.

One then gets

$$(p^2+m^2-\mu^2)\psi_p = \frac{8g}{\pi^2}\sqrt{p^2+m^2}\int dqq^2\psi_q. \quad (7)$$

We recall that nonlocality in (7) is of a symbolic character as soon as we use the pointlike four-fermion interaction.

From (4), (5), and (7), we see that, in the dilute strong-coupling limit, the chemical potential $\mu$ plays the role of the eigenvalue of the Klein–Gordon equation. The point $\mu = m - \mathcal{E}_B$ corresponds to a negative nonrelativistic chemical potential, typical for the "molecular" limit of the BCS–BEC crossover. The phase diagram in the $(n_p/\bar{n}_p, \mu)$ plane has two symmetric branches corresponding to quarks and antiquarks. Note also that the system described by Eqs. (4), (5), and (7) possesses an "exciton-like" instability.

I am grateful for discussions and remarks from N.O. Agasian, T.D. Lee, E.V. Shuryak, D.T. Son, M.A. Stephanov, and A.M. Tsvelik. Special thanks go to T.Yu. Matveeva for the help in preparing the article. Financial support from BNL and grant no. Ssc-1774-2003 is gratefully acknowledged. We thank INT (Seattle) for its hospitality and the Department of Energy for the support during the Workshop INT-04-1.

It is a pleasure and a honor to submit this paper to the Yurii Simonov Festschrift.

## REFERENCES

1. K. Rajagopal and F. Wilczek, in *At the Frontier of Particle Physics. Handbook of QCD*, Ed. by M. Shifman (World Sci., Singapore, 2001), Vol. 3, p. 2061; M. Alford, Annu. Rev. Nucl. Part. Sci. **51**, 131 (2001); G. Nardulli, Riv. Nuovo Cimento **25**, 1 (2002).
2. J. Berges and K. Rajagopal, Nucl. Phys. B **538**, 215 (1999).
3. D. M. Eagles, Phys. Rev. **186**, 456 (1969); A. J. Leggett, in *Modern Trends in the Theory of Condensed Matter* (Springer-Verlag, Berlin, 1980), p. 13.
4. P. Nozieres and S. Schmitt-Rink, J. Low Temp. Phys. **59**, 195 (1985).
5. P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin/Cummings Publ., Menlo Park, Cal., 1983), p. 105.
6. P. W. Anderson and W. F. Brinkman, in *Proceedings of the 15th Scottish University Summer School in Physics* (Acad. Press, New York, 1975), p. 315.
7. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
8. B. O. Kerbikov, hep-ph/0110197; Yad. Fiz. **65**, 1972 (2002) [Phys. At. Nucl. **65**, 1918 (2002)].
9. B. O. Kerbikov, hep-ph/0106324.
10. G. Baym, *Lectures on Quantum Mechanics* (W. A. Bejamin, New York, 1969), p. 54.

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 892–898. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 927–934.*


# Path Integrals and Degrees of Freedom in Many-Body Systems and Relativistic Field Theories*

**F. Palumbo****

*INFN, Laboratori Nazionali di Frascati, Italy*

Received June 2, 2004; in final form, October 29, 2004

**Abstract**—The identification of physical degrees of freedom is sometimes obscured in the path-integral formalism, and this makes it difficult to impose some constraints or to do some approximations. I review a number of cases where the difficulty is overcome by deriving the path integral from the operator form of the partition function after such identification has been made. © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

It is a great pleasure for me to contribute to this volume in honor of Prof. Yu.A. Simonov. This is an occasion for me to remember again the origin of our friendship, which has extended over the years well beyond what it would appear from our joint papers.

The subject of my contribution is focused on my recent interests in several problems which have a common feature: the identification of degrees of freedom in a path integral. Indeed, there are many situations in which such identification is helpful or necessary. An old example is the thermodynamics of gauge theories [1], but recently I met with many others. The first I will discuss here is how to find actions exactly equivalent to the standard ones but closer to the continuum at finite lattice spacing [2]. This includes the definition of the couplings of the chemical potential, an issue of particular importance in QCD [3]. The chemical potential is used to fix the expectation value of some charge operator. Fixing instead the value of the charge, namely, selecting a specific charge sector, is somewhat more difficult, but interesting pieces of information can be obtained by an expansion in a given charge sector of the fermion determinant in series of the number of fermions [4]. The last problem I will consider is how to describe the low-lying excitations of fermionic systems, both relativistic and nonrelativistic, by means of effective bosons, in short, how to bosonize them [5]. All these problems require the identification of the relevant degrees of freedom—in the last case, to determine the structure of the composite bosons in terms of the fermionic constituents; in the first case, to show the equivalence of different actions.

Now the path-integral formalism is widely used because of its flexibility and the possibility of numerical applications, but the identification of degrees of freedom is not always easy, while it can be conveniently achieved in the operator form of the partition function. Therefore, in all of the above problems, I will first identify the relevant degrees of freedom in the Fock space, where the partition function is defined; then I will introduce the appropriate constraints or approximations; and finally I will derive the constrained or approximated path integral.

As I said, all the subjects I mentioned have a common feature in the role played by the identification of degrees of freedom, but are otherwise very different. Therefore, the motivations for their investigation are given separately in the relevant sections. In Section 2, I will report the results I will use later about the standard derivation of the path integral from the operator form of the partition function. In Section 3, I will show how to derive an action different from the standard one and closer to the continuum in the nonrelativistic case. In Section 4, I will carry out the corresponding derivation for relativistic field theories, confining myself to the couplings of the chemical potential. In Section 5, I will discuss the case of a given charge sector. I will present a general method of bosonization valid for relativistic and nonrelativistic theories in Section 6 and my conclusions in Section 7.

## 2. THE STANDARD DERIVATION OF THE EUCLIDEAN PATH INTEGRAL FROM THE PARTITION FUNCTION

Let me introduce some definitions. I denote by $\tau$ the temporal lattice spacing, by $N_0$ the number of temporal sites, by $x_0$ or $t$ the temporal component of the site position vector $x$, by $T$ the temperature, by $\mu$

*This article was submitted by the author in English.

** e-mail: palumbof@lnf.infn.it

the chemical potential, by $\hat{Q}$ the (electric, baryon, ...) charge operator, and by $\hat{\mathcal{T}}(x_0)$ the transfer matrix.

In the nonrelativistic case, $\mathcal{T}$ is expressed in terms of the Hamiltonian,

$$\mathcal{T} = \exp(-\tau H), \tag{1}$$

and the Hamiltonian is the generator of continuous time translations. In relativistic theories, instead, only the transfer matrix is known in general, and the above equation can be used to define a Hamiltonian, but only as the generator of discrete translations by the time spacing $\tau$. Another important difference is that the nonrelativistic interactions are generally quartic in the fields, while the relativistic ones are quadratic. Both features contribute to make the derivation of the Euclidean path integral different in the two cases.

$\hat{\mathcal{T}}$ is defined in terms of particle–antiparticle creation–annihilation operators $\hat{c}^\dagger, \hat{d}^\dagger, \hat{c}, \hat{d}$ acting in a Fock space. It depends on the time coordinate $x_0$ only through the dependence on it of other fields (for instance, gauge fields). In fact, the creation and annihilation operators do not depend on $x_0$. They depend on the spatial coordinates $\mathbf{x}$ and on the internal quantum numbers (Dirac, color, and flavor indices in the case of QCD), comprehensively represented by $I, J, \ldots$.

In the transfer matrix formalism, often one has to deal with quantities at a given (Euclidean) time $x_0$. For this reason, we adopt a summation convention over spatial coordinates and intrinsic indices at fixed time. So, for instance, for an arbitrary matrix $M$, we will write

$$\hat{c}^\dagger M(x_0)\hat{c} = \sum_{\mathbf{x},\mathbf{y},I,J} \hat{c}^\dagger_{\mathbf{x},I} M_{\mathbf{x},I;\mathbf{y},J}(x_0)\hat{c}_{\mathbf{y},J}. \tag{2}$$

In this notation, the charge operator $\hat{Q}$ has the form

$$\hat{Q} = \hat{c}^\dagger\hat{c} - \hat{d}^\dagger\hat{d}. \tag{3}$$

The grand canonical partition function can be written as a time-ordered product

$$Z = \mathrm{Tr}\left\{\exp\left(\frac{\mu}{T}\hat{Q}\right)\prod_{x_0}\hat{\mathcal{T}}(x_0)\right\}, \tag{4}$$

which, using the relation $T^{-1} = \tau N_0$ and assuming the conservation of $\hat{Q}$, is conveniently rewritten as

$$Z = \mathrm{Tr}\left\{\prod_{x_0}\left[\hat{\mathcal{T}}(x_0)\exp\left(\mu\tau\hat{Q}\right)\right]\right\}. \tag{5}$$

The standard way [6, 7] to obtain the path-integral form of $Z$ is to write all the operators in normal order and introduce between the factors in Eq. (5) the identity

$$\mathcal{I} = \int [dc^+ dc\, dd^+ dd]\exp(-c^+c - d^+d)|cd\rangle\langle cd|, \tag{6}$$

where the basis vectors are coherent states

$$|cd\rangle = |\exp(-c\hat{c}^\dagger - d\hat{d}^\dagger)\rangle. \tag{7}$$

The $c^+$, $c$, $d^+$, $d$ are holomorphic/Grassmann variables and satisfy periodic/antiperiodic boundary conditions in time for bosons/fermions, respectively [6]. They have the label of the time slice, in which the identity operator is introduced. For the other indices, they are subject to the same convention as the creation and annihilation operators. The main property of coherent states is that they are eigenstates of the annihilation operators

$$\hat{c}|cd\rangle = c|cd\rangle. \tag{8}$$

To get the functional form of the partition function, it is only necessary to evaluate the matrix elements $\langle c_1 d_1|\mathcal{T}|cd\rangle$. This, as anticipated in the Introduction, must be done in different ways for nonrelativistic and relativistic theories.

The first case appears more difficult because of the quartic interactions, but since in general there are no ultraviolet divergencies (a detailed discussion of the departure from the standard form of the path integral in the presence of singular potentials can be found in [8]), we can make without any error the approximation

$$\exp(-\tau\hat{H}) \sim 1 - \tau\hat{H}. \tag{9}$$

Let us then consider a many-body system with the Hamiltonian written in normal form

$$\hat{H}(\hat{a}^\dagger, \hat{a}) = \sum_{\mathbf{x},\mathbf{y}}\left\{\hat{a}^\dagger_{\mathbf{x}} h_{\mathbf{x},\mathbf{y}}\hat{a}_{\mathbf{y}} + \hat{a}^\dagger_{\mathbf{x}}\hat{a}^\dagger_{\mathbf{y}} v_{\mathbf{x},\mathbf{y}}\hat{a}_{\mathbf{y}}\hat{a}_{\mathbf{x}}\right\}. \tag{10}$$

The standard expression of the Euclidean path integral associated to this Hamiltonian is

$$Z = \int [da^* da]\exp(-S), \tag{11}$$

where

$$S = \tau\sum_{t=1}^{N_0}\left\{a^*_{t+1}\nabla_t a_t + H(a^*_{t+1}, a_t)\right\} \tag{12}$$

is the action and I denoted by

$$\nabla_t f_t = \frac{1}{\tau}(f_{t+1} - f_t) \tag{13}$$

the right time discrete derivative. Notice the time splitting between the fields and their conjugates,

which implies a departure from the classical expression not only in derivative but also in potential terms. Needless to say, neglecting this time splitting introduces finite errors [6].

In the relativistic case, the transfer matrix can be written as [7]

$$\mathcal{T}_{x_0} = T^\dagger_{x_0} T_{x_0}, \tag{14}$$

where

$$T = \exp\left(\hat{c}^\dagger M \hat{c} + \hat{d}^\dagger M^T \hat{d}\right) \exp\left(\hat{c} N \hat{d}\right) \tag{15}$$

(here, the superscript $T$ means transposed and I do not need to specify the matrices $M$, $N$). The matrix elements of $\mathcal{T}$ can be evaluated exactly [7], yielding the standard form of the Euclidean path integral.

## 3. ANTINORMAL ORDERING: A DIFFERENT ACTION IN THE NONRELATIVISTIC CASE

In the standard form of the action, the time is split in the fields and their conjugates. This is an artifact which makes the equations unnecessarily different from continuous ones, somewhat more complicated, and somewhat confusing. In gauge theories, for instance, such time splitting introduces a coupling of the chemical potential to the temporal gauge fields which has been considered of physical significance, while it can be altogether avoided in a different derivation of the path integral. I will first illustrate such a derivation in the nonrelativistic case.

Instead of the normal order, I write the Hamiltonian in antinormal order (all the annihilation operators to the left of the creation ones)

$$\hat{H} = \sum_{\mathbf{x},\mathbf{y}} \left\{ h_0 + \hat{a}_\mathbf{x} h'_{\mathbf{x},\mathbf{y}} \hat{a}^\dagger_\mathbf{y} + \hat{a}_\mathbf{x} \hat{a}_\mathbf{y} v_{\mathbf{x},\mathbf{y}} \hat{a}^\dagger_\mathbf{y} \hat{a}^\dagger_\mathbf{x} \right\}, \tag{16}$$

where

$$h_0 = N^3 \left(-\sigma h_{\mathbf{x},\mathbf{x}} + w\right), \tag{17}$$
$$h'_{\mathbf{x},\mathbf{y}} = \sigma h_{\mathbf{x},\mathbf{y}} - w\delta_{\mathbf{x},\mathbf{y}}.$$

In the above equations,

$$w = \sigma v_{\mathbf{x},\mathbf{x}} + \sum_\mathbf{z} v_{\mathbf{x},\mathbf{z}}, \tag{18}$$

$N^3$ is the number of spatial sites, and $\sigma = -1$ for fermions and $+1$ for bosons. I assume that the potential is sufficiently regular for $w$ to exist, otherwise a regularization or a more drastic change in our procedure is needed. Then I expand the transfer matrix and in each term I insert the identity between the creation and annihilation operators. For the rightmost factor before taking the trace, I move the creation operators to the leftmost position. In this way, the term with the time derivative remains unchanged, but in the other terms, all the fields appear at the same time,

$$S' = \tau \sum_t \left\{ h_0 + a^*_{t+1} \nabla_t a_t + (h' - h) a^*_t a_t \right. \tag{19}$$
$$\left. + H(a^*_t, a_t) \right\}.$$

We can easily check on solvable models that this action gives the right results and that the terms arising from the rearrangement in antinormal ordering cannot be neglected.

In the case of fermions, the time splitting between the fields and their conjugates in the potential terms can be avoided in the same way as for bosons, but also in a simpler way. In fact, the Grassmann fields, unlike the holomorphic variables, are independent of their conjugates, so that the simple transformation

$$a^*_{t+1} \to a^*_t \tag{20}$$

eliminates the time splitting everywhere, with the obvious exception of the term with the time derivative, which is changed into the left one.

Because in this case we have two different derivations of the path integral, by their comparison, we can get nontrivial identities.

## 4. ANTINORMAL ORDERING: A DIFFERENT COUPLING OF THE CHEMICAL POTENTIAL IN RELATIVISTIC FIELD THEORIES

In relativistic field theories, the artificial time splitting affects only the coupling of the chemical potential. In the Hasenfratz–Karsch–Kogut formulation [9], which is the standard one, for Wilson fermions such coupling takes the form

$$\delta S = 2K \sum_x \overline{q} \left\{ \left[\exp\mu - 1\right] P^{(+)}_0 U_0 T^{(+)}_0 \right. \tag{21}$$
$$\left. + \left[\exp(-\mu) - 1\right] P^{(-)}_0 T^{(-)}_0 U^{(+)}_0 \right\} q,$$

where $q$ is the quark field, $K$ is the hopping parameter, $U_0$ the temporal link variable, and

$$P^{(\pm)}_0 = \frac{1}{2}\left(\mathbb{1} \pm \gamma_0\right), \tag{22}$$
$$T^{(\pm)}_0 f(x_0) = f(x_0 \pm 1).$$

Because of gauge invariance, in the presence of the time splitting, a coupling with the temporal links is needed, and this led to the conclusion that a nonvanishing contribution of the chemical potential must necessarily involve a Polyakov loop. But Creutz showed in a toy model [10] that this is not true, and I will report [4] in the sequel how to avoid such an artifact in the full-fledged QCD.

I write the exponential of the charge in the following way:

$$\exp(\mu a_0 \hat{Q}) = \int [dc^+ dc\, dd^+ dd] \times \exp(\delta S - c^+ c - d^+ d)|cd\rangle\langle cd|, \quad (23)$$

where

$$\delta S = (1 - \cosh(\mu a_0))\,(c^+ c + d^+ d) + \sinh(\mu a_0)(c^+ c - d^+ d). \quad (24)$$

The expression of $\delta S$ is obtained by expanding the exponential of the charge operator, putting all the terms in antinormal form, inserting in each monomial the unity between the set of annihilation and the set of creation operators, and replacing them by their Grassmannian eigenvalues. For the rightmost exponential of the charge, before taking the trace, one has to move the creation operators to the left of all the operators appearing under the trace.

After this, the construction of the path integral proceeds in the standard way, and we get the standard action with the exception of the coupling of the chemical potential, where all the fields appear at the same time:

$$\delta S = \sum_x \overline{q}\big[\,(1 - \cosh(\mu a_0)) + \sinh(\mu a_0)\gamma_0\big] Bq. \quad (25)$$

Here, I only need to say that the matrix $B$ does not depend on $U_0$. What is important is to notice that the quark field and its conjugate are at the same time, and then the temporal Wilson variable disappears.

## 5. EXPANSION OF THE FERMION DETERMINANT IN THE NUMBER OF FERMIONS IN A GIVEN CHARGE SECTOR

The use of the chemical potential is a way to impose a given expectation value for some conserved charge. The alternative option of selecting a given sector of the charge in the path integral presents additional difficulties, but something can be learned by a series expansion of the fermion determinant [5]. To be concrete, we will refer to the case of QCD, but the method can be applied to other cases with appropriate modifications.

In the absence of any condition on the baryon number, the quark determinant is

$$\det Q = \int [d\overline{q}dq] \exp S_q, \quad (26)$$

where $S_q$ is the quark action and $Q$ is the quark matrix. My strategy is to write $\det Q$ as the trace of the transfer matrix acting in the quark Fock space, impose the restriction to a given baryonic sector, and then rewrite the trace as the determinant of a modified quark matrix. The round-trip is done by mapping the Grassmann algebra generated by the quark fields into the Fock space following the construction of Lüscher [7]. But while his paper is based on the mere existence of the map, in order to enforce the projection, I will make use of a concrete realization by means of coherent states.

The first step is then to write the unconstrained determinant as a trace in the Fock space,

$$\det Q = \mathrm{Tr}\,\hat{\mathcal{T}}. \quad (27)$$

The second step is to impose the restriction to a sector with baryon number $n_B$ by inserting in the trace the appropriate projection operator $\hat{P}_{n_B}$,

$$\det Q|_{n_B} = \mathrm{Tr}\left(\hat{\mathcal{T}}\hat{P}_{n_B}\right) \int [dx^+ dx\, dy^+ dy] \times \exp\left(-x^+ x - y^+ y\right) \langle x, y|\hat{\mathcal{T}}\hat{P}_{n_B}|-x, -y\rangle, \quad (28)$$

which will be expressed in terms of the determinant of a modified quark matrix. The kernel $\langle x, y|\hat{\mathcal{T}}\hat{P}_{n_B}|-x, -y\rangle$ has the integral form

$$\langle x, y|\hat{\mathcal{T}}\hat{P}_{n_B}|-x, -y\rangle = \int [dz^+ dz dw^+ dw] \times \exp\left(-z^+ z - w^+ w\right) \langle x, y|\hat{\mathcal{T}}|z, w\rangle \times \langle z, w|\hat{P}_{n_B}|-x, -y\rangle. \quad (29)$$

The expression of the kernel $\langle x, y|\hat{\mathcal{T}}|z, w\rangle$ will not be reported here, while that of $\hat{P}_{n_B}$ can easily be derived,

$$\langle z, w|\hat{P}_{n_B}|-x, -y\rangle = \sum_{r=0}^{\infty}(-1)^{n_B} \times \langle(\hat{y}w^+)^r(\hat{x}z^+)^{(n_B+r)}(x\hat{x}^+)^{n_B+r}(y\hat{y}^+)^r\rangle \times \frac{1}{((n_B + r)!r!)^2}. \quad (30)$$

Since $\hat{x}^+$ and $\hat{y}^+$ are creation operators of quarks and antiquarks, respectively, we see that the $r$th term of this series gives the gauge-invariant contribution of $n_B$ valence quarks plus $r$ quark–antiquark pairs.

Needless to say, for $n_B = 0$, $\det Q|_{n_B}$ does not reduce to the unconstrained determinant: Indeed, also baryonic states are present in the unconstrained determinant, while they are absent in $\det Q|_{n_B=0}$. In QCD at nonvanishing temperature, it makes a difference whether we impose or not the condition $n_B = 0$. In view of the relatively low value of the critical temperature with respect to the nucleon mass, however, we do not expect significant effects from the

restriction to a given baryon sector unless we go to exceedingly high temperatures.

Let me now proceed to derive our final result. By evaluating the vacuum expectation values appearing in the last equation, we express the kernel of the projection operator in terms of Grassmann variables only,

$$\langle z,w|\hat{P}_{n_B}|-x,-y\rangle \tag{31}$$
$$=\sum_{r=0}^{\infty}(-1)^{n_B}\frac{1}{(n_B+r)!r!}(z^+x)^{n_B+r}(w^+y)^r.$$

To evaluate the integral of Eq. (29), I rewrite the above equation in exponential form

$$\langle z,w|\hat{P}_{n_B}|x,y\rangle=\sum_{r=0}^{\infty}\frac{1}{(n_B+r)!r!} \tag{32}$$
$$\times\frac{\partial^{n_B+r}}{\partial j_1^{n_B+r}}\frac{\partial^r}{\partial j_2^r}\exp(-j_1z^+x-j_2w^+y)|_{j_1=j_2=0}.$$

The integrals of Eqs. (28) and (29) are Gaussian and we get the constrained determinant in terms of the determinant of a modified quark matrix

$$\det Q|_{n_B}=\sum_{r=0}^{\infty}\frac{1}{(n_B+r)!r!}\frac{\partial^{n_B+r}}{\partial j_1^{n_B+r}} \tag{33}$$
$$\times\frac{\partial^r}{\partial j_2^r}\det\left(Q+\delta Q_1+\delta Q_2\right)|_{j_1=j_2=0}.$$

The explicit form of the variations $\delta Q_1$ and $\delta Q_2$ of the quark matrix is not important here, but we warn the reader that there is an error in their expression in [14].

## 6. BOSONIZATION IN MANY-BODY SYSTEMS AND RELATIVISTIC FIELD THEORIES

The low-energy collective excitations of many-fermion systems can be described by effective bosons. Well-known examples are the Cooper pairs of superconductivity, the bosons of the interacting boson model of nuclear physics, the chiral mesons, and the quark pairs of color superconductivity in QCD. In all these cases, the effective bosons are generated by attractive interactions, but effective bosons can arise also in the presence of repulsive forces, as in the Hubbard model [11]. Some of the effective bosons are Goldstone bosons, and then there is a general theory which tells that they live in the coset space of the group which is spontaneously broken and dictates how they are related to the original fields [12]. But there is no general procedure to reformulate the fermionic theory in terms of the effective bosonic degrees of freedom, even though there are several recipes for specific cases which are reviewed in [13]. A more flexible approach is based on the Hubbard–Stratonovich transformation which linearizes the fermionic interaction by introducing bosonic auxiliary fields which are then promoted to physical life. The typical resulting structure is that of chiral theories [14]. But in such an approach, an energy scale emerges naturally, and only excitations of lower energy can be described by the auxiliary fields. Moreover, in renormalizable relativistic field theories like QCD, the fermion Lagrangian is quadratic to start with, so that the Hubbard–Stratonovich transformation cannot be used. One can add quartic interactions as irrelevant operators, and this can help in numerical simulations, but has not led so far to a formulation of low-energy QCD in terms of chiral mesons.

I present a new approach [5] to bosonization which does not suffer from the above limitations and can be applied to theories with quartic and quadratic interactions as well. It is based on the evaluation of the partition function restricted to the bosonic composites of interest. By rewriting the partition function so obtained in functional form, we get the Euclidean action of the composite bosons from which in the nonrelativistic theories we can derive the Hamiltonian. In this way, I derived the interacting boson model from a nuclear Hamiltonian. In the case of pure pairing, I reproduce the well-known results for the excitations corresponding to the addition and removal of pairs of fermions, as well as for the seniority excitations which are inaccessible by the Hubbard–Stratonovich method. Indeed, at least in this example, this theory does not have the structure of a chiral expansion.

For the relativistic case, an investigation is in progress [15].

Let me start by defining the composites in terms of the fermion operators $\hat{c}$:

$$\hat{b}_J^\dagger=\frac{1}{2}\hat{c}^\dagger B_J^\dagger\hat{c}^\dagger=\frac{1}{2}\sum_{m_1,m_2}\hat{c}_{m_1}^\dagger\left(B_J^\dagger\right)_{m_1,m_2}\hat{c}_{m_2}^\dagger. \tag{34}$$

In the above equation, $m$ represents all the fermion intrinsic quantum numbers and position coordinates and $J$ represents the quantum numbers of the composites. I assume all the structure matrices $B_J$ to have one and the same dimension, which I denote by $2\Omega$. The fermionic operators have canonical anticommutation relations, while for the composites

$$[\hat{b}_{J_1},\hat{b}_{J_2}^\dagger]=\frac{1}{2}\mathrm{Tr}(B_{J_1}B_{J_2}^\dagger)-\hat{c}^\dagger B_{J_2}^\dagger B_{J_1}\hat{c}. \tag{35}$$

It is then natural to require the normalization

$$\mathrm{Tr}(B_{J_1}^\dagger,B_{J_2})=2\delta_{J_1,J_2}. \tag{36}$$

A convenient way to get the Euclidean path integral from the trace of the transfer matrix is to use

coherent states of composites. Therefore, I introduce the operator

$$\mathcal{P} = \int db^* db (\langle b|b\rangle)^{-1} |b\rangle\langle b|, \tag{37}$$

where

$$|b\rangle = |\exp(b\cdot \hat{b}^\dagger)\rangle. \tag{38}$$

I adopted the convention

$$b\cdot\hat{b}^\dagger = \sum_J b_J \hat{b}_J^\dagger. \tag{39}$$

If $\hat{b}$ were operators of elementary bosons, $\mathcal{P}$ would be the identity in the boson Fock space. I would like $\mathcal{P}$ to be the identity in the fermion subspace of the composites. To see the action of $\mathcal{P}$ on composite operators, let us first consider the case where there is only one composite with structure function satisfying the equation

$$B^\dagger B = \frac{1}{\Omega}\mathbb{1}. \tag{40}$$

Then we find

$$\langle b|b\rangle^{-1} = \left(1 + \frac{1}{\Omega} b_1^* b\right)^{-\Omega} \tag{41}$$

and

$$\langle b_1|(\hat{b}^\dagger)^n\rangle = C_n (b^*)^n, \tag{42}$$

where

$$C_n = \frac{\Omega!}{(\Omega-n)!\Omega^n} \tag{43}$$
$$= \left(1-\frac{1}{\Omega}\right)\left(1-\frac{2}{\Omega}\right)\cdots\left(1-\frac{n-1}{\Omega}\right).$$

Now we can determine the action of $\mathcal{P}$ on the composites

$$\mathcal{P}|(\hat{b}^\dagger)^n\rangle = \left(1-\frac{n}{\Omega}\right)^{-1}\left(1-\frac{n+1}{\Omega}\right)^{-1}|(\hat{b}^\dagger)^n\rangle, \tag{44}$$

which shows that $\mathcal{P}$ behaves approximately like the identity with an error of the order of $n/\Omega$. It is perhaps worthwhile noticing that, in the limit of infinite $\Omega$, we recover exactly the expressions valid for elementary bosons, in particular,

$$\langle b_1|b\rangle = \left(1+\frac{1}{\Omega}b_1^* b\right)^{\Omega} \to \exp(b_1^* b), \quad \Omega\to\infty. \tag{45}$$

It might appear that the treatment of states with $n \sim \Omega$ is precluded, but this is not true. Indeed, if we are interested in states with $n = \overline{n} + \nu$ for an arbitrary reference state $\overline{n}$, we redefine $\mathcal{P}$ according to

$$\mathcal{P}_{\overline{n}} = \frac{(\Omega-\overline{n})^2}{\Omega^2}\mathcal{P}_0. \tag{46}$$

We then have

$$\mathcal{P}_{\overline{n}}|(\hat{b}^\dagger)^n\rangle = \left(1-\frac{\nu}{\Omega-\overline{n}}\right)^{-1} \times \left(1-\frac{\nu+1}{\Omega-\overline{n}}\right)^{-1}|(\hat{b}^\dagger)^n\rangle, \tag{47}$$

which shows that $\mathcal{P}_{\overline{n}}$ behaves like the identity in the neighborhood of the reference state up to an error of order $\nu/(\Omega-\overline{n})$, namely, the measure $\langle b|b\rangle^{-1}$ is essentially uniform.

In the general case of many composites, we have

$$\langle b_1|b\rangle = \left[\det\left(\mathbb{1} + \beta_1\beta\right)\right]^{1/2}, \tag{48}$$

where the matrix $\beta$ is

$$\beta = b\cdot B^\dagger \tag{49}$$

and

$$\langle b_1|(\hat{b}_{I_0}^\dagger)^{n_0}\ldots(\hat{b}_{I_i}^\dagger)^{n_i}\rangle = \frac{\partial^{n_0}}{\partial x_0^{n_0}}\cdots\frac{\partial^{n_i}}{\partial x_i^{n_i}} \times \exp\left\{\frac{1}{2}\mathrm{Tr}\ln[\mathbb{1} + (x\cdot B^\dagger)(b_1^*\cdot B)]\right\}\Big|_{x=0}. \tag{50}$$

We must now make an assumption which replaces Eq. (40), namely, that all the eigenvalues of the matrices $B_J^\dagger B_J$ are much smaller than $\Omega$. Then we find again that $\mathcal{P}$ approximates the identity with an error of order $1/\Omega$.

Now we are equipped to carry out the program outlined at the beginning. The first step is the evaluation of the partition function $Z_C$ restricted to fermionic composites. To this end, we divide the inverse temperature in $N_0$ intervals of spacing $\tau$,

$$\frac{1}{T} = N_0\tau, \tag{51}$$

and write

$$Z_c = \mathrm{Tr}\left(\mathcal{P}\mathcal{T}\right)^{N_0}, \tag{52}$$

where $\mathcal{T}$ is the transfer matrix. In the nonrelativistic case, $\mathcal{T}$ is expressed in terms of the Hamiltonian

$$\mathcal{T} = \exp\left(-\tau\hat{H}\right), \tag{53}$$

and the Hamiltonian is the generator of continuous time translations. In relativistic field theories, instead, only the transfer matrix is known in general, and the above equation can be used to define a Hamiltonian, but only as the generator of discrete translations by the time spacing $\tau$.

At this point, we must evaluate the matrix element $\langle b_1|\mathcal{T}|b\rangle$ and to do this we must distinguish between relativistic field theories and many-body systems. In the first case, the transfer matrix is a product of exponentials of quadratic forms in the fermion operators [7], and the matrix element can be directly and exactly evaluated [15]. In the second case, one must at an intermediate stage expand with respect to the time spacing $\tau$. This does not introduce any error because one can retain all the terms which give finite contributions in the limit $\tau \to 0$. We will report here only the nonrelativistic calculation. The most general Hamiltonian can be written

$$\hat{H} = \hat{c}^\dagger h_0 \hat{c} - \sum_K g_K \frac{1}{2}\hat{c}^\dagger F_K^\dagger \hat{c}^\dagger \frac{1}{2}\hat{c} F_K \hat{c}, \qquad (54)$$

where $K$ represents all the necessary quantum numbers. The single-particle term includes the single-particle energy with matrix $e$, any single-particle interaction with external fields described by the matrix $\mathcal{M}$, and the chemical potential $\mu$,

$$h_0 = e + \mathcal{M} - \mu. \qquad (55)$$

Therefore, we will be able to solve the problem of fermion–boson mapping by determining the interaction of the composite bosons with external fields. Assuming for the potential form factors the normalization

$$\mathrm{Tr}(F_K^\dagger F_K) = 2\Omega \qquad (56)$$

and setting

$$\Gamma_t = (\mathbb{1} + \beta_t^* \beta_{t-1})^{-1}, \qquad (57)$$

we get the Euclidean action

$$S = \tau \sum_t \left\{ \frac{1}{2\tau}\mathrm{Tr}[\ln(\mathbb{1} + \beta_t^*\beta_t) - \ln\Gamma_t] \right. \qquad (58)$$

$$- H_1 + \frac{1}{4}\sum_K g_K \mathrm{Tr}\left[\mathrm{Tr}(\Gamma_t \beta_t^* F_K^\dagger)\mathrm{Tr}(\Gamma_t F_K \beta_{t-1})\right.$$

$$\left. - 2\mathrm{Tr}\left(\Gamma_t F_K^\dagger F_K\right) - \mathrm{Tr}[\Gamma_t \beta_t^* F_K^\dagger, \Gamma_t F_K \beta_{t-1}]_+\right]$$

$$\left. + \frac{1}{2}\mathrm{Tr}\left[\beta_t^*(\beta_{t-1}h^T + h\beta_{t-1})\right]\right\},$$

where $[\ldots,\ldots]_+$ is an anticommutator. This action differs from that of elementary bosons because

(i) the time derivative terms (contained in the first line) are noncanonical;

(ii) the coupling of the chemical potential (which appears in $h$) is also noncanonical, since it is not quadratic in the boson fields;

(iii) the function $\Gamma$ becomes singular when the number of bosons is of order $\Omega$, which reflects the Pauli principle.

We remind the reader that the only approximation done concerns the operator $\mathcal{P}$. Therefore, these are to be regarded as true features of compositeness.

The bosonization of the system that we considered has thus been accomplished. In particular, the fermionic interactions with external fields can be expressed in terms of the bosonic terms which involve the matrix $\mathcal{M}$ (appearing in $h$), and the dynamical problem of the interacting (composite) bosons can be solved within the path-integral formalism. Part of this problem is the determination of the structure matrices $B_J$. This can be done by expressing the energies in terms of them and applying a variational principle which gives rise to an eigenvalue equation.

The Hamiltonian can be derived by standard procedures.

## 7. CONCLUSION

I showed that there are a number of problems which can be easily dealt with in the operator form of the partition function. Only afterwards can it be given the functional form of the path integral which is more convenient for many purposes.

There are other examples that I left out for different reasons. Among these, I would like to mention the problem of the restriction of gauge theories to physical states, which has not yet found a general and satisfactory solution.

## REFERENCES


1. L. Dolan and R. Jackiw, Phys. Rev. D **9**, 3320 (1974).
2. F. Palumbo (in press).
3. F. Palumbo, Phys. Rev. D **69**, 074508 (2004).
4. F. Palumbo, Nucl. Phys. B **643**, 391 (2002); hep-lat/0202021.
5. F. Palumbo, nucl-th/0405045.
6. J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, New York, 1988).
7. M. Lüscher, Commun. Math. Phys. **54**, 283 (1977).
8. H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics and Polymer Physics* (World Sci., Singapore, 1990).
9. P. Hasenfratz and F. Karsch, Phys. Lett. B **125B**, 308 (1983); J. Kogut, M. Matsuoka, M. Stone, *et al.*, Nucl. Phys. B **225**, 93 (1983).
10. M. Creutz, Found. Phys. **30**, 487 (2000).
11. M. Cini and G. Stefanucci, cond-mat/0204311 v. 1.
12. S. Weinberg, *The Quantum Theory of Fields* (Cambridge Univ. Press, Cambridge, 1996).
13. A. Klein and E. R. Marshalek, Rev. Mod. Phys. **63**, 375 (1991).
14. V. A. Miransky, *Dynamical Symmetry Breaking in Quantum Field Theories* (World Sci., Singapore, 1993).
15. S. Caracciolo and F. Palumbo (in press).

*Physics of Atomic Nuclei, Vol. 68, No. 5, 2005, pp. 899–908. From Yadernaya Fizika, Vol. 68, No. 5, 2005, pp. 935–944.*


# Isochronous PDEs*

**M. Mariani[1] and F. Calogero[2]**

Received May 13, 2004

**Abstract**—A number of well-known evolution PDEs are modified so that they then possess many solutions which are isochronous, i.e., completely periodic, with a fixed period that does not depend on the initial data (for large sets of such data). © *2005 Pleiades Publishing, Inc.*

## 1. INTRODUCTION

Recently, a simple "trick"—amounting essentially to a change of independent and dependent variables—was introduced [1–3] that associates to a given evolution equation a family of deformed equations parametrized by a real constant $\omega$ to which (for positive $\omega$) the period

$$T = \frac{2\pi}{\omega} \tag{1.1}$$

is associated. For vanishing $\omega$, $\omega = 0$, the original equation is recovered; for positive $\omega$, $\omega > 0$, to which attention is hereafter confined without loss of generality, these "$\omega$-modified" equations possess many solutions which are isochronous—namely, completely periodic with period $T$ or possibly with a period that is a simple multiple or fraction of $T$. In the context of the initial-value problem, these solutions generally emerge out of open sets of initial data having the full dimensionality of the phase space underlining the problem at hand. In several recent papers, this phenomenology has been explored and exploited, or just mentioned, in an ODE context (mainly, though not exclusively, in a many-body problem context) [1–21] and in a few recent papers in a PDE context [3, 17, 21–23]. The purpose and scope of this paper is to apply the trick to a number of well-known, autonomous, evolution PDEs which thereby yield, also autonomous (at least as regards the dependence on the time variable), $\omega$-modified equations that, we believe, deserve to be exhibited and—at least some of them—shall warrant further investigation and are likely to become useful theoretical tools, in view of the remarkable phenomenology featured by their solutions. The following section is devoted to certain, mainly notational, preliminaries, the main purpose of which is to minimize subsequent repetitions. The nonlinear evolution PDEs are then listed, with minimal comments, in Section 3, both in their unmodified avatars (with appropriate—if inevitably incomplete—references) and in their $\omega$-modified versions. Occasionally, some solutions are also reported and tersely discussed. The last section contains some final comments.

## 2. NOTATION AND PRELIMINARIES

The independent variables of the unmodified evolution PDE are denoted as $\underline{\xi} \equiv (\xi_1, \ldots, \xi_N)$ and $\tau$; the dependent variable of the unmodified evolution PDE is denoted as $w\left(\underline{\xi};\tau\right) \equiv w\left(\xi_1, \ldots, \xi_N; \tau\right)$, and if a second dependent variable also enters, it is denoted as $\tilde{w}\left(\underline{\xi};\tau\right) \equiv \tilde{w}\left(\xi_1, \ldots, \xi_N; \tau\right)$; upper case letters, $W\left(\underline{\xi};\tau\right) \equiv W\left(\xi_1, \ldots, \xi_N; \tau\right)$, $\tilde{W}\left(\underline{\xi};\tau\right) \equiv \tilde{W}\left(\xi_1, \ldots, \xi_N; \tau\right)$, are used for matrices. The independent variables of the $\omega$-modified evolution PDE are denoted as $\underline{x} \equiv (x_1, \ldots, x_N)$ and $t$; the dependent variable of the $\omega$-modified evolution PDE is denoted as $u\left(\underline{x};t\right) \equiv u\left(x_1, \ldots, x_N; t\right)$, and if a second dependent variable also enters, it is denoted as $\tilde{u}\left(\underline{x};t\right) \equiv \tilde{u}\left(x_1, \ldots, x_N; t\right)$; and again upper case letters, $U\left(\underline{x};t\right) \equiv U\left(x_1, \ldots, x_N; t\right)$, $\tilde{U}\left(\underline{x};t\right) \equiv \tilde{U}\left(x_1, \ldots, x_N; t\right)$, are used for matrices. The relations among the (independent and dependent) variables of the unmodified evolution PDE and the $\omega$-modified evolution PDE are given by the following formulas (the trick):

$$\tau = \frac{\exp\left(i\omega t\right) - 1}{i\omega}, \tag{2.1}$$

$$\xi_n = \xi_n(t) = x_n \exp\left(i\mu_n \omega t\right), \quad n = 1, \ldots, N, \tag{2.2}$$

*This article was submitted by the authors in English.

[1]Dottorato in Matematica, Università di Roma "La Sapienza," Roma, Italy; e-mail: mariani@mat.uniroma1.it

[2]Dipartimento di Fisica, Università di Roma "La Sapienza," and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy; e-mail: francesco.calogero@roma1.infn.it

$$u\left(\underline{\xi};t\right)=\exp\left(i\lambda\omega t\right)w\left(\underline{\xi};\tau\right), \tag{2.3a}$$

$$\tilde{u}\left(\underline{\xi};t\right)=\exp\left(i\tilde{\lambda}\omega t\right)\tilde{w}\left(\underline{\xi};\tau\right), \tag{2.3b}$$

with analogous formulas [see (2.3)] in the matrix case. Here and hereafter, constants such as $\mu_n$, $\lambda$, $\tilde{\lambda}$, $\alpha$, $\beta$ (Greek letters) denote rational numbers (not necessarily positive), which whenever necessary shall be properly assigned, while Latin letters such as $a$, $b$, $c$ denote complex (or, as the case may be, real) constants (sometimes, we keep such constants even when they could be eliminated by trivial rescaling transformations; and, of course, by such transformations, additional such constants might instead be introduced). When $N=1$, we drop the index $n$; namely, we write $\xi$ instead of $\xi_1$, $x$ instead of $x_1$, and $\mu$ instead of $\mu_1$; and for $N=2$, we also, to simplify the notation, write $\eta$ instead of $\xi_2$, $y$ instead of $x_2$, and $\nu$ instead of $\mu_2$. Note that this transformation (2.1)–(2.3) entails that, at the "initial" time, $\tau=t=0$, the change of variables disappears altogether:

$$\underline{\xi}(0)=\underline{x},\qquad u\left(\underline{x};0\right)=w\left(\underline{\xi};0\right), \tag{2.4}$$
$$\tilde{u}\left(\underline{\xi};0\right)=\tilde{w}\left(\underline{\xi};0\right).$$

Hereafter, subscripted variables denote partial differentiations, $w_\tau\equiv\partial w\left(\underline{\xi};\tau\right)/\partial\tau$, $u_{x_n}\equiv\partial u\left(\underline{x};t\right)/\partial x_n$, and so on. Let us emphasize—obvious as this may be—that, since the transition from an (unmodified) PDE satisfied by $w\left(\underline{\xi};\tau\right)$ to the corresponding ($\omega$-modified) PDE satisfied by $u\left(\underline{x};t\right)$ is performed via the explicit change of variables (2.1)–(2.3) (the trick), properties such as integrability or solvability, if possessed by the unmodified PDE satisfied by $w\left(\underline{\xi};\tau\right)$, carry over to the corresponding ($\omega$-modified) PDE satisfied by $u\left(\underline{x};t\right)$, which generally has in addition the property of isochronicity, as defined above. Let us moreover note that the property of isochronicity of the $\omega$-modified evolution PDE, which does not require that the original unmodified PDE from which it has emerged be itself integrable, implies that, in some open set of its phase space, the $\omega$-modified equation is generally integrable indeed, in some sense, superintegrable (for a discussion of this question in the ODE context, where the concepts of integrability and indeed superintegrability can be given a more precise meaning, see [17] and [19]). Let us end this section by pointing out that, in most cases, the $\omega$-modified equations are complex; they can of course be rewritten in real form by introducing the real and imaginary parts (or, instead, the amplitudes and phases) of all the quantities that enter in these PDEs and by then considering the two, generally coupled, real PDEs that are obtained from each complex PDE by considering separately its real and imaginary parts; below, in a few cases, we also exhibit the system of real evolution PDEs obtained in this manner.

## 3. ISOCHRONOUS PDEs

In this section, we display, with minimal commentary, a list of nonlinear evolution PDEs, each of them first in its unmodified avatar, then in its $\omega$-modified version. It is remarkable that so many well-known autonomous evolution PDEs possess $\omega$-modified versions which are as well autonomous, at least as regards their time dependence; in several cases, this appears to be due to some minor miracles, inasmuch as the number of relevant parameters $\lambda$, $\mu$, $\nu$, $\tilde{\lambda}$, ... is smaller than the equations they are required to satisfy in order to guarantee the autonomous character of the $\omega$-modified equations, yet nontrivial parameters satisfying these conditions do exist. All the $\omega$-modified evolution PDEs displayed below possess many isochronous solutions, but only in some (rare) cases do we exhibit below examples of these solutions. A discussion of each of these nonlinear evolution PDEs would indeed require much more space. Let us also emphasize that the list reported below includes only some kind of representative instances of this phenomenology; obviously, many more examples could be added. The list is arranged in a user-friendly manner, being ordered according to the following taxonomic rules: of primary importance is the number of independent variables; next, the number of dependent variables; next, the order of the differential equation, but with primary attention to the time variable; then, the type of nonlinearity [except when an equation is presented as a special case of a more general equation—see, for instance, (3.9) and (3.34)].

The following unmodified $(1+1)$-dimensional "generalized shock-type" PDE is integrable, indeed solvable (see, for instance, [24]):

$$w_\tau=aw^\alpha w_\xi. \tag{3.1a}$$

By setting $\mu=1-\alpha\lambda$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t-i\lambda\omega u+i\left(\alpha\lambda-1\right)\omega xu_x=au^\alpha u_x. \tag{3.1b}$$

The general solution to the initial-value problem for this PDE (3.1b) is given, in implicit form, by the following formula:

$$u(x;t)=e^{i\lambda\omega t}u_0\Bigg(\exp\left[i(1-\alpha\lambda)\omega t\right] \times\left\{x+a\frac{1-e^{-i\omega t}}{i\omega}\left[u(x;t)\right]^\alpha\right\}\Bigg), \tag{3.1c}$$

where of course $u_0(x)=u(x;0)$.

The following unmodified $(1+1)$-dimensional "generalized Burgers–Hopf" PDE (see [25]) reads

$$w_\tau = aww_\xi + b\left(w^\alpha w_\xi\right)_\xi . \tag{3.2a}$$

By setting $\lambda = 1/(2-\alpha)$ and $\mu = (1-\alpha)/(2-\alpha)$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t + i\frac{1}{\alpha-2}\omega u + i\frac{1-\alpha}{\alpha-2}\omega x u_x = auu_x + b\left(u^\alpha u_x\right)_x . \tag{3.2b}$$

Note that, for $\alpha = 1$, this PDE becomes autonomous also with respect to the space variable $x$, while for $\alpha = 0$ the PDE (3.2a) becomes the standard (solvable) Burgers–Hopf equation.

The following unmodified $(1+1)$-dimensional dispersive KdV-like PDE is integrable, indeed solvable (see, for instance, [26]):

$$w_\tau = w_{\xi\xi\xi} + 3\left[w_{\xi\xi}w^2 + 3w\left(w_\xi\right)^2\right] + 3w_\xi w^4. \tag{3.3a}$$

By setting $\lambda = 1/6$, $\mu = 1/3$, and (for notational simplicity) $\Omega = \omega/6$, one gets the corresponding $\Omega$-modified evolution PDE

$$u_t - i\Omega u - 2i\Omega x u_x = u_{xxx} + 3\left(u_{xx}u^2 + 3uu_x^2\right) + 3u_x u^4. \tag{3.3b}$$

The symmetry properties, and some explicit solutions, of the following unmodified $(1+1)$-dimensional "generalized KdV" equation have been investigated recently [27]:

$$w_\tau = a\left(w^\alpha\right)_{\xi\xi\xi} + b\left(w^\beta\right)_\xi . \tag{3.4a}$$

By setting $\lambda = 2/(3\beta-\alpha-2)$ and $\mu = (\beta-\alpha)/(3\beta-\alpha-2)$, one gets the $\omega$-modified evolution PDE

$$u_t - \frac{2i\omega}{3\beta-\alpha-2}u - \frac{i(\beta-\alpha)\omega}{3\beta-\alpha-2}xu_x = a\left(u^\alpha\right)_{xxx} + b\left(u^\beta\right)_x . \tag{3.4b}$$

We assume here of course that $3\beta-\alpha-2 \neq 0$. Particularly interesting is the case with $\alpha = \beta$, when this $\omega$-modified PDE becomes autonomous also in the space variable $x$. In the even more special case with $\alpha = \beta = 2$ and by setting $u = u_1 + iu_2$, $a = c_1 + ic_2$, and $b = c_3 + ic_4$, we rewrite this evolution PDE as a system of two coupled PDEs:

$$\begin{aligned} u_{1_t} + \omega u_2 &= \left[c_1\left(u_1^2 - u_2^2\right) - 2c_2u_1u_2\right]_x \\ &\quad + \left[c_3\left(u_1^2 - u_2^2\right) - 2c_4u_1u_2\right]_{xxx}, \\ u_{2_t} - \omega u_1 &= \left[c_2\left(u_1^2 - u_2^2\right) + 2c_1u_1u_2\right]_x \\ &\quad + \left[c_4\left(u_1^2 - u_2^2\right) + 2c_3u_1u_2\right]_{xxx}. \end{aligned} \tag{3.4c}$$

Here, we assume of course that the two dependent variables $u_1 \equiv u_1(x,t)$ and $u_2 \equiv u_2(x,t)$ are real and that the four arbitrary constants $c_1$, $c_2$, $c_3$, and $c_4$ are real as well.

The following unmodified $(1+1)$-dimensional "Schwarzian KdV" equation is integrable (see, for instance, [28]):

$$w_\tau = w_{\xi\xi\xi} + a\frac{w_{\xi\xi}^2}{w_\xi}. \tag{3.5a}$$

By setting $\mu = 1/3$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t - i\lambda\omega u - i\frac{\omega}{3}xu_x = u_{xxx} + a\frac{u_{xx}^2}{w_x}. \tag{3.5b}$$

The following unmodified $(1+1)$-dimensional "Cavalcante–Tenenblat" equation is integrable (see, for instance, [29]):

$$w_\tau = a\left(\frac{1}{\sqrt{w_\xi}}\right)_{\xi\xi} + b\left(w_\xi\right)^{3/2}. \tag{3.6a}$$

By setting $\lambda = 0$, $\mu = 2/3$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t - i\frac{2\omega}{3}xu_x = a\left(\frac{1}{\sqrt{u_x}}\right)_{xx} + b\left(u_x\right)^{3/2}. \tag{3.6b}$$

The KdV class of unmodified $(1+1)$-dimensional integrable evolution PDEs (see, for instance, [30]) reads

$$w_\tau = \Lambda^m w_\xi, \quad m = 1, 2, \ldots, \tag{3.7a}$$

where $\Lambda$ is the integro-differential operator (depending on the dependent variable $w(\xi;\tau)$) that acts on a generic (twice-differentiable and integrable at infinity) function $\phi(\xi)$ as follows:

$$\Lambda\phi(\xi) = \phi_{\xi\xi}(\xi) - 4w(\xi;\tau)\phi(\xi) + 2w_\xi(\xi;\tau)\int_\xi^\infty \phi\left(\xi'\right)d\xi'. \tag{3.7b}$$

By setting $\lambda = 2/(2m+1)$, $\mu = 1/(2m+1)$, and (for notational simplicity) $\Omega_m = \omega/(2m+1)$, one gets the corresponding class of $\Omega_m$-modified evolution PDEs

$$u_t - i\Omega_m\left(2u + xu_x\right) = L^m u_x, \quad m = 1, 2, \ldots, \tag{3.7c}$$

where $L$ is the integro-differential operator (depending on the dependent variable $u(x;t)$) analogous to

$\Lambda$, namely, the operator that acts on a generic (twice-differentiable and integrable at infinity) function $f(x)$ as follows:

$$Lf(x) = f_{xx}(x) - 4u(x;t)f(x) + 2u_x(x;t)\int_x^\infty f(x')\,dx'. \tag{3.7d}$$

For $m = 1$, the PDE (3.7a) becomes the well-known KdV equation

$$w_\tau + w_{\xi\xi\xi} = 6ww_\xi, \tag{3.8a}$$

and the corresponding $\omega$-modified equation reads

$$u_t + u_{xxx} - i\frac{\omega}{3}(2u + u_x) = 6uu_x. \tag{3.8b}$$

The unmodified $(1+1)$-dimensional "Monge–Ampere" integrable PDE (see, for instance, [24, 31]) reads

$$w_{\tau\tau}w_{\xi\xi} - w_{\xi\tau}^2 = 0. \tag{3.9a}$$

The corresponding $\omega$-modified PDE is in this case $t$-autonomous for any choice of $\lambda$ and $\mu$

$$u_{tt}u_{xx} - (u_{tx})^2 + i\omega[-(2\lambda+1)u_tu_{xx} + 2(\lambda+\mu)u_{tx}u_x] + \omega^2[-\lambda(\lambda+1)uu_{xx} + \mu(\mu-1)xu_xu_{xx} + (\lambda+\mu)^2(u_x)^2] = 0. \tag{3.9b}$$

The general solution to this PDE (3.9b) is given in two steps: first, for any arbitrary function $F(r)$, find the function $r(x;t)$ from the nondifferential equation:

$$r(x;t) = xe^{i(\mu-1)\omega t} - e^{-i\omega t}F[r(x;t)]; \tag{3.9c}$$

then, for any arbitrary function $G(r)$ and constant $a$, evaluate the solution

$$u(x;t) = e^{i\lambda\omega t}\int_a^t dt' e^{i\omega t'}G[r(x;t')]. \tag{3.9d}$$

A class of explicit solutions to this PDE (3.9b) is

$$u(x;t) = e^{i(\lambda+\mu)\omega t}xf\left[\frac{\sin(\omega t/2)}{xe^{i(\mu-1/2)\omega t}}\right], \tag{3.9e}$$

where $f(z)$ is an arbitrary function. Three particularly neat cases of the $\omega$-modified evolution PDE (3.9b) are worth explicit display:

for $\lambda = \mu = 0$,

$$u_{tt}u_{xx} - u_{tx}^2 - i\omega u_tu_{xx} = 0; \tag{3.9f}$$

for $\lambda = -1$, $\mu = 1$,

$$u_{tt}u_{xx} - u_{tx}^2 + i\omega u_tu_{xx} = 0; \tag{3.9g}$$

for $\lambda = -1/2$, $\mu = 1/2$,

$$u_{tt}u_{xx} - u_{tx}^2 + \left(\frac{\omega}{2}\right)^2(u - xu_x)u_{xx} = 0. \tag{3.9h}$$

The following unmodified $(1+1)$-dimensional solvable PDE reads

$$w_{\tau\tau}w_\xi - w_{\tau\xi}w_\tau = 0. \tag{3.10a}$$

The corresponding $\omega$-modified PDE is in this case $t$-autonomous for any choice of $\lambda$ and $\mu$

$$u_{tt}u_x - u_{tx}u_t + i\omega[(-\lambda+\mu+1)u_tu_x + \lambda uu_{tx} - \mu xu_{tx}u_x + \mu xu_tu_{xx}] + \omega^2[\lambda(\mu-1)uu_x + (-\lambda+\mu-2)\mu x(u_x)^2 + \lambda\mu xuu_{xx}] = 0. \tag{3.10b}$$

The general solution to this PDE (3.10b) reads

$$u(x;t) = e^{i\lambda\omega t}f\left[g\left(xe^{i\mu\omega t}\right) + e^{i\omega t}\right], \tag{3.10c}$$

where $f(z)$ and $g(z)$ are two arbitrary analytic functions.

The following unmodified $(1+1)$-dimensional "Boussinesq" equation is integrable (see, for instance, [32]):

$$w_{\tau\tau} = (w_{\xi\xi\xi} + ww_\xi)_\xi. \tag{3.11a}$$

By setting $\lambda = 1$ and $\mu = 1/2$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt} - i\omega u - \frac{i\omega}{2}xu_x = (u_{xxx} + uu_x)_x. \tag{3.11b}$$

The following unmodified $(1+1)$-dimensional "nonlinear wave" equation (see, for instance, [24]) reads

$$w_{\tau\tau} = (w^\alpha w_\xi)_\xi. \tag{3.12a}$$

By setting $\mu = 1 - \alpha\lambda/2$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt} - i(2\lambda+1)\omega u_t - i\omega(2-\alpha\lambda)xu_{tx} - \lambda(\lambda+1)\omega^2u - (2-\alpha\lambda)\left(\lambda+1-\frac{\alpha\lambda}{4}\right)\omega^2xu_x - \left(1-\frac{\alpha\lambda}{2}\right)^2\omega^2x^2u_{xx} = (u^\alpha u_x)_x. \tag{3.12b}$$

Two special cases of this nonlinear PDE warrant explicit display:

$$u_{tt} + \left(\frac{\omega}{2}\right)^2u = \left(\frac{u_x}{u^4}\right)_x \tag{3.12c}$$

corresponding to $\alpha = -4$, $\lambda = -1/2$;

$$u_{tt} - 5\omega u_t - 6\omega^2u = (uu_x)_x \tag{3.12d}$$

corresponding to $\alpha = 1$, $\lambda = 2$ (some solutions to this $\omega$-modified PDEs are exhibited in [21]).

Another class (out of many possible ones) of unmodified $(1+1)$-dimensional "nonlinear wave" equations reads

$$w_{\tau\tau} = \sum_k \frac{a_k}{w^{3+\alpha_k}}\left(\frac{\partial^{p_k}w}{\partial\xi^{p_k}}\right)^{\alpha_k}, \tag{3.13a}$$

where the numbers $p_k$ are nonnegative integers, or possibly just integers: here and hereafter, we indicate

$$\frac{\partial^{-p}}{\partial\xi^{-p}}f(\xi)=\int^{\xi}d\xi_p\int^{\xi_p}d\xi_{p-1}\ldots\int^{\xi_2}f(\xi_1)d\xi_1.$$

By setting $\lambda=-1/2$ and $\mu=0$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt}+\left(\frac{\omega}{2}\right)^2u=\sum_k\frac{a_k}{u^{3+\alpha_k}}\left(\frac{\partial^{p_k}u}{\partial x^{p_k}}\right)^{\alpha_k}. \quad (3.13b)$$

The following unmodified $(1+1)$-dimensional system of two coupled PDEs is integrable (see, for instance, [33]):

$$w_\tau=w_{\xi\xi}+\tilde w^2,\quad \tilde w_\tau=w_{\xi\xi}. \quad (3.14a)$$

By setting $\lambda=\tilde\lambda=1$ and $\mu=1/2$, one gets the corresponding $\omega$-modified system:

$$u_t-i\omega u-\frac{i\omega}{2}xu_x=u_{xx}+\tilde u^2, \quad (3.14b)$$
$$\tilde u_t-i\omega\tilde u-\frac{i\omega}{2}x\tilde u_x=u_{xx}.$$

The following unmodified $(1+1)$-dimensional system of two coupled PDEs is integrable (see, for instance, [34, 33]):

$$w_\tau=a\left(w\tilde w\right)_\xi,\quad \tilde w_\tau=\left(bw+c\tilde w^2\right)_\xi. \quad (3.15a)$$

By setting $\lambda=2(1-\mu)$ and $\tilde\lambda=1-\mu$, one gets the corresponding $\omega$-modified system:

$$u_t-2i(1-\mu)\omega u-i\mu\omega xu_x=a\left(u\tilde u\right)_x, \quad (3.15b)$$
$$\tilde u_t-i(1-\mu)\omega\tilde u-i\mu\omega x\tilde u_x=\left(bu+c\tilde u^2\right)_x.$$

The following unmodified $(1+1)$-dimensional "Zakharov–Shabat" system of two coupled PDEs is integrable (see [35, 36]):

$$w_\tau+w_{\xi\xi}=w^2\tilde w,\quad \tilde w_\tau-\tilde w_{\xi\xi}=-\tilde w^2w. \quad (3.16a)$$

By setting $\tilde\lambda=1-\lambda$ and $\mu=1/2$, one gets the corresponding $\omega$-modified system:

$$u_t-i\lambda\omega u-i\frac{\omega}{2}xu_x+u_{xx}=u^2\tilde u, \quad (3.16b)$$
$$\tilde u_t-i(1-\lambda)\omega\tilde u-i\frac{\omega}{2}x\tilde u_x-\tilde u_{xx}=-u\tilde u^2.$$

The following unmodified $(1+1)$-dimensional "Wadati–Konno–Ichikawa" system of two coupled PDEs is integrable (see [37]):

$$w_\tau=a\left(\frac{w}{\sqrt{1+w\tilde w}}\right)_{\xi\xi},\quad \tilde w_\tau=b\left(\frac{\tilde w}{\sqrt{1+w\tilde w}}\right)_{\xi\xi}. \quad (3.17a)$$

By setting $\tilde\lambda=-\lambda$ and $\mu=1/2$, one gets the corresponding $\omega$-modified system:

$$u_t-i\lambda\omega u-i\frac{\omega}{2}xu_x=a\left(\frac{u}{\sqrt{1+u\tilde u}}\right)_{xx}, \quad (3.17b)$$
$$\tilde u_t+i\lambda\omega\tilde u-i\frac{\omega}{2}x\tilde u_x=b\left(\frac{\tilde u}{\sqrt{1+u\tilde u}}\right)_{xx}.$$

The following unmodified $(1+1)$-dimensional "Landau–Lifshitz" system of two coupled PDEs is integrable (see, for instance, [38, 33]):

$$w_\tau=-\sin(w)\tilde w_{\xi\xi}-2\cos(w)w_\xi\tilde w_\xi \quad (3.18a)$$
$$+(a-b)\sin(w)\cos(\tilde w)\sin(\tilde w),$$
$$\tilde w_\tau=\frac{w_{\xi\xi}}{\sin(w)}-\cos(w)\left(\tilde w_\xi\right)^2$$
$$+\cos(w)\left(a\cos^2(\tilde w)+b\sin^2(\tilde w)+c\right).$$

By setting $\lambda=\tilde\lambda=0$ and $\mu=1/2$, one gets the corresponding $\omega$-modified system

$$u_t-\frac{i\omega}{2}xu_x=-\sin(u)\tilde u_{xx}-2\cos(u)u_x\tilde u_x \quad (3.18b)$$
$$+(a-b)\sin(u)\cos(\tilde u)\sin(\tilde u),$$
$$\tilde u_t-\frac{i\omega}{2}x\tilde u_x=\frac{u_{xx}}{[\sin(u)]}-\cos(u)\tilde u_x^2$$
$$+\cos(u)\left(a\cos^2(\tilde u)+b\sin^2(\tilde u)+c\right).$$

Note the simplification if $a=b$ and, moreover, if $c=-a$.

The following unmodified $(2+1)$-dimensional integrable PDE (see, for instance, [24]) reads

$$w_\tau=aw_\eta+bww_\xi. \quad (3.19a)$$

By setting $\lambda=0$ and $\mu=\nu=1$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t-i\omega\left(xu_x+yu_y\right)=au_y+buu_x. \quad (3.19b)$$

The following unmodified $(2+1)$-dimensional integrable PDE (see, for instance, [24]) reads

$$w_\tau=aw_\eta+b\left(w_\xi\right)^2. \quad (3.20a)$$

By setting $\lambda=0$, $\mu=1/2$, and $\nu=1$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t-i\omega\left(\frac{x}{2}u_x+yu_y\right)=au_y+b\left(u_x\right)^2. \quad (3.20b)$$

The following unmodified $(2+1)$-dimensional PDE reads

$$w_\tau=a\left(w_\xi w_\eta-ww_{\xi\eta}\right)^\alpha. \quad (3.21a)$$

By setting $\lambda=1/(2\alpha-1)$ and $\mu=\nu=0$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t-\frac{i\omega}{2\alpha-1}u=a\left(u_xu_y-uu_{xy}\right)^\alpha. \quad (3.21b)$$

A (rather trivial) separable solution of this PDE reads

$$u(x,y;t) = \exp\left(\frac{i\omega t}{2\alpha-1}\right) f(x)g(y), \qquad (3.21c)$$

where $f(x)$ and $g(y)$ are two arbitrary functions.

The following unmodified $(2+1)$-dimensional PDE reads

$$w_{\tau\tau} = a(w_\xi w_\eta - w w_{\xi\eta})^\alpha. \qquad (3.22a)$$

By setting $\lambda = 2/(2\alpha-1)$ and $\mu = \nu = 0$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt} - \frac{2\alpha+3}{2\alpha-1} i\omega u_t - \frac{2(2\alpha+1)}{(2\alpha-1)^2}\omega^2 u = a(u_x u_y - u u_{xy})^\alpha. \qquad (3.22b)$$

A (rather trivial) separable solution of this PDE reads

$$u(x,y;t) = \left[b\exp\left(\frac{2i\omega t}{2\alpha-1}\right) + c\exp\left(\frac{2\alpha+1}{2\alpha-1} i\omega t\right)\right] f(x)g(y), \qquad (3.22c)$$

where $a$ and $b$ are two arbitrary constants and $f(x)$ and $g(y)$ are two arbitrary functions.

The following unmodified $(2+1)$-dimensional system of two coupled PDEs is integrable (see, for instance, [39]):

$$w_\tau + w_{\xi\xi\xi} = (w\tilde{w})_\xi, \quad \tilde{w}_\eta = w_\xi. \qquad (3.23a)$$

By setting $\lambda = \nu + 1/3$, $\tilde{\lambda} = 2/3$, $\mu = 1/3$, and, for notational convenience, $\Omega = \omega/3$, one gets the corresponding $\Omega$-modified system:

$$u_t + u_{xxx} - i(3\nu+1)\Omega u - i\Omega x u_x - 3i\nu\Omega y u_y = (u\tilde{u})_x, \quad \tilde{u}_y = u_x. \qquad (3.23b)$$

The following unmodified $(2+1)$-dimensional "long-wave equation" system of two coupled PDEs is integrable (see, for instance, [40, 41]):

$$w_{\tau\eta} + \tilde{w}_{\xi\xi} = \frac{1}{2}\left(w^2\right)_{\xi\eta}, \qquad (3.24a)$$
$$\tilde{w}_\tau + w_{\xi\xi} = \left(w\tilde{w} + w_{\xi\eta}\right)_\xi.$$

By setting $\lambda = 1/2$, $\tilde{\lambda} = 0$, $\mu = 1/2$, $\nu = -1/2$, and, for notational convenience, $\Omega = \omega/2$, one gets the corresponding $\omega$-modified system:

$$u_{ty} - i\Omega u_y - i\Omega\left(x u_{xy} - y u_{yy}\right) + \tilde{u}_{xx} = \frac{1}{2}\left(u^2\right)_{xy}, \qquad (3.24b)$$
$$\tilde{u}_t - i\Omega(x\tilde{u}_x - y\tilde{u}_y) + u_{xx} = (u\tilde{u} + u_{xy})_x.$$

The following unmodified $(2+1)$-dimensional system of two coupled PDEs is integrable (see, for instance, [42]):

$$w_\tau + w_{\xi\xi\eta} = (w^2)_\eta + w_\xi\tilde{w}, \quad \tilde{w}_\xi = w_\eta. \qquad (3.25a)$$

By setting $\tilde{\lambda} = 1 - \lambda/2$, $\mu = \lambda/2$, and $\nu = 1-\lambda$, one gets the corresponding $\omega$-modified system:

$$u_t - i\lambda\omega u - \frac{i\lambda\omega}{2} x u_x - i(1-\lambda)\omega y u_y + u_{xxy} = (u^2)_y + u_x\tilde{u}, \quad \tilde{u}_x = u_y. \qquad (3.25b)$$

A nontrivial family of solutions of this system (3.25b) reads as follows:

$$u(x,y;t) = \frac{-4a^2 e^{i\lambda\omega t}}{\left\{\cosh\left[axe^{i\lambda\omega t/2} - f(t) - g\left(ye^{i(1-\lambda)\omega t} - be^{i\omega t}\right)\right]\right\}^2}, \qquad (3.25c)$$
$$\tilde{u}(x,y;t) = \frac{e^{i\lambda\omega t/2}}{a}\left\{ - f'(t) + e^{i\omega t} g'\left(ye^{i(1-\lambda)\omega t} - be^{i\omega t}\right)\left[i\omega b - 4a^2 + \left(\frac{2a}{\cosh\left[axe^{i\lambda\omega t/2} - f(t) - g(ye^{i(1-\lambda)\omega t} - be^{i\omega t})\right]}\right)^2\right]\right\},$$

where $a$ and $b$ are two arbitrary constants and $f(t)$ and $g(z)$ are arbitrary functions (and of course $f'(t)$ and $g'(z)$ denote their derivatives). Conditions sufficient to guarantee that this solution will be isochronous are given in [21].

The following unmodified $(2+1)$-dimensional "matrix KP" system of two coupled matrix PDEs is integrable (see, for instance, [43]):

$$W_\tau + W_{\xi\xi\xi} - 3\tilde{W}_\eta = 3\left(W^2\right)_\xi + 3i\left[W, \tilde{W}\right], \quad \tilde{W}_\xi = W_\eta. \qquad (3.26a)$$

Here, $W \equiv W(\xi,\eta;\tau)$ and $\tilde{W} \equiv \tilde{W} \equiv \tilde{W}(\xi,\eta;\tau)$ are matrices (of course, of the same rank), and the no-

tation $\left[W,\tilde{W}\right]$ denotes their commutator. By setting $\lambda=2/3$, $\tilde{\lambda}=1$, $\mu=1/3$, and $\nu=1$, one gets the corresponding $\omega$-modified system:

$$U_t+U_{xxx}-3\tilde{U}_y-\frac{2}{3}i\omega U \tag{3.26b}$$
$$-\frac{1}{3}i\omega xU_x-i\omega yU_y=3\left(U^2\right)_x+3i\left[U,\tilde{U}\right],$$
$$\tilde{U}_x=U_y.$$

The following class of unmodified $(N+1)$-dimensional PDEs,

$$\frac{\partial^{m+1}w}{\partial\tau\partial\xi^m}=f(w), \tag{3.27a}$$

where $m$ is a positive integer (or possibly just an integer) and $f(w)$ is an arbitrary analytic function, gets transformed, by setting $\lambda=0$ and $\mu=-1/m$, into the corresponding $\omega$-modified evolution PDE

$$\frac{\partial^{m+1}u}{\partial t\partial x^m}+i\omega\frac{\partial^m u}{\partial x^m}+i\frac{\omega}{m}x\frac{\partial^{m+1}u}{\partial x^{m+1}}=f(u). \tag{3.27b}$$

For instance, for $m=1$ and $f(w)=\exp(aw)$, Eq. (3.27a) becomes the Liouville equation

$$w_{\tau\xi}=\exp(aw), \tag{3.28a}$$

and the corresponding $\omega$-modified evolution PDE (3.27b) reads

$$u_{tx}+i\omega u_x+i\omega xu_{xx}=\exp(au). \tag{3.28b}$$

The general solution to this $\omega$-modified Liouville equation reads

$$u(x;t)=g(xe^{-i\omega t})+f(t) \tag{3.28c}$$
$$-\frac{2}{a}\ln\left\{be^{-i\omega t}\int_{x_0}^{x}dy\exp\left[ag\left(ye^{-i\omega t}\right)\right]\right.$$
$$\left.-\frac{a}{2b}\int_0^t ds\exp\left[i\omega s-af(s)\right]\right\},$$

where $b$ is an arbitrary (nonvanishing) complex constant, $x_0$ is an arbitrary real constant, $g(x)$ is an arbitrary analytic function, and $f(t)$ is as well an arbitrary function of the real time variable $t$, but it must of course be periodic with period $T$ [see (1.1)] for the isochronicity property to hold (other conditions on this general solution sufficient to guarantee this are given in [21]).

The following unmodified $(N+1)$-dimensional PDE reads

$$\frac{\partial^{m+2}w}{\partial\xi^m\partial\tau^2}=f(w), \tag{3.29a}$$

where $m$ is an integer and $f(w)$ is an analytic function. By setting $\lambda=0$ and $\mu=-2/m$, one gets the corresponding $\omega$-modified evolution PDE

$$\frac{\partial^{m+2}u}{\partial t^2\partial x^m}+3i\omega\frac{\partial^{m+1}u}{\partial t\partial x^m}+4i\omega x\frac{\partial^{m+2}u}{\partial t\partial x^{m+1}} \tag{3.29b}$$
$$-2\omega^2\frac{\partial^m u}{\partial x^m}-\frac{2}{m}\left(3+\frac{2}{m}\right)\omega x\frac{\partial^{m+1}u}{\partial x^{m+1}}$$
$$-\left(\frac{2}{m}\right)^2\omega^2x^2\frac{\partial^{m+2}u}{\partial x^{m+2}}=f(u).$$

The following unmodified $(N+1)$-dimensional PDE reads

$$w_{\tau\tau}=w_\tau\sum_{n=1}^{N}\left[a_n\left(w_{\xi_n}\right)^{\alpha_n}\right] \tag{3.30a}$$
$$+\sum_{n=1}^{N}\left[b_n\left(w_{\xi_n}\right)^{2(\beta\alpha_n-1)}w_{\xi_n\xi_n}\right].$$

By setting $\lambda=\beta-1$ and $\mu_n=1-\beta+1/\alpha_n$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt}-i\omega\left[(2\beta-1)u_t\right. \tag{3.30b}$$
$$\left.+\sum_{n=1}^{N}\left(1-\beta+\frac{1}{\alpha_n}\right)x_nu_{tx_n}\right]$$
$$-\omega^2\left[\beta(\beta-1)u+\beta\sum_{n=1}^{N}\left(1-\beta+\frac{1}{\alpha_n}\right)x_nu_{x_n}\right]$$
$$=\left\{u_t-i\omega\left[(\beta-1)u\right.\right.$$
$$\left.\left.+\sum_{n=1}^{N}\left(1-\beta+\frac{1}{\alpha_n}\right)x_nu_{x_n}\right]\right\}$$
$$\times\sum_{n=1}^{N}\left[a_n\left(u_{x_n}\right)^{\alpha_n}\right]+\sum_{n=1}^{N}\left[b_n\left(u_{x_n}\right)^{2(\beta\alpha_n-1)}u_{x_nx_n}\right].$$

The following unmodified $(N+1)$-dimensional "nonlinear diffusion" PDE (see, for instance, [24]) reads

$$w_\tau=w^\alpha\sum_{n=1}^{N}w_{\xi_n\xi_n}. \tag{3.31a}$$

By setting $\mu_n=(1-\alpha\lambda)/2$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t-i\lambda\omega u-i\frac{1-\alpha\lambda}{2}\omega\sum_{n=1}^{N}x_nu_{x_n} \tag{3.31b}$$

$$= u^{\alpha} \sum_{n=1}^{N} u_{x_n x_n}.$$

Note the simplification for $\lambda = 1/\alpha$.

The following unmodified $(N+1)$-dimensional "nonlinear heat equation with a source" (see, for instance, [24]) reads

$$w_\tau = \sum_{n=1}^{N} a_n \left(w^{\alpha_n} w_{\xi_n}\right)_{\xi_n} + b w^{\beta}. \tag{3.32a}$$

By setting $\lambda = 1/(\beta-1)$ and $\mu_n = (\beta - \alpha_n)/[2 \times (\beta-1)]$, one gets the corresponding $\omega$-modified evolution PDE

$$u_t - \frac{i\omega}{\beta-1} u - \frac{i\omega}{2} \sum_{n=1}^{N} \frac{(\beta-\alpha_n)}{(\beta-1)} x_n u_{x_n} \tag{3.32b}$$
$$= \sum_{n=1}^{N} a_n \left(u^{\alpha_n} u_{x_n}\right)_{x_n} + b u^{\beta}.$$

The following unmodified $(N+1)$-dimensional solvable "Bateman" PDE (see, for instance, [44]) reads

$$\det \begin{pmatrix} 0 & w_\tau & w_{\xi_1} & \cdots & w_{\xi_n} \\ w_\tau & w_{\tau\tau} & w_{\xi_1\tau} & \cdots & w_{\xi_n\tau} \\ w_{\xi_1} & w_{\tau\xi_1} & w_{\xi_1\xi_1} & \cdots & w_{\xi_n\xi_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{\xi_n} & w_{\tau\xi_n} & w_{\xi_1\xi_n} & \cdots & w_{\xi_n\xi_n} \end{pmatrix} = 0. \tag{3.33a}$$

The corresponding $\omega$-modified PDE is in this case $t$-autonomous for any choice of $\lambda$ and $\mu_n$. For $\lambda = \mu_n = 0$, it reads

$$\det \begin{pmatrix} 0 & u_t & u_{x_1} & \cdots & u_{x_n} \\ u_t & u_{tt} & u_{x_1 t} & \cdots & u_{x_n t} \\ u_{x_1} & u_{x_1 t} & u_{x_1 x_1} & \cdots & u_{x_n x_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{x_n} & u_{t x_n} & u_{x_1 x_n} & \cdots & u_{x_n x_n} \end{pmatrix} \tag{3.33b}$$
$$- i\omega u_t \det \begin{pmatrix} 0 & u_{x_1} & \cdots & u_{x_n} \\ u_{x_1} & u_{x_1 x_1} & \cdots & u_{x_n x_1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{x_n} & u_{x_1 x_n} & \cdots & u_{x_n x_n} \end{pmatrix} = 0.$$

The general solution $u \equiv u(\underline{x}; t)$ to this PDE (3.33b) is given by the implicit formula

$$(e^{i\omega t} - 1) f_0(u) + \sum_{k=1}^{n} x_k f_k(u) = c, \tag{3.33c}$$

where the $N+1$ functions $f_k(z)$, $k = 0, 1, \ldots, N$, are arbitrary. For $N = 1$, the unmodified $(1+1)$-dimensional Bateman equation reads

$$w_{\tau\tau} (w_\xi)^2 + w_{\xi\xi} w_\tau^2 - 2 w_\xi w_\tau w_{\xi\tau} = 0, \tag{3.34a}$$

and the $\omega$-modified version of this equation reads

$$u_{tt} u_x^2 + u_{xx} u_t^2 - 2 u_{xt} u_x u_t \tag{3.34b}$$
$$+ i\omega [2\lambda u (u_x u_{xt} - u_t u_{xx}) + (2\mu - 1)(u_x)^2 u_t]$$
$$+ \omega^2 [\lambda^2 u (u_x)^2 - \lambda^2 u^2 u_{xx} + \lambda (2\mu - 1) u (u_x)^2$$
$$+ \mu(\mu - 1) x (u_x)^3] = 0.$$

The general solution to this equation (3.34b) reads (in implicit form)

$$(e^{i\omega t} - 1) f\left(e^{-i\lambda\omega t} u(x;t)\right) \tag{3.34c}$$
$$+ x e^{i\mu\omega t} g\left(e^{-i\lambda\omega t} u(x;t)\right) = c,$$

with $f(z)$ and $g(z)$ two arbitrary functions which can be easily determined in terms of the initial data, say, $u_0(x) = u(x;0)$ and $u_1(x) = u_t(x;0)$. By setting $\lambda = 0$, $\mu = 1/2$, and (for notational convenience) $\Omega = \omega/2$, the $\omega$-modified $(1+1)$-dimensional Bateman equation (3.34b) takes the simple (real) form

$$u_{tt} u_x^2 + u_{xx} u_t^2 - 2 u_{xt} u_x u_t = \Omega^2 x u_x^3. \tag{3.34d}$$

Note that, if $u(x;t)$ is a solution to this PDE, $v(x;t) = f\left[u(ax; t-b)\right]$ is also a solution, with $f(z)$ an arbitrary function and $a$ and $b$ two arbitrary constants. The initial-value problem for this equation is solved by the implicit formula

$$u(x;t) \tag{3.34e}$$
$$= u_0 \left( \frac{\tan(\Omega t)}{\Omega} \frac{u_1\left(u_0^{(\mathrm{inv})}\left[u(x;t)\right]\right)}{u_0'\left(u_0^{(\mathrm{inv})}\left[u(x;t)\right]\right)} + \frac{x}{\cos[\Omega t]} \right),$$

where of course $u_0^{(\mathrm{inv})}(z)$ (respectively, $u_0'(z)$) are the inverse (respectively, the derivative) of the function $u_0(z)$ (namely, $u_0^{(\mathrm{inv})}\left[u_0(x)\right] = x$, $u_0'(x) = du_0(x)/dx$). And two explicit solutions of this equation (3.34d) read as follows:

$$u(x;t) = f\left\{ \frac{c_1 x + c_2 \cos\left[\Omega(t - c_3)\right]}{c_4 x + c_5 \cos\left[\Omega(t - c_6)\right]} \right\}, \tag{3.34f}$$

$$u(x;t) = f\left\{ \frac{c_1 \cos\left[\Omega(t - c_2)\right]}{\cos\left[2\Omega(t - c_3)\right] + \cos\left[2\Omega(c_2 - c_3)\right]} x \right. \tag{3.34g}$$

$$+ c_4 \tan\left[\Omega(t + c_2 - 2c_3)\right] \Big\}.$$

Here, $f(z)$ denotes an arbitrary (twice differentiable) function, and the constants $c_k$ are arbitrary. Clearly, these solutions are real if the function $f(z)$ and the constants $c_k$ are themselves real, and the conditions that the function $f(z)$ and the constants $c_k$ must satisfy in order to guarantee the isochronicity of these solutions are rather obvious.

The following unmodified $(N+1)$-dimensional PDE reads

$$w_{\tau\tau} = w_\tau^2 \tag{3.35a}$$
$$\times f\left(w, w_{\xi_1}, \ldots, w_{\xi_N}, w_{\xi_1\xi_1}, \ldots, w_{\xi_N\xi_N}, w_{\xi_1\xi_1\xi_1} \ldots\right),$$

where $f(w,w_{x_1},\ldots,w_{\xi_N},w_{\xi_1\xi_1},\ldots,w_{\xi_N\xi_N},w_{\xi_1\xi_1\xi_1}\ldots)$ is an arbitrary analytic function. By setting $\lambda = \mu_n = 0$, one gets the corresponding $\omega$-modified evolution PDE

$$u_{tt} - i\omega u = u_t^2 \tag{3.35b}$$
$$\times f\left(u, u_{x_1}, \ldots, u_{x_N}, u_{x_1x_1}, \ldots, u_{x_Nx_N}, u_{x_1x_1x_1} \ldots\right).$$

## 4. OUTLOOK

As mentioned in the introductory Section 1, each of the $\omega$-modified equations reported above would require a separate investigation to exhibit the existence and characterize the properties of isochronous solutions. This shall eventually be done, probably mainly in the context of possible applications of these equations. Throughout this paper, we focused on $\omega$-modified equations that feature many isochronous solutions. By a variant of the trick (2.1)–(2.3), it is in some cases possible to generate equations that feature many solutions which are completely periodic not only in the time variable $t$ but also in the space variable $x$. We only exhibit here a single example, obtained from the unmodified PDE (3.1a) via the following change of variables:

$$\tau = \frac{e^{i\omega t} - 1}{i\omega}, \tag{4.1a}$$

$$\xi = \frac{e^{ikx} - 1}{ik}, \tag{4.1b}$$

$$u(x;t) = e^{i\lambda\omega t} e^{i\rho kx} w(\xi;\tau), \tag{4.1c}$$

where, to apply this transformation to the unmodified PDE (3.1a), we set $\lambda = 1/\alpha$ and $\rho = -1/\alpha$, while $\omega$ and $k$ are two real (indeed, without loss of generality, positive) constants that determine the basic period in the (real) time variable $t$ [see (1.1)] and the basic period, $L = 2\pi/k$, in the (also real) space variable $x$. The modified PDE corresponding to (3.1a) then reads

$$u_t - \frac{i\omega}{\alpha} u = a u^\alpha \left(u_x + \frac{ik}{\alpha} u\right), \tag{4.2a}$$

and the general solution, in implicit form, to this PDE reads

$$u(x;t) = e^{i\omega t/\alpha} \left\{1 + \frac{ak}{\omega}\left(1 - e^{-i\omega t}\right)\left[u(x;t)\right]^\alpha\right\}^{1/\alpha} \tag{4.2b}$$
$$\times\, u_0\left(x - \frac{i}{k}\log\left\{1 + \frac{ak}{\omega}\left(1 - e^{-i\omega t}\right)\left[u(x;t)\right]^\alpha\right\}\right),$$

where of course the initial datum, $u_0(x) = u(x;0)$, should itself be periodic with period $L$ (or some appropriate integer multiple, or fraction, of $L$), in order for this solution to be periodic for all time with period $L$ (or some appropriate integer multiple, or fraction, of $L$)—in addition of course to being periodic in $t$ with period $T$ [see (1.1)] or with some integer multiple of $T$, depending on the analyticity properties of $u_0(z)$ as a function of the complex variable $z$. An explicit special case of this implicit equation (corresponding to $u_0(x) = e^{iqx}$) reads

$$u(x;t) = b e^{i\omega t/\alpha} e^{iqx} \tag{4.2c}$$
$$\times \left\{1 + \frac{ak}{\omega}\left(1 - e^{-i\omega t}\right)\left[u(x;t)\right]^\alpha\right\}^{1/\alpha + q/k},$$

yielding, for $q = -k/\alpha$, the trivial solution to (4.2a)

$$u(x;t) = b \exp\left[i\frac{(\omega t - kx)}{\alpha}\right]. \tag{4.2d}$$

For $q = k/\alpha$, one obtains instead from (4.2c) the explicit solution to (4.2a)

$$u(x;t) = e^{i(\omega t - kx)/\alpha} \Bigg\{\frac{\omega}{2a^2 b k^2 \left(e^{i\omega t} - 1\right)^2} \tag{4.2e}$$
$$\times \Big[\omega - 2abe^{ikx}(e^{i\omega t} - 1)$$
$$- \sqrt{\omega^2 - 4abk\omega e^{ikx}(e^{i\omega t} - 1)}\Big]\Bigg\}^{1/\alpha}.$$

Of course, many other explicit solutions could be exhibited in specific cases, for instance, in the special case with $q = -k$ and $\alpha = 2$

$$u(x;t) \tag{4.2f}$$
$$= \pm\sqrt{\frac{\omega \pm \sqrt{4ab^2 k\omega e^{-2ikx}(e^{i\omega t} - 1) + \omega^2}}{2ak\left(1 - e^{i\omega t}\right)}}.$$

Finally, we also rewrite below the evolution PDE (4.2a) with $\alpha = 1$ in real form, setting $u = u_1 + iu_2$ and $a = c_1 + ic_2$, where the two dependent variables $u_1 \equiv u_1(x;t)$ and $u_2 \equiv u_2(x;t)$, as well as the two constants $c_1$ and $c_2$, are of course now real:

$$u_{1_t} + \omega u_2 = c_1\left[u_1 u_{1_x} + u_2 u_{2_x} - 2k u_1 u_2\right] \tag{4.3}$$
$$- c_2\left[u_1 u_{2_x} + u_2 u_{1_x} + k\left(u_1^2 - u_2^2\right)\right],$$

$$u_{2_t} - \omega u_1 = c_2 \left[u_1 u_{1_x} + u_2 u_{2_x} - 2k u_1 u_2\right] + c_1 \left[u_1 u_{2_x} + u_2 u_{1_x} + k\left(u_1^2 - u_2^2\right)\right].$$

## REFERENCES


1. F. Calogero, J. Math. Phys. **38**, 5711 (1997).
2. F. Calogero, Lect. Notes Phys. M **66** (2001).
3. F. Calogero and J.-P. Françoise, J. Nonlinear Math. Phys. **9**, 99 (2002).
4. F. Calogero, in *Proceedings of the Workshop on Calogero–Moser–Sutherland Models, Montreal, 1997*, Ed. by J. F. van Diejen and L. Vinet (Springer, 2000), p. 93.
5. F. Calogero and J.-P. Françoise, Ann. Inst. Henri Poincaré **1**, 173 (2000).
6. F. Calogero and J.-P. Françoise, Inverse Probl. **17**, 1 (2001).
7. F. Calogero, Phys. Lett. A **293**, 146 (2002).
8. F. Calogero, J. Phys. A **35**, 985 (2002).
9. F. Calogero, J. Phys. A **35**, 4249 (2002).
10. F. Calogero and M. Sommacal, J. Nonlinear Math. Phys. **9**, 483 (2002).
11. F. Calogero, J.-P. Françoise, and M. Sommacal, J. Nonlinear Math. Phys. **10**, 157 (2003).
12. F. Calogero, J. Phys. A **35**, 3619 (2002).
13. F. Calogero and J.-P. Françoise, Teor. Mat. Fiz. **137**, 1663 (2003).
14. F. Calogero and V. I. Inozemtsev, J. Phys. A **35**, 10365 (2002).
15. F. Calogero, J. Phys. A **36**, 7291 (2003).
16. F. Calogero and J.-P. Françoise, in *Proceedings of the Workshop on Superintegrability in Classical and Quantum Systems, Montreal, 2002* (in press).
17. F. Calogero, in *Proceedings of the NATO Advanced Research Workshop, Cadiz, 2002*, Ed. by A. B. Shabat, A. Gonzalez-Lopez, M. Manas, *et al.*, NATO Sci. Ser., II. Math., Phys. and Chem. **132**, 49 (2004).
18. F. Calogero, J. Nonlinear Math. Phys. **11**, 1 (2004).
19. F. Calogero, Appl. Anal. (in press).
20. F. Calogero, J. Nonlinear Math. Phys. **11**, 208 (2004).
21. M. Mariani, in *Tesi di Laurea in Fisica, Dipartimento di Fisica, Università di Roma "La Sapienza," 2003*.
22. F. Calogero, London Math. Soc. Lect. Notes **295**, 9 (2003).
23. F. Calogero and M. Mariani, *A Modified Schwarzian Korteweg de Vries Equation in* $2+1$ *Dimensions with Lots of Periodic Solutions* (in press).
24. N. H. Ibragimov, *CRC Handbook of Lie Group Analysis of Differential Equations* (CRC Press, 1994–1996).
25. E. Hopf, Comm. Pure Appl. Math. **3**, 201 (1950).
26. F. Calogero, J. Math. Phys. **28**, 538 (1987).
27. P. Bracken, Preprint No. 78541-2999 (Department of Mathematics, University of Texas at Edinburg, USA).
28. I. Dorfman, *Dirac Structures and Integrability of Nonlinear Evolution Equations* (Wiley, New York, 1993).
29. J. A. Cavalcante and K. Tenenblat, J. Math. Phys. **29**, 1044 (1988).
30. F. Calogero and A. Degasperis, *Spectral Transform and Solitons I* (North-Holland, Amsterdam, 1982).
31. D. B. Fairlie and A. N. Leznov, J. Geom. Phys. **16**, 385 (1995).
32. A. P. Fordy and J. Gibbons, J. Math. Phys. **22**, 1170 (1981).
33. J. P. Wang, J. Nonlinear Math. Phys. **9** (Suppl. 1), 213 (2002).
34. M. J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering* (Cambridge Univ. Press, Cambridge, 1991).
35. V. E. Zakharov and A. B. Shabat, Zh. Éksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1971)].
36. M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
37. M. Wadati, K. Konno, and Y. H. Ichikawa, J. Phys. Soc. Jpn. **47**, 1698 (1979).
38. T. van Bemmelen and P. Kersten, J. Math. Phys. **32**, 1709 (1991).
39. M. Boiti, J. J.-P. Leon, M. Manna, and F. Pempinelli, Inverse Problems **2**, 271 (1987).
40. M. Boiti, J. J.-P. Leon, and F. Pempinelli, Inverse Probl. **3**, 371 (1987).
41. M. Wang, Y. Zhou, and Z. Li, J. Nonlinear Math. Phys. **5**, 120 (1998).
42. M. S. Bruzòn, M. L. Gandarias, C. Muriel, *et al.*, Theor. Math. Phys. **137**, 1367 (2003).
43. B. B. Kadomtsev and B. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Sov. Phys. Dokl. **15**, 539 (1970)].
44. M. Euler, N. Euler, O. Lindblon, and L.-E. Perrson, Symmetry in Nonlinear Math. Phys. **1**, 185 (1997).